ISSN 1853-709X

Fascículo 10

Cursos y seminarios de matemática

Serie A

Roque G. Carranza

Probabilidades y estadística

Departamento de Matemática

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

2011

Cursos y Seminarios de Matemática – Serie A

Fascículo 10

ISSN 1853-709X (Versión Electrónica)
ISSN 0524-9643 (Versión Impresa)

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Ciudad Universitaria – Pabellón I
(1428) Ciudad de Buenos Aires
Argentina.
http://www.dm.uba.ar
e-mail. secre@dm.uba.ar
tel/fax: (+54-11)-4576-3335

FASCICULO **10**

cursos
y seminarios
de matemática

Roque G. Carranza

PROBABILIDADES Y ESTADISTICA

UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES — DEPARTAMENTO DE MATEMATICA
1961

ADVERTENCIA PRELIMINAR.

El presente fascículo es, con algunas ampliaciones, el desarrollo del curso de Elementos de Probabilidades y Estadística del segundo año de las licenciaturas en Matemáticas y en Física. La ubicacion del curso en los primeros años obedece a la necesidad de dar lo antes posible una introducción a los problemas y a los métodos el Cálculo de Probabilidades, cuyas aplicaciones crecen casi día por día y que no sólo es en la actualidad uno de los campos de investigación matemática que aparecen como más promisorios, sino que también abre mayores perspectivas de trabajo profesional a los futuros licenciados.

Naturalmente, esta ubicación temprana plantea el problema de la limitación de los recursos matemáticos de que disponen los alumnos, pero sin embargo se ha tratado de presentar la teoría de las variables aleatorias discretas en la forma más rigurosa y completa compatible con dicha limitación, lo que incluye un tratamiento del Problema Central del Límite más extenso y detallado que el corriente en los textos de introducción y una exposición de la importante teoría de las Cadenas de Markoff. Se ha completado esta parte con la discusión de algunos ejemplos de aplicación, especialmente a la Biología y a la Física. Se prosigue después en forma predominantemente heurística, con los elementos de la teoría de las variables aleatorias continuas, y con una introducción lo bastante extensa a la inferencia estadística como para contener al menos una explicación de los métodos más usuales. Es a esta parte que se han hecho algunos agregados a los temas que se tratan ordinariamente en el curso, especialmente en lo que se refiere a la deducción de la ley de probabilidad del coeficiente de regresión.

Buenos Aires, Junio de 1961.

INTRODUCCION

I.- LOS FENOMENOS ALEATORIOS

I.1.- El cálculo de probabilidades, tal como se lo expondrá en este curso y de acuerdo con las tendencias contemporáneas, tiene por finalidad la de constituir un modelo matemático, apto para describir e investigar el comportamiento de una clase especial y bien determinada de fenómenos: los llamados f e n o m e n o s a l e a t o r i o s.

Antes de comenzar con el desarrollo propio del curso será conveniente, por lo tanto, caracterizar de una manera suficientemente precisa que se entiende por fenómeno aleatorio, y que resultados permite obtener en su estudio el cálculo de probabilidades.

La experiencia indica que existe en la naturaleza una gran cantidad de fenómenos, que presentan las siguientes dos características comunes:

a) a pesar de repetirse en igualdad de condiciones, no se dan siempre los mismos resultados, lo que hace impredecible el resultado de una única experiencia;

b) si después de realizar un gran número de observaciones se cuentan las veces en que se ha dado un mismo resultado, y se calcula su frecuencia relativa (número de veces en que se ha presentado el resultado en cuestión, dividido por el número total de observaciones) es posible comprobar que repitiendo la serie larga de observaciones, la frecuencia relativa es constante, o aproximadamente constante(a).

Posiblemente el ejemplo más simple de los fenómenos caracterizados por este comportamiento, a los que denominaremos fenómenos aleatorios, lo constituye el más sencillo de los juegos de azar: el tiro de la moneda.

Aunque en cada tiro se cuide de arrojar la misma moneda, colocándola a la misma distancia de una superficie, con el mismo ángulo e imprimiéndole igual velocidad y aceleración en todos los tiros, es imposible predecir cual de las dos caras quedará hacia arriba al inmovilizarse sobre la superficie.

Sin embargo, si después de realizar un gran número de tiros, anotando cada vez el resultado, se calcula la frecuencia relativa conque ha aparecido una de las caras, y se repite la experiencia otra serie numerosa de veces, se podrá comprobar que la nueva frecuen

(a) La característica de impredecibilidad señalada enteriormente excluye una clase amplia de fenómenos de frecuencia relativa constante: los fenómenos periódicos. La combinación de la impredicibilidad y de la constancia de la frecuencia relativa, parece caracterizar de una manera muy peculiar el orden con que se presentan las distintas alternativas observables en un fenómeno aleatorio. Sobre la caracterización de ese orden es que Von Mises (20), intentó fundamentar rigurosamente el cálculo de probabilidades (el número se refiere a la lista bibliográfica agregada al final), pero en la actualidad parece preferible una fundamentación axiomática de otro tipo.

cia relativa diferirá muy poco de la calculada en el primer experimento.

Si el experimento se repite muchas veces, se encontrará que las frecuencias relativas son mas estables cuanto más largas sean las series de tiros que constituyen cada experiencia. Este comportamiento autoriza a decir, en un sentido que se precisará más adelante (en el capítulo IV), que la frecuencia relativa con que se presenta cada una de las caras de la moneda tiende a un límite, cuando el número de observaciones crece indefinidamente.

I.2.- A pesar de incertidumbre existente para predecir el resultado de un único tiro con una moneda, la constancia de las frecuencias, tanto más acentuada cuanto mayor sea el número de tiros, permite llegar a ciertas conclusiones útiles.

Si el jugador antes de cada tiro efectúa una apuesta de magnitud a, y obtiene un premio b si sale, por ejemplo, el resultado cara, perdiendo si se da el resultado opuesto (es decir pagando la apuesta sin compensación alguna), al cabo de n tiros su ganancia neta será:

I.1) $$g = r.b - n.a$$

en donde r es el número de veces que se ha presentado el resultado cara, el producto de r por b será por lo tanto el total de los premios percibidos, y el producto de n por a lo que ha abonado como apuesta en las n jugadas realizadas.

Si el número n de tiros, o jugadas, es lo bastante grande como para que la frecuencia relativa del resultado cara.

1.2) $$f = \frac{r}{n}$$

sea suficientemente próxima al valor límite $f^\wedge$ que experiencias anteriores han establecido como característica de la moneda utilizada, se podrá reemplazar a r en la expresión de g por el producto $f^\wedge.n$, obteniéndose.

1.3) $$g = n(f^\wedge.b - a)$$

y el jugador ganará o perderá según que el signo del paréntesis sea positivo o negativo.

En consecuencia, conociendo por experiencias anteriores el valor de $f^\wedge$ y las reglas del juego que establecen el monto de las apuestas y su relación con el premio, el jugador podrá determinar con anticipación si le conviene intervenir o no en un juego con una moneda efectuando una apuesta constante a. Si bien la previsión del resultado está condicionada a la realización de un gran número de tiros, no cabe negar que representa un progreso considerable frente a la incertidumbre de predecir el resultado en un solo tiro.

En los casos corrientes (monedas simétricas y equilibradas) $f^\wedge = 1/2$. Al jugador le resultará conveniente intervenir con una apuesta constante a, si las reglas del juego establecen un premio b tal que su producto por 1/2 sea mayor que el monto a

de la apuesta. Si dicho producto resulta menor que la apuesta, sabe por anticipado que no le conviene jugar. Si, en cambio, el producto de b por 1/2 es igual a la apuesta, el resultado le será indiferente.

En los juegos bancados de los casinos, se establecen las reglas de manera que el casino resulte beneficiado en una serie larga de jugadas.

Tomando como ejemplo el caso de la ruleta, en que la frecuencia límite de aparición de cada uno de los números es de $\frac{1}{37}$ y el jugador, en caso de que salga el número al que ha apostado, percibe como premio una suma igual a treinta y seis veces su apuesta, la aplicación de (1.3) conduce a

$$g = n\left(\frac{1}{37}.36a - a\right) = na\left(\frac{36}{37} - 1\right) < 0 \qquad (1.4)$$

Si el jugador efectúa una apuesta constante a, el juego le resultará a la larga desfavorable. Cuando se efectúan solamente series cortas de jugadas, la inestabilidad de las frecuencias relativas no permite afirmar que el paréntesis sea negativo, y el jugador podrá ganar. Pero desde el punto de vista del casino, que considera a todos los jugadores que juegan contra él, como uno solo que juega un número muy grande de veces, la situación será siempre favorable, pués entonces las frecuencias relativas diferirán poco de su límite, y la pérdida del conjunto de los jugadores es su ganancia.

I.3.- Sin embargo, la caracterización de la situación anterior es incompleta, pués parte de considerar que la frecuencia f (frecuencia relativa experimental), es estrictamente igual a $f^\wedge$ (frecuencia relativa límite), o que sus diferencias no alcanzan a alterar el signo paréntesis de (1.3) y (1.4).

Para considerar concluyente esta discusión, es necesario acotar la magnitud de las desviaciones de f con respecto a $f^\wedge$ que pueden esperarse, en función del número n de jugadas, y por otra parte, queda abierta la cuestión de si una política de apuestas adecuada, que vincule el monto de la que se realiza antes de una jugada con el resultado de la jugada o jugadas anteriores (por ejemplo, doblando la apuesta cada vez que se pierde), permitirá al jugador asegurarse un resultado siempre favorable, cualquiera sea la relación entre el producto $f^\wedge . b$ y la apuesta a. Todos estos problemas hallan solución en el cálculo de probabilidades, que confirman la discusión realizada en el párrafo I.2, precisando convenientemente los enunciados implicados.

I.4.- El caso de los juegos de azar tiene importancia histórica, porque de su estudio surgió el cálculo de probabilidades y es útil en una exposición del mismo, porque permite formular ejemplos sencillos para ilustrar sus principales resultados, sin suponer el conocimiento detallado de otras cuestiones, a menudo muy complejas.

Pero si los juegos de azar fueran la única aplicación del cálculo de probabilida

des, la importancia de este sería muy limitada. En realidad, y en cambio, los fenómenos aleatorios caracterizados en el párrafo I.1 constituyen una amplia clase, con representación en muy diversos campos de la experiencia, por lo que las conclusiones a que llegaron los matemáticos de los siglos XVII y XVIII, que iniciaron el estudio de los juegos de azar, han servido de fundamento a una rama de las matemáticas que a medida que transcurre el tiempo cuenta cada vez con más aplicaciones importantes.(a)

Cronológicamente, los primeros fenómenos aleatorios estudiados después de los juegos de azar fueron los que se presentan en la actividad del seguro.

Por ejemplo, cuando se relacionan las estadísticas de mortalidad, que registran las defunciones que ocurren en una cierta población y en un período determinado, clasificándolas por edades, con las estadísticas de población, también clasificadas por edades, es posible comprobar que la proporción de habitantes de una determinada edad que fallecen antes de alcanzar la edad siguiente, es una cantidad tanto más constante cuanto mayor es la población estudiada.

El fenómeno puede interpretarse exactamente de la misma manera que en el caso de lo juegos de azar. Para cada habitante de una cierta edad no es posible predecir si fallecerá antes de alcanzar la edad siguiente, o si sobrevivirá, lo mismo que en un único tiro de moneda no es posible predecir cuál lado se obtendrá . Pero para un grupo grande de habitantes de una misma edad es, en cambio, posible afirmar con mucha aproximación la proporción de los que fallecerán antes de alcanzar la edad siguiente, así como en un número suficientemente grande de tiros de una moneda es posible predecir muy aproximadamente la frecuencia relativa de aparición de uno de los lados.

La diferencia, aparentemente muy importante, de que en el caso de la mortalidad el conjunto de observaciones se realiza simultáneamente (en este contexto simultáneo quiere decir en un mismo período de tiempo, grande o pequeño, pero determinado) sobre objetos diversos (los habitantes de cierta edad), mientras que en el tiro de una moneda el conjunto de observaciones corresponde a la reiteración de la observación en el transcurso del tiempo sobre un mismo objeto, c a r e c e d e s i g n i f i c a c i ó n d e s d e e l p u n t o d e v i s t a d e l c á l c u l o d e p r o b a b i l i d a d e s.

Todos los asegurados de una misma edad se consieran como un único jugador que apuesta contra el asegurador una cantidad fija, la prima p del seguro, tantas veces como asegurados hay. El resultado de cada jugada está dado por el hecho de que el asegurado correspondiente sobreviva a la edad en cuestión, o fallezca antes de cumplir la edad inmediata siguiente. En el caso de sobrevivir, gana el asegurador, que percibe el importe de la apuesta (prima del seguro). En caso de fallecimiento, pierde el asegurador, que

(a) Para una historia de los orígenes y desarrollo inicial del cálculo de probabilidades hasta la contribución de Laplace, a fines del siglo XVIII y principios del XIX, ver el libro de TODHUNTER (23).

debe abonar como premio el importe de la póliza, que se denominará monto.

La fijación del importe de la prima, para que el asegurador no pierda y sea posibles el funcionamiento del seguro de vida, se funda en una discusión similar a la de los juegos de azar.

El importe de los pagos que debe efectuar el asegurador es igual al producto del número de fallecimientos r por el monto M de la pólizas, mientras que su recaudación total es igual al producto del número n de asegurados por la prima p.

El pago neto P, será igual a la diferencia de ambos productos

(1.5) $$P = r.M - n.p$$

Cuando el número n de asegurados es bastante grande como para reemplazar r por el producto de n y el límite de la frecuencia relativa $f^\wedge$, se tiene

(1.6) $$P = n(f^\wedge.M - p)$$

Si el importe de la prima p es mayor que el producto del límite de la frecuencia relativa por el monto M de la póliza, el asegurador recaudará una suma superior a la que debe abonar, y el pago neto será negativo, es decir, obtendrá un beneficio. Pero si la prima es inferior a dicho producto, la situación se invierte, y deberá efectuar realmente un pago.

La solución equitativa para ambos parece ser, de acuerdo con esta discusión la igualdad de la prima con el producto del monto por la frecuencia relativa límite, de manera que los asegurados no paguen en exceso, ni el asegurador deba pagar más de lo que recauda. El beneficio del asegurador consistiría en el uso del dinero que tiene en depósito en el lapso transcurrido entre el cobro de la prima y el pago de las pólizas de los asegurados fallecidos.

Naturalmente, en este caso interesa tanto o más que en los juegos de azar el estudio de las desviaciones de las frecuencias observadas con respecto a la frecuencia límite, y saber qué ocurre cuando se trata de un conjunto de montos y de primas de distintos importes, correspondientes a las apuestas variables, que será lo más frecuente en la práctica, problemas a los que también da respuesta el cálculo de probabilidades.

1.5.- La necesidad de estudiar en detalle el comportamiento de los fenómenos aleatorios resulta aparente, para tratar de descubrir nuevas constantes asintóticas del tipo del límite de las frecuencias relativas, lo que permitiría formular otras predicciones para un número grande de observaciones. Pero también es necesario el estudio de las desviaciones, no solo por las razones dadas anteriormente, pués hay casos en que su análisis tiene importancia de por sí y no sólo para precisar dichas predicciones.

Un ejemplo sencillo permite ilustrar una aplicacion importante de este tipo de análisis.

Si en vez de una población de asegurados se trata de una población de enfermos, atacados por un cierto mal y sometidos a un determinado tratamiento, la experiencia indica que la proporción de enfermos que reaccionan de determinada manera (por ejemplo, baja la temperatura después de transcurridas 24 horas de la iniciación del tratamiento), tiende a ser constante al aumentar el número de enfermos tratados.

Cuando se trata de una enfermedad desconocida, sobre la que se experimentan distintos tratamientos, generalmente el número de casos de que se dispone no es tan grande para llegar a determinaciones suficientemente precisas de la proporción de enfermos que reaccionan de determinada manera frente a un tratamiento. Simplemente, se acepta que habrá una frecuencia límite que no se determina, y se compara los resultados obtenidos en dos grupos de pocos enfermos con dos tratamientos distintos. El problema consiste en separar las diferencias entre las frecuencias observadas de acuerdo a que se deban a la posible diferencia de esos límites.

El tratamiento sistemático de problemas de esta clase es una de las más grandes contribuciones recientes del cálculo de probabilidades, que aparece así como herramienta fundamental del método experimental, pues lo que acaba de decirse con respecto a enfermedades puede generalizarse a casi cualquier campo de aplicación.

I.6.- Finalmente, y para no alargar esta introducción, la física de las partículas ofrece un amplio campo de aplicación del cálculo de probabilidades. Fué utilizando sus métodos como Boltzmann pudo fundamentar estadísticamente la termodinámica y obtener notables progresos. También con auxilio del cálculo de probabilidades Planck pudo resolver el problema de hallar la ley que sigue la radiación del cuerpo negro, para no mencionar más que dos jalones históricos.

Los problemas de física son más difíciles de formular desde el punto de vista del cálculo de probabilidades que los que presentan los juegos de azar, debido a que requieren un conocimiento de los fenómenos que en los juegos se reemplaza por la enunciación de las reglas.

Por esta causa, en el curso se hará amplio uso de los ejemplos de juegos, debiendo limitarse los de física a aquellos que puedan formularse sin requerir estudios especiales.

PRIMERA PARTE

AXIOMATICA ELEMENTAL Y TEOREMAS FUNDAMENTALES

II.- AXIOMATICA PARA SISTEMAS FINITOS DE SUCESOS.

II.1.- Como se vió en el capítulo anterior, en la descripción concreta de un fenómeno aleatorio intervienen elementos que carecen de significación desde el punto de vista del cálculo de probabilidades. Basta recordar, para ilustrar suficientemente el punto, que el hecho que las observaciones sean repetidas sobre un mismo objeto o realizadas cada una sobre un objeto diferente (caso del tiro de una moneda y de la mortalidad en una población de edad definida, respectivamente), no establece diferencia alguna en el tratamiento por medio del cálculo de probabilidades.

Para construir un modelo matemático libre de los detalles a que obliga la especialización en un campo determinado, es necesario aislar las nociones primarias que interesan, de todo lo que tienen significación unicamente para la identificación en la experiencia del problema que se estudia.

En el cálculo de probabilidades, la noción primaria es la de s u c e s o : la ocurrencia o la no ocurrencia de una propiedad en una observación, y no tiene nada que ver con la propiedad en sí, ni con el fenómeno particular que se estudia.

La similitud señalada anteriormente entre el tratamiento de las series de observaciones correspondientes a tiros sucesivos de una moneda, y las correspondientes a las observaciones simultáneas de supervivencia o mortalidad de sujetos de una población bien definida, se debe a que se trata de fenómenos aleatorios en cada uno de los cuales se distinguen solamente dos sucesos. Todos los demás elementos de descripción, propios de cada uno de ellos, no intervienen. El tratamiento formal desde el punto de vista del cálculo de probabilidades, y la interpretación de sus resultados, son idénticos.

Cualquier otro fenómeno aleatorio de alternativa simple se tratará de la misma manera, como por ejemplo el sexo en los nacimientos.

La experiencia enseña que la frecuencia relativa de los nacimientos del sexo masculino entre los nacidos vivos (y por lo tanto también la de los nacimientos del sexo femenino) tiende a ser constante cuando el número de observaciones aumenta suficientemente. Este valor límite, o tasa de masculinidad, que se encuentra en la proximidad de 0.51 puede ser utilizado para hacer una estimación de la composición según el sexo, de una población humana futura.

Otro caso importante de fenómeno aleatorio simple es la inspección de materiales. Cuando existe una especificación técnica que permite decidir si un determinado artículo

(por ejemplo una pieza de máquina o un producto terminado) está o no de acuerdo con ella, en ciertas condiciones que serán estudiadas detalladamente más adelante (en la cuarta y última parte, al tratar la inferencia estadística), es posible interpretar la extracción de muestras (número limitado de objetos pertenecientes a una población más numerosa), como un fenómeno aleatorio. Utilizando el cálculo de probabilidades, y conocida la composición de la población total, en cuanto a la proporción de elementos que cumplen con la especificación (y por lo tanto también la de los que no cumplen con la misma), es posible predecir la frecuencia relativa con que apareceran muestras de una composición dada. Partiendo de estos resultados sera posible, por métodos rigurosos, llegar a una inducción o inferencia sobre la composición de la población partiendo de la inspección de un número limitado de objetos. Los enunciados a que se llegará se traducirán en reglas de acción práctica (aceptar o rechazar una partida de acuerdo con la inspección de una muestra), susceptibles de una interpretación precisa en cuanto a sus consecuencias (limitación de la magnitud de los errores de diversos tipos que pueden cometerse por aplicación de esas reglas). Este campo de la inspección de materiales es uno de los más recientes y fructíferos del cálculo de probabilidades.

II.2.- ESPACIO FUNDAMENTAL DE PROBABILIDADES. Si se identifica a un fenómeno aleatorio con un conjunto o espacio, a cada uno de cuyos elementos corresponde una observación posible del mismo, la noción de suceso corresponde a la de subconjunto contenido en ese espacio.

El fenómeno aleatorio del tiro de una moneda estaría representado por un conjunto o espacio, dividido en dos subconjuntos, correspondientes cada uno a la aparición de una de las figuras de la moneda, cuando ésta queda inmóvil después de efectuado el tiro.

Lo mismo ocurre en el caso de la mortalidad. El conjunto que representa todas las observaciones posibles sobre seres humanos de una misma edad, experimenta una partición en dos subconjuntos o sucesos, que corresponden al fallecimiento o a la supervivencia. De idéntica manera se procede en el caso del sexo o de la inspección de materiales.

Los fenómenos aleatorios de alternativa simple, en los que se distinguen solamente dos sucesos, constituyen una clase que no contiene sino parcialmente al conjunto de los fenómenos aleatorios. Es posible que el estudio de un fenómeno aleatorio revele la existencia de más de dos alternativas, lo que requerirá subdivisiones más complicadas del espacio fundamental de probabilidades.

Por ejemplo, en el tiro de un dado, hay seis resultados posibles, lo que implica

una subdivisión del espacio fundamental en seis subconjuntos.

Es posible imaginar fenómenos aleatorios en los que se distinguirían infinitos sucesos, pero en esta primera parte del curso, y en la siguiente, nos limitaremos principalmente al caso de que solamente exista un número finito de sucesos, dejando para la tercera parte el estudio de los fenómenos con infinitas alternativas o sucesos.

II.3.- SUCESO CIERTO. La identificación de un suceso con un subconjunto de un espacio permite una representación sencilla que hace posible establecer, de una manera clara y susceptible de interpretación matemática, todas las nociones auxiliares.

En primer lugar, es posible considerar el mismo espacio fundamental como uno de los conjuntos en que se lo puede subdividir, y que se identifica con el suceso cierto, o sea al que se presenta en todas las observaciones.

II.4.- RELACIONES ENTRE SUCESOS. Dada la identificación entre sucesos y subconjuntos de un espacio, a las relaciones posibles entre sucesos corresponderan relaciones entre subconjuntos de un mismo espacio, y reciprocamente.

Si se denomina tanto a los subconjuntos como a los sucesos con letras mayúsculas, es posible establecer el siguiente paralelismo:

SUBCONJUNTOS	SUCESOS
1) Dado un subconjunto A, queda univocamente determinado otro subconjunto llamado complementario de A, formado por todos los puntos del espacio que no pertenecen a A, y que se conviene en denominar $A^{\wedge}$ Por definición, A y $A^{\wedge}$ no tienen puntos en común.	1) Dado un suceso A, existe otro suceso $A^{\wedge}$, denominado el opuesto de A, que corresponde a la no observación de A. Por definición, A y $A^{\wedge}$ no pueden observarse simultaneamente.
2) Dados dos subconjuntos A y B, se dice que son rampantes, cuando tienen puntos comunes a ambos.	2) Dados dos sucesos A y B, se dice que son compatibles cuando pueden observarse simultáneamente. En el caso del tiro de un dado, los sucesos par y múltiplo de tres son compatibles porque pueden observarse simultaneamente (en el resultado 6).
3) Dados dos subconjuntos A y B, se dice que son aislados, o no rampantes, cuando no tienen puntos en común.	3) Dados dos sucesos, se dice que son excluyentes cuando es imposible observarlos simultaneamente. En el caso del tiro de un dado, los suce-

sos número par y cinco son excluyentes.

II.5.- OPERACIONES CON SUCESOS. Dado un sistema de sucesos en un fenómeno aleatorio, que corresponde en el espacio fundamental de probabilidades a un sistema de subconjuntos, es posible definir nuevos sucesos, mediante operaciones que corresponden a las que son posibles entre subconjuntos.

1) Dados dos subconjuntos A y B, se llama unión de los mismos ($A \cup B$) al conjunto de los elementos del espacio que pertenecen a uno o a otro subconjunto. En el caso de que ambos subconjuntos sean rampantes, habrá elementos que pertenecen simultaneamente a los dos, y otros que pertenecen solamente a uno de ellos. Si los subconjuntos son aislados, los elementos de su unión perteneceran cada uno de ellos solamente a uno de los subconjuntos.

1) Dados dos sucesos A y B, se llama suma lógica ($A+B$), o simplemente suma al suceso que consiste en la observación de uno u otro. Si los sucesos son compatibles, las observaciones del suceso suma pueden corresponder a la aparición simultánea de los dos, o a la de solamente uno de ellos. En el caso del dado, la suma de los sucesos número par y múltiplo de tres es otro suceso, que comprende la observación simultánea de ambos (en el resultado 6), o la individual (en los resultados 3 y 2, por ejemplo). Si los sucesos son excluyentes, el suceso suma está formado por observaciones en las que aparece uno sólo de los sucesos.

2) Dados dos subconjuntos A y B, se llama intersección de los mismos ($A \cap B$) al conjunto formado por los puntos comunes a ambos. En el caso de subconjuntos no rampantes, la intersección es el conjunto nulo o vacío.

2) Dados dos sucesos A y B, se llama producto de los mismos (simbólicamente $A.B$) al suceso formado por la observación simultánea de ellos. Si A y B son excluyentes, el producto $A.B$ es el suceso imposible. En el ejemplo anterior del dado, el producto del suceso par y del suceso múltiplo de 3 es el resultado 6.

II.6.- AXIOMATICA ELEMENTAL. Una vez establecida la noción de sucesos con sus nociones auxiliares, o sea establecido el objeto sobre el cual se desarrolla el cálculo de

probabilidades, es posible formular la noción de probabilidad en forma axiomática.

A cada suceso se le asigna un número (probabilidad), dotado de ciertas propiedades sugeridas por las de las frecuencias relativas. Estas propiedades, son un conjunto mínimo, a partir del cual se pueden deducir logicamente nuevas propiedades, algunas de las cuales coinciden con las que muestran las frecuencias relativas, y otras que no son directamente intuitivas.

El procedimiento es, basicamente, el mismo que se sigue para fundamentar axiomaticamente la geometría. Una vez postulada la existencia de los entes (puntos, rectas y planos) de que se ocupará, las propiedades de los cuerpos en el espacio sugieren un conjunto mínimo de relaciones entre dichos entes, a partir de las cuales se deducen logicamente todas las demás, algunas de las cuales son intuitivas y otras no.

Un conjunto mínimo de propiedades, suficiente para deducir de él todos los resultados fundamentales del cálculo de probabilidades (en el caso de un número finito de sucesos), está dado por los siguientes axiomas:

AXIOMA I : La probabilidad es un número no negativo (mayor o igual a cero).

AXIOMA II : La probabilidad del suceso cierto es igual a la unidad.

AXIOMA III : La probabilidad del suceso suma de dos sucesos excluyentes A y B es la suma de las probabilidades de A y de B :

(2.1) $$P(A+B) = P(A) + P(B).$$

Es de notar que así como en la geometría se postula la existencia de puntos, rectas y planos, sin definir en qué consisten, siendo posible varias representaciones concretas, con los que los teoremas de la geometría se convierten en enunciados sobre el espacio de la intuición, de distinto contenido segun la transcripción que se haga de los entes geométricos, tampoco en el cálculo de probabilidades se definen los sucesos, ni se dice cuanto valen las probabilidades.

Unicamente se postula su existencia y sus propiedades, y su representación concreta sobre un fenómeno aleatorio permite llegar a conclusiones sobre el mismo. El modelo matemático así construído será util si las conclusiones a que permite llegar son adecuadas para la interpretación o análisis del fenómeno, de la misma forma en que el modelo de la geometría de Euclides es útil en el espacio de la intuición, por ejemplo para estudiar las propiedades de la superficie terrestre en una cierta transcripción de los puntos, rectas y planos, a condición que las distancias sean limitadas. Si las distancias son mayores, será necesario cambiar la transcripción.

Cuando se trata de la ubicación de los cuerpos en el espacio, el número de interpretaciones diversas del modelo es, al menos en la geometría elemental muy limitado, en cambio en el cálculo de probabilidades las diversas interpretaciones de sus enunciados son sumamente variadas, lo que hace, que su estudio sea, al comienzo, un tanto des-

concertante hasta llegar a familiarizarse, con el nuevo sistema de ideas.

II.7.- Es fácil ver como las demás propiedades de las frecuencias relativas aparecen por deducción, como consecuencia lógica de la axiomática elemental formulada anteriormente.

Por ejemplo, no fué necesario incluir entre los axiomas que la probabilidad además de ser un número real y no negativo, no puede ser mayor que uno.

Esta última propiedad, evidente en las frecuencias relativas, se demuestra como teorema para las probabilidades.

Dado un suceso A ,conjuntamente con su opuesto $A^{\wedge}$, queda definida una partición del espacio E:

(2.2) $$E = A + A^{\wedge}$$

Como un suceso y su opuesto son excluyentes, por aplicación del AXIOMA III, se tiene:

(2.3) $$P(E) = P(A) + P(A^{\wedge})$$

Por el AXIOMA II , P(E) es igual a la unidad, y como por el AXIOMA I tanto P(A) como $P(A^{\wedge})$ son números no negativos, para que su suma valga la unidad, ambos son menores que ella, o uno de ellos es igual a 1 y el otro a 0.

Como A es un suceso genérico cualquiera, queda demostrado que P(A) no puede ser mayor que la unidad.

II.8.- Un problema interesante lo constituye la determinación de la probabilidad del suceso imposible, al que se denominará con la 0 mayúscula.

Teniendo en cuenta (2.3) y el AXIOMA II

(2.4) $$P(A^{\wedge}) = 1 - P(A)$$

o sea que la probabilidad del suceso opuesto es igual al complemento con respecto a la unidad de la probabilidad del suceso A.

Considerando como suceso a todo el espacio, el suceso opuesto será el suceso imposible, pues el complementario de un conjunto con respecto a si mismo es el conjunto vacío, o suceso imposible, y aplicando (2.4) se tiene:

(2.5) $$P(0) = 1 - P(E) = 1 - 1 = 0$$

o sea que la probabilidad del suceso imposible es nula. Tanto los resultados (2.4) como (2.5) seran de utilización frecuente en lo sucesivo.

II.9.- Como se hizo notar al formular la axiomática elemental, ésta establece solamente relaciones entre probabilidades, pero no indica cuanto valen. Practicamente la única probabilidad que puede calcularse por la sola aplicación de la axiomática es la del suceso imposible. En cada problema habrá que partir de los datos de la experien-

cia para conocer los valores numéricos de ciertas probabilidades, con ayuda de las cuales y con aplicación de la axiomática se calculan aquellas que interesan.

Pero el procedimiento para determinar una probabilidad partiendo de la experiencia es un problema muy delicado.

Una primera idea es utilizar la estabilidad de las frecuencias relativas al crecer el número de las observaciones, pero naturalmente sera necesario hacer un análisis prolijo no sólo de esta estabilidad, sino también de los enunciados que pueden hacerse fundándose en ella, todo lo cual es materia de la i n f e r e n c i a e s t a d í s t i c a que se estudiará en la última parte del curso, por requerir previamente el desarrollo de los elementos del cálculo de probabilidades, en forma deductiva a partir de la axiomática.

II.10.- SISTEMAS COMPLETOS DE SUCESOS EXCLUYENTES. SISTEMAS DE LAPLACE. Históricamente, las primeras probabilidades que se calcularon, y que estaban vinculadas a problemas de juegos de azar, se determinaron en base a un método auxiliar importante. Por ejemplo, en el caso del tiro de una moneda, se distinguen dos sucesos, que son excluyentes entre sí, y que recubren todo el espacio fundamental de probabilidades. En otros términos, forman lo que se convendrá en designar como un s i s t e m a c o m p l e t o d e s u c e s o s e x c l u y e n t e s , y por aplicación directa del AXIOMA III la suma de sus probabilidades será igual a la unidad.

Pero por otra parte, consideraciones de simetría hacen pensar que las probabilidades de cada una de las figuras son iguales entre sí, lo que es confirmado en la experiencia por el cómputo de las frecuencias relativas, que para series largas de tiros dan resultados aproximadamente iguales; por lo tanto, parece razonable asignar a la aparición de cada una de las figuras la probabilidad $\frac{1}{2}$.

La extensión del método a fenómenos aleatorios que no son de alternativa simple, requiere la demostración previa de un TEOREMA DE ADICION: La probabilidad de la suma de n sucesos excluyentes (n finito), es igual a la suma de las probabilidades de los sucesos.

(2.6) $$P(A_1+A_2+\ldots+A_n) = P(A_1)+P(A_2)+\ldots+P(A_n)$$

La demostración se hace por inducción completa. Si es válido para $n-1$, por aplicación directa del AXIOMA III será válido para n :

(2.7) $$P(A_1+\ldots+A_n) = P\left[(A_1+\ldots+A_{n-1})+A_n\right] = P(A_1+\ldots A_{n-1})+P(A_n) = P(A_1)+\ldots+P(A_n)$$

y como evidentemente es valido para $n=3$, es cierto para todo n finito.

Dado un sistema completo de n sucesos excluyentes, si consideraciones de simetría o verificaciones experimentales hacen aceptable la hipótesis de que los sucesos son equiprobables, la aplicación del teorema de adición permite entonces deducir

que la probabilidad de cada uno de los sucesos es igual a $\frac{1}{n}$.

Siguiendo a Richter (21), se convendrá en llamar a un sistema completo de sucesos excluyentes, que además son equiprobables, s i s t e m a d e s u c e s o s d e L a p l a c e , o en forma abreviada s i s t e m a d e L a p l a c e .

En el caso de un dado equilibrado (o simétrico), la hipótesis de que la aparición de cada una de las caras tiene igual probabilidad se verifica experimentalmente por la igualdad aproximada de las frecuencias relativas. En este caso, el sistema de Laplace está formado por seis sucesos excluyentes, y la probabilidad de cada uno de ellos será igual a $\frac{1}{6}$

Inmediatamente se pueden deducir, con ayuda del AXIOMA III, las probabilidades de otros sucesos:

$$P(\text{par}) = P(2+4+6) = 3 . \frac{1}{6} = \frac{1}{2}$$

(2.8)

$$P(\dot{3}) = P(3+6) = 2 . \frac{1}{6} = \frac{1}{3}$$

El problema no queda terminado con esta determinación de las probabilidades. Será necesario comprobar la validez del resultado, utilizando los métodos que proporciona la inferencia estadística, pero hasta tanto se establecen los fundamentos de esta última, se utilizará este procedimiento para construir algunos ejemplos ilustrativos.

II.11.- Es muy frecuente, en las aplicaciones elementales del cálculo de probabilidades, que se presente el caso de poder construir un sistema de Laplace, ya sea por consideraciones de simetría o por observación de las frecuencias relativas.

Pero es tambien muy frecuente que el cómputo efectivo del número de sucesos excluyentes equiprobables se vea complicado por la dificultad de la enumeración de los resultados distintos.

Un problema histórico, en realidad anterior a la constitución del cálculo de probabilidades como rama de las matemáticas, es el siguiente (a) :

Un noble italiano "jugador profesional y matemático aficionado" (segun Fisher) había notado por observaciones prolongadas de un juego consistente en arrojar simultaneamente tres dados, que la suma 10 aparecía más frecuentemente que la suma 9. Este hecho aparecía en contradicción con un análisis realizado, segun el cual era indiferente apostar al 9 o al 10.

El razonamiento en que fundaba la indiferencia de la apuesta, consistía en contar las formas distintas en que se podían descomponer 9 y 10 en tres sumandos enteros:

SUMA 9		SUMA 10	
126	225	136	145
135	234	226	235
144	333	244	334

(a) Galileo Galilei: "Considerazzioni sopra il giuco dei dadi" (reseñado por FISHER (5), pag. 18 y 19).

Para ambos resultados se obtenía el mismo número de descomposiciones, lo que parecía indicar que la frecuencia de aparición sería la misma. Pero no satisfecho con las desviaciones que aparecían en la práctica, el noble italiano sometió el problema a Galileo, quien halló la falacia del razonamiento probando que cada una de las descomposiciones no tenía que presentarse con igual frecuencia que las demás.

Para demostrarlo, Galileo pintó los tres dados, cada uno con diferente color, y encontró que, por ejemplo, la descomposición 126 puede formarse de seis maneras distintas:

126	261
162	612
216	621

donde cada columna corresponde a un dado individualizado por su color.

Cuando en la descomposición hay dos números iguales, dicha descomposición se puede presentar solamente de tres maneras distintas, y cuando los tres números son iguales, hay una sola.

Por lo tanto, la suma 9 aparece de 25 maneras distintas, y la suma 10 de 27.

Galileo no conocía el cálculo de probabilidades, que todavía no se había inventado, y se limitó a computar el número de maneras en que podían hacerse las descomposiciones, considerando como distintas dos descomposiciones no solamente cuando estan formadas por números distintos, sino también cuando siendo iguales los números, estos se encuentran en dados distintos. En otros términos, los dados son distinguibles.

Hoy es posible llegar al mismo resultado sin recurrir a pintar los dados, utilizando el cálculo combinatorio. El número de maneras distintas en que puede presentarse una suma de tres sumandos, considerando como distinta una suma en que los sumandos son iguales cuando estos tienen ubicación distinta, es el de permutaciones de tres elementos. Cuando los tres elementos son distintos, el número de permutaciones es factorial de 3. Cuando dos elementos son iguales, el número de permutaciones es factorial de tres dividido factorial de dos, y cuando los tres elementos son iguales entre si, hay una sola permutación.

En términos de probabilidades, hay 6^3, o sea 216 resultados distintos, lo que resulta de multiplicar los seis posibles resultados de un dado por los seis posibles del segundo y los seis posibles del tercero. Si los dados son simétricos, se puede concluir que los resultados son equiprobables, o sea se tiene un sistema de Laplace, formado por 216 sucesos, y la probabilidad de cada suceso es igual a $\frac{1}{216}$.

Computando el numero de veces en que se presentan suma 9 y suma 10, con ayuda del cálculo combinatorio, y multiplicando dichos números de veces por la probabilidad individual $\frac{1}{216}$, lo que resulta de aplicar el teorema de adición, se tiene finalmente:

(2.9) $$P(9) = \frac{25}{216} \qquad P(10) = \frac{27}{216}$$

El cálculo combinatorio aparece aquí como una herramienta auxiliar, de uso indicado cada vez que es posible llegar a la conclusión que un suceso es suma de otros excluyentes, que, a su vez forman parte de un sistema de Laplace.

Es necesario tener presente, sin embargo, que el cálculo de probabilidades no es un capítulo de aplicaciones de la combinatoria, como podría desprenderse de la manera en que se lo presenta en algunos textos. Si bien hace uso de ella, su problemática es distinta, y en las aplicaciones a que modernamente se lo ha extendido, la importancia de la combinatoria es bastante limitada, en oposición a lo que sucedía cuando su principal aplicación eran los juegos de azar.

II.11.- Algunos ejemplos ayudarán a comprender la aplicación del cálculo combinatorio a problemas de probabilidades, así como ciertos problemas de otro tipo que es necesario tener en cuenta.

Dado un mazo de baraja, con 52 cartas y cuatro colores, de 13 cartas cada uno, se plantea el problema de calcular la probabilidad de extracciones de ciertos grupos de cartas. Por ejemplo ¿Cuál es la probabilidad de que, una vez extraídas tres cartas se encuentre que estas son ases?

La primera cuestión a considerar, es si se trata o no de un fenómeno aleatorio. Es evidente que si no se establecen ciertas reglas, no será posible encontrar frecuencias relativas constantes al repetir la extracción, o bien se encontrará que estas frecuencias relativas varían segun el método de extracción.

Para ponerse en condiciones de tratar la extracción de un grupo de cartas como un fenómeno aleatorio bien definido hay que introducir reglas como la siguiente: una vez extraídas tres cartas y anotados los resultados, hay que volverlas al mazo y barajar cuidadosamente antes de repetir la extracción.

La experiencia indica que si se cumple dicha regla, la frecuencia relativa con que aparece cada grupo de cartas es aproximadamente constante.

Esta cuestión previa de asegurarse si un fenómeno es aleatorio o no, es esencial no sólo en los juegos de azar, sino en casos más importantes. Por ejemplo, en la inspección de materiales, es menester asegurarse previamente que la extracción de muestras es un fenómeno aleatorio, introduciendo ciertas reglas que ya no son tan sencillas de establecer como en el caso de un mazo de cartas, lo que obligará a estudiar el problema más cuidadosamente al tratar la inferencia estadística.

En los libros clásicos (como el de Laplace), y en otros textos más modernos pero que han conservado el estilo de tratamiento, este problema no era considerado como muy importante, pues aparecía salvado al considerar restringido el cálculo de probabilidades a las extracciones de bolillas de urnas o bolilleros, de composición dada y con ciertas reglas. Las aplicaciones de otro tipo se consideraban simplemente como asi-

milables a las extracciones de bolillas. Pero actualmente, en que las aplicaciones se han extendido enormemente, resultaría absurdo presentar el cálculo de probabilidades en esa forma, y es necesario darle la importancia debida al problema de la aleatorización de las observaciones, como se lo designa.

Salvada esta cuestión previa, y establecidas las reglas de reposición de las cartas al mazo y barajado ante de repetir la extracción, el problema consiste en determinar un sistema de Laplace, en el que la aparición de tres ases aparezca como suma de un número definido de los sucesos que lo componen.

En el ejemplo dado, el sistema de Laplace estaría formado por todos los grupos distintos de tres cartas que se pueden extraer, cuyo número es el de combinaciones de de 52 elementos tomados de tres en tres. La probabilidad de cada uno de los grupos será entonces el inverso de ese número:

(2.10) $$P = \frac{1}{\binom{52}{3}}$$

Como hay cuatro ases, el número de grupos de tres ases que se pueden formar será el de combinaciones de cuatro elementos tomados de a tres, y finalmente la probabilidad buscada será:

(2.11) $$P = \frac{\binom{4}{3}}{\binom{52}{3}} = \frac{4!\,(52-3)!\,3!}{52!\,(4-3)!\,3!} = \frac{4!\,49!}{52!} = \frac{1}{5525}$$

Es importante tener en cuenta que el problema debe formularse con toda precisión. No es lo mismo la probabilidad de extraer tres ases (uno cualquiera de los grupos de tres que se pueden formar con cuatro ases), que la de extraer tres ases determinados.

Igualmente se procede para calcular la probabilidad de extraer tres cartas que pertenecen al mismo palo o color, que resulta ser:

(2.12) $$P = \frac{\binom{13}{3}}{\binom{52}{3}} = \frac{11}{850}$$

La probabilidad de que entre las tres cartas no haya dos de un mismo palo será:

(2.13) $$P = \frac{\binom{4}{3}.13^3}{\binom{52}{3}} = \frac{169}{425}$$

teniendo en cuenta que si bien con tres palos se pueden formar 13^3 grupos de tres cartas, cada una de un palo, hay $\binom{4}{3}$ grupos de tres palos que se pueden formar con cuatro. palos

II.12.- Ejemplos importantes de aplicación del cálculo combinatorio a problemas de probabilidades, se encuentran en física.

Dado un sistema de n partículas (por ejemplo, moléculas de un gas), es posible clasificarlas, al menos idealmente, en k grupos o c e l d a s , confor

me a una cierta magnitud. Un caso de clasificación como el indicado se obtiene cuando se utiliza como criterio la velocidad, dividiendo el conjunto de valores de esta última en k intervalos, y asignando una partícula a un intervalo si la velocidad de la que se supone animada se encuentra dentro de ese intervalo. En general, el criterio de clasificación es más complicado, y a menudo interviene en él una multiplicidad de variables.

Una configuración o distribución del sistema de partículas se caracteriza porque r_1 partículas se encuentran en la primera celda, r_2 en la segunda, y así sucesivamente, debiendo cumplirse:

(2.14) $$r_1 + r_2 + \ldots + r_k = n$$

Sin embargo, deberá precisarse con claridad qué se entiende por distribución distinta de otra. Si, por ejemplo, se considera que las partículas son distinguibles, en el sentido de que dos distribuciones que tienen iguales r_j en (2.14) son distintas si dos moléculas se permutan en su ubicación en un par de celdas distintas, el número de distribuciones a considerar será k^n, ya que cada partícula tiene k posibilidades.

Una vez contadas todas las distribuciones, si las hipótesis físicas permiten considerar que son igualmente probables, la probabilidad de cada una será:

(2.15) $$P = \frac{1}{k^n}$$

Este resultado corresponde a la estadística física llamada de Maxwell-Boltzmann, y será considerada con más detalle en el capítulo siguiente.

Si se considera que dos distribuciones o configuraciones son distintas únicamente cuando los sistemas de valores de los r_j en (2.14) no son iguales, pero continúan siendo igualmente probables, el número total y por consiguiente la probabilidad, varían. Esta hipótesis equivale a considerar que las partículas no son distinguibles, en el sentido antes explicado.

Para calcular el número de distribuciones distintas, con esta definición, conviene hacer uso de un artificio simple: se consideran las partículas dispuestas linealmente, estando en sucesión las que se encuentran en la misma celda e introduciendo un elemento de separación entre los grupos de partículas correspondientes a cada celda. Habrá $k-1$ elementos de separación entre los k grupos, y el número de distribuciones distintas será el de combinaciones de $n+k-1$ elementos en grupos de $k-1$:

(2.16) $$\binom{n+k-1}{k-1}$$

La probabilidad de cada distribución será, por lo tanto:

(2.17) $$P = \frac{1}{\binom{n+k-1}{k-1}}$$

Esta estadística es llamada de Bose-Einstein.

FERMI - DIRAC

Tambien se ha encontrado necesario estudiar un tercer tipo de estadística física, en el que se introduce la hipótesis auxiliar de que dos partículas no se pueden encontrar simultáneamente en la misma celda (principio de exclusión), además de continuar considerando que las partículas son indistinguibles. Evidentemente, $n<k$ (n no puede ser mayor que k, por el principio de exclusión, y si fuera igual, como las partículas son indistinguibles, habría una sola distribución).

El número de distribuciones o configuraciones distintas será entonces:

(2.18) $$\binom{k}{n}$$

y la probabilidad correspondiente:

(2.19) $$P = \frac{1}{\binom{k}{n}}$$

Cual de estas estadisticas se aplica a un problema dado, es asunto de la física, pero es interesante señalar que además de las tres estadísticas indicadas, es perfectamente posible estudiar una cuarta, en la que las partículas serían distinguibles (como en la Maxwell-Boltzmann) y valdría el principio de exclusión. Esta última estadística aún no ha encontrado aplicación en la física.

	No exclusión	exclusión
Distinguibles	M - B $p = \frac{1}{k^n}$	?
Indistinguibles	B - E $p = \frac{1}{\binom{n+k-1}{k-1}}$	F - D $p = \frac{1}{\binom{k}{n}}$

No hay en Física algún sistema de partículas donde se pueda aplicar una teoría estadística así ;... por ahora.

27-8-64

III.- T E O R E M A S F U N D A M E N T A L E S

III.1.- En el caso de que dos sucesos no sean excluyentes, el AXIOMA III no es directamente aplicable, y para establecer la forma de operar en ese caso es necesario demostrar un teorema.

El procedimientos de demostración se funda en descomponer el suceso suma en sucesos excluyentes, después de lo cual podrá recién aplicarse el AXIOMA III.

De acuerdo con lo visto en el capítulo anterior, un suceso y su opuesto determinan una partición simple del espacio de probabilidades, o sea que cada uno de los puntos del mismo pertenecen a uno u otro de los subconjuntos representativos, que por definición son excluyentes.

Dados dos sucesos A y B que no son excluyentes, es decir que sus subconjuntos representativos tienen parte común no vacía, es posible hacer de cada uno de ellos una descomposición en suma de sucesos excluyentes, aprovechando la partición simple determinada por el otro suceso y su opuesto:

$$A = AB + AB^{\wedge} \qquad (3.1)$$
$$B = BA + BA^{\wedge}$$

A su vez, el suceso suma de A y B admite también una descomposición en suma de sucesos excluyentes, esta vez de tres y no de dos como en los casos anteriores:

$$A + B = AB^{\wedge} + AB + A^{\wedge}B \qquad (3.2)$$

ya que cada uno de los puntos del espacio pertenece a uno de los sucesos o a su opuesto, y no caben otras combinaciones que las indicadas en (3.2), porque de las cuatro intersecciones no necesariamente vacías que pueden formarse con A, $A^{\wedge}$, B y $B^{\wedge}$. tomados de dos en dos, hay que prescindir de $A^{\wedge}B^{\wedge}$, cuyos puntos no pertenecen ni a A ni a B, y por lo tanto tampoco a A+B.

Aplicando ahora el AXIOMA III a (3.1) y (3.2), se tiene:

$$P(A) = P(AB) + P(AB^{\wedge})$$
$$P(B) = P(AB) + P(BA^{\wedge}) \qquad (3.3)$$
$$P(A + B) = P(AB^{\wedge}) + P(AB) + P(A^{\wedge}B)$$

de donde resulta inmediatamente:

$$P(A + B) = P(A) + P(B) - P(AB) \qquad (3.4)$$

o sea el buscado TEOREMA GENERAL DE ADICION (o de la probabilidad total):

L a p r o b a b i l i d a d d e l s u c e s o s u m a d e d o s e s i g u a l a l a s u m a d e l a s p r o b a b i l i d a d e s , m e n o s l a p r o b a b i l i d a d d e l a r e a l i z a c i ó n c o n j u n t a .

Es evidente que el teorema demostrado contiene como caso particular el AXIOMA III

pues este último corresponde a que la realización conjunta sea el suceso imposible, el que tiene, como se vió en el capítulo anterior, probabilidad nula.

Teniendo en cuenta estos resultados, en la axiomática formulada hubiera sido posible reemplazar el AXIOMA III por el siguiente: La probabilidad de la suma de dos sucesos (excluyentes o no), es menor o igual que la suma de las probabilidades:

(3.5) $$P(A + B) \leq P(A) + P(B)$$

conocida como desigualdad de BOOLE. Una vez demostrado que para que valga el signo de igualdad es necesario substraer P(AB), el AXIOMA III se demuestra como teorema, y la nueva axiómática es equivalente a la anterior.

III.2.- Como primer ejemplo de aplicación se calculará la probabilidad de obtener, en el tiro de un dado equilibrado, un resultado par o múltiplo de tres. Ambos sucesos no son excluyentes, pues se dan simultaneamente en el resultado 6.

Tanto la probabilidad de obtener un resultado par como la de un múltiplo de tres han sido calculadas en el capítulo anterior, encontrando que $P(2) = \frac{1}{2}$ y $P(3) = 1/3$, mientras que $P(\dot{6}) = 1/6$, por lo tanto, aplicando (3.4)

(3.6) $$P(\dot{2} + \dot{3}) = P(\dot{2}) + P(\dot{3}) - P(6) = 1/2 + 1/3 - 1/6 = 2/3$$

que es la probabilidad buscada.

III.3.- PROBABILIDAD CONDICIONAL. En el ejemplo que se acaba de dar de la aplicación del teorema general de adición, existe una gran simplificación, debido a que la probabilidad de la realización simultánea de los sucesos par y múltiplo de tres era conocida con anterioridad. Pero en la mayoría de los casos no ocurre así, y sería necesario disponer de un método para calcular la probabilidad de la ocurrencia simultánea de dos sucesos, partiendo de las probabilidades de cada uno de ellos.

Este problema no tiene solución general pero, en compensación, su consideración permite extender notablemente la esfera de aplicación del cálculo de probabilidades.

En el ejemplo ya tratado del juego con un dado, la probabilidad de obtener un seis es un número que representa la frecuencia relativa límite (en el sentido que se precisará en el capítulo siguiente), en una sucesión ilimitada de tiros. Pero si de dicha sucesión se excluyen todos los resultados impares, y luego se calcula la frecuencia relativa del resultado seis dentro de esta sucesión restringida, se obtiene lo que se conviene en designar como la frecuencia relativa condicional, dado un resultado par, del resultado seis.

Como la experiencia indica que estas nuevas frecuencias relativas gozan de la misma propiedad de aumentar progresivamente su estabilidad al aumentar el número de observaciones, parece razonable completar la axiomática formulada en el capítulo anterior,

introduciendo la noción de p r o b a b i l i d a d c o n d i c i o n a l , cuya definición sería:

(3.7) $$P(B/A) = \frac{P(AB)}{P(A)}$$

Esta definición de probabilidad condicional de B, dado A, que como puede verificarse inmediatamente resulta de establecer una relación entre las probabilidades igual a la que existe entre las frecuencias relativas correspondientes, y es válida sólo cuando $P(A) \neq 0$, permite obtener números que satisfacen la axiomática formulada, por lo que su designación de probabilidades no es arbitraria.

En efecto, tanto P(AB) como P(A) son números no negativos, y su cociente también será no negativo, o sea se cumple el AXIOMA I.

La probabilidad condicional del espacio, P(E/A), es igual al cociente de la probabilidad de la intersección del espacio con A, por la probabilidad de este último. Como la intersección del espacio con A es el mismo A, el cociente es igual a la unidad, con lo que se cumple el AXIOMA II.

Finalmente, dados dos sucesos B y C, excluyentes entre sí, se tiene que:

$$P(B + C/A) = P(B/A) + P(C/A)$$

igualdad que se deduce expresando los sumandos del segundo miembro segun (3.7) y teniendo en cuenta que el numerador de la fracción suma que resulta es igual a P (B+C).A , como consecuencia de la propiedad distributiva de la intersección con respecto a la unión de conjuntos, verificandose por lo tanto el AXIOMA III.

III.4.- TEOREMA DE MULTIPLICACION (o de probabilidad compuesta). Aplicando la definición (3.7) a la probabilidad condicional de A dado B:

(3.8) $$P(A/B) = \frac{P(AB)}{P(B)}$$

supuesto que $P(B) \neq 0$

Combinando (3.7) y (3.8), se tiene el TEOREMA DE MULTIPLICACION

(3.9) $$P(B/A).P(A) = P(A/B).P(B) = P(AB)$$

III.5.- DEPENDENCIA E INDEPENDENCIA ALEATORIA. De (3.9) se deduce que cuando $P(B/A) = P(B)$, también $P(A/B) = P(A)$, lo que se enuncia diciendo que l o s s u c e s o s A y B s o n i n d e p e n d i e n t e s e n t r e s í, expresión perfectamente justificada, pues trasladándose a la frecuencia relativa correspondiente a las probabilidades, se ve que la que corresponde a uno de los sucesos no resulta alterada si para su cálculo se toman en cuenta solamente las observaciones en las que aparece el otro.

En términos tal vez más intuitivos, la posibilidad de la aparición de un suceso (A) que es independiente con (B), no es alterada por la aparición de este último, y recíprocamente.

Nótese que se ha dicho que (A) es independiente c o n (B) y no d e (B), lo que se debe a que el concepto está establecido por (3.9) en forma de relación recíproca, y no de causa a efecto. Si se hubiera empleado la locución "independiente de (B)", podría originarse la interpretación errónea que en el caso de dependencia (A) es causa de (B), o (B) efecto de (A), y recíprocamente.

Cuando hay independencia, (3.9) se transforma en

(3.10) $$P(AB) = P(A).P(B)$$

con lo que se obtiene, p a r a e l c a s o p a r t i c u l a r d e i n d e p e n d e n c i a s o l a m e n t e, la regla cuya conveniencia se señaló al comentar en el párrafo (3.3) los problemas que se podían presentar en la aplicación del teorema general de adición.

Pero en cambio, aparece la posibilidad de una aplicación importante del cálculo de probabilidades a la investigación de las posibles relaciones de dependencia de dos sucesos entre sí. El método en forma esquemática, consiste en calcular P(AB) a partir de P(A) y P(B), en la forma dada por (3.10), y observar después si las frecuencias relativas se apartan significativamente de dicha probabilidad (con los métodos que se indicarán al tratar la inferencia estadística, en la parte cuarta).

III.6.- El teorema de multiplicación permite en muchos casos calcular probabilidades por métodos mas simples de los que resultan de aplicar la combinatoria en la forma explicada en el capítulo anterior.

La idea básica consiste en considerar el espacio fundamental de probabilidades, sobre el que para aplicar el cálculo combinatorio debería buscarse un sistema de Laplace como un producto cartesiano de otros dos espacios de probabilidades, o sea que cada elemento del conjunto producto está formado por pares, cada uno de cuyos integrantes pertenece a uno de los conjuntos, y todos los pares que se pueden formar de esta manera hallan su representación en el conjunto producto (un caso especial de producto cartesiano es el formado por los puntos de un rectángulo en el plano, que es producto de los conjuntos formados por dos de cualesquiera de sus lados que se corten).

Una vez descompuesto el espacio fundamental de probabilidades en el producto cartesiano de otros dos, sobre uno de los cuales se efectúa la partición correspondiente al suceso A, con la asignación de sus probabilidades (o sea, se establece la distribución de probabilidades inducida por A), y sobre el otro la correspondiente al suceso B, o sea se lo descompone en dos subconjuntos representativos de B y B' respectivamente, asignandoles las probabilidades condicionales de A, a cada uno de los subconjuntos del espacio producto correspondiente a una cierta combinación de A con B o con B', corresponde al producto cartesiano de subconjuntos de los espacios factores, y la probabilidad es igual al producto de las probabilidades. Si interesara la combinación de A' con B o con B',

el procedimiento es el mismo, sólo que hay que tomar entonces la distribución de probabilidades en B y B', c o n d i c i o n a l de A'.

Sea, por ejemplo, calcular la probabilidad de que en un mazo de 52 cartas (dividido en cuatro palos o colores) se extraigan simultáneamente dos y sean del mismo palo (supuesto establecidas las correspondientes reglas de devolución al mazo de los pares extraídos y mezcla de éste antes de una nueva extracción).

Para aplicar el cálculo combinatorio, se determina el sistema de Laplace correspondiente, calculando el número de grupos de dos cartas que pueden formarse con 52, y luego se multiplicaría la inversa de ese número (probabilidad de cada uno de los pares) por el número de grupos de dos cartas que pueden formarse con trece (que es el número de cartas que posee cada palo o color), lo que equivale a aplicar el teorema de adición demostrado en (2.7), obteniéndose:

$$P = \frac{\binom{13}{2}}{\binom{52}{2}} = \frac{13!\cdot 50!\ 2!}{2!\ 11!\ 52!} = \frac{12.13}{51.52} = \frac{1}{17} \tag{3.11}$$

Utilizando la noción de espacio producto, los cálculos se simplifican considerablemente. El espacio fundamental de probabilidades estaría dado por el producto cartesiano de otros dos espacios, sobre el primero de los cuales se distinguirían dos sucesos (pertenecer o no al palo en cuestión), con probabilidades 1/4 y 3/4, y sobre el segundo se tendría tambien dos sucesos (pertenecer o no al palo dado, con probabilidades condicionales 12/51 y 39/51 respectivamente).

El suceso cuya probabilidad se trata de calcular está formado por el producto cartesiano de dos sucesos cuyas probabilidades son 1/4 y 12/51, y aplicando (3.9), se tiene:

$$P = \frac{1}{4}\cdot\frac{12}{51} = \frac{1}{17} \tag{3.12}$$

III.7.- EXTENSION DEL TEOREMA DE LA PROBABILIDAD TOTAL. Teniendo en cuenta que

$$A_1 + A_2 + \dots + A_{n-1} + A_n = (A_1 + \dots + A_{n-1}) + A_n \tag{3.13}$$

o sea la propiedad asociativa de la unión de conjuntos, el teorema general de adición, o de la probabilidad total, se extiende facilmente al caso de un número arbitrario n (finito) de sucesos

En el caso de n=3 , aplicando (3.13) y (3.4), se obtiene:

$$P(A_1 + A_2 + A_3) = P(A_1 + A_2) + P(A_3) - P\left[A_3(A_1 + A_2)\right] \tag{3.14}$$

Como $A_3(A_1 + A_2) = A_3A_1 + A_3A_2$, aplicando nuevamente (3.4), al primer y último sumando del segundo miembro de (3.14):

$$P(A_1 + A_2 + A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_2A_3) - P(A_1A_3) + P(A_1A_2A_3) \tag{3.15}$$

En el caso general se tiene:

$$(3.16)\quad P(A_1+A_2+\dots+A_n)=P(A_1)+P(A_2)+P(A_n)-P(A_1A_2)-\dots-P(A_{n-1}A_n)+$$
$$+P(A_1A_2A_3)+\dots+P(A_{n-2}A_{n-1}A_n)+\dots+(-1)^n.P(A_1A_2\dots A_n)$$

III.8.- GENERALIZACION DEL TEOREMA DEL PRODUCTO. Procediendo por inducción, y limitandonos por sencillez al caso de independencia de todos los sucesos, entre si, se tiene a partir de (3.10):

$$(3.17)\qquad P(A_1A_2\dots A_n)=P(A_1).P(A_2).P(A_3)\dots P(A_n)$$

En efecto, aplicando la propiedad asociativa de la interseccion de conjuntos, si el teorema es válido para $n-1$ es válido para n:

$$(3.18)\quad P(A_1A_2\dots A_n)=P[(A_1A_2\dots A_{n-1})A_n]=P(A_1A_2\dots A_{n-1})P(A_n)=P(A_1)P(A_2)\dots P(A_n)$$

y se ha demostrado su validez para $n=2$, lo que se emplea precisamente en (3.18).

III.9.- Estos teoremas de adición y de multiplicación fueron aislados y enunciados por primera vez por Laplace. Los autores anteriores a él ciertamente los emplearon, pero siguiendo la costumbre de la época, se ocupaban de resolver problemas particulares, sin dedicar mayor atención a analizar los métodos empleados, para deducir reglas simples de aplicación general.

En su aplicación, a fin de evitar errores que conducen a lo que en un tiempo fueron consideradas paradojas del cálculo de probabilidades, es necesario tener en cuenta, que el teorema de adición se refiere a probabilidades definidas sobre un mismo espacio fundamental, mientras que el del producto se aplica en realidad al espacio producto.

Sea por ejemplo, el caso de una operación de bombardeo realizada sobre dos blancos distintos. La experiencia indica que debido a las defensas antiaéreas existentes, la frecuencia con que un bombardero es derribado sobre uno de los blancos es 0,7 , y 0,4 sobre el otro. Se pregunta: ¿Cual es la probabilidad de ser derribado en una operación de bombardeo sobre los dos blancos?

Una aplicación precipitada del teorema de adición conduciría a la suma de ambas probabilidades (identificadas en su valor numérico con las frecuencias relativas dadas) , con lo que se obtendría la probabilidad 1,1 , mayor que la unidad y evidentemente absurda.

Aunque los sucesos son excluyentes (si un aparato es derribado sobre un blanco no puede serlo sobre el otro), lo que parece justificar la suma, las probabilidades no estan definidas sobre el mismo espacio, y por lo tanto no son sumables.

El espacio fundamental correspondiente a la operación de bombardeo sobre los dos blancos, está constituido por el producto cartesiano de los espacios correspondientes al bombardeo sobre cada uno. Una vez definidos los conjuntos correspondientes a ser derri

bado o no sobre el primer blanco, con probabilidades 0,7 y 0,3 , en el primer espacio, en el segundo espacio hay que considerar las probabilidades condicionales correspondientes. Para el caso de ser derribado sobre el primer blanco, la probabilidad condicional de ser derribado sobre el segundo en la misma operación de bombardeo es evidentemente nula. Pero en el caso de no ser derribado sobre el primer blanco, la probabilidad condicional de serlo sobre el segundo es evidentemente 0,4 . En el espacio producto, la probabilidad buscada es, por lo tanto:

(3.19) $$P = 0,7 + 0,4 . 0,3 = 0,82$$

Como en el espacio producto, el subconjunto representativo de ser derribado sobre uno u otro de los blancos tiene por complementario del de no ser derribado sobre ninguna, cuya probabilidad es $0,3 . 0,6 = 0,18$, sacar otra manera de calcular la probabilidad buscada es simplemente restar de la unidad la probabilidad de la realización simultánea de los sucesos opuestos:

(3.20) $$P = 1 - 0,18 = 0,82$$

Este segundo método es tambien aplicable con frecuencia en lugar de teorema de adición, cuando por el número de sumandos se complica su uso. Por ejemplo ¿Cual es la probabilidad de que aparezca al menos un 6 en cinco tiros de un dado equilibrado?

La probabilidad de que no aparezca ninguno en un tiro es 5/6 . Como los resultados en tiros sucesivos son independientes, la probabilidad de que no aparezca ninguno en cinco tiros es $(5/6)^5$, y su opuesto es precisamente el sucesc de que aparezca al menos un 6, cuya probabilidad será entonces:

(3.21) $$P = 1 - (5/6)^5$$

La sola inspeccion de la fórmula (3.16), en comparación con (3.21), basta para comprobar la utilidad de esta forma de aplicación del teorema del producto en substitución del de adición (a).

III.10.- APLICACION DE LA NOCION DE SUCESOS INDEPENDIENTES. En el siglo XIX, Mendel investigó las leyes de la herencia, observando el fenómeno en plantas. Los resultados que encontró fueron sorprendentes, al menos desde el punto de vista de las ideas dominantes en su tiempo.

Si se cruzaban dos plantas en que las flores eran, por ejemplo, rojas y blancas, de acuerdo a la opinión comun entonces, debería obtenerse plantas con flores rosadas, en las que ambos caracteres aparecerían mezclados. Pero Mendel encontró casos en que no sucedía así, obteniendo plantas en que las flores eran de un color de los dos. Uno de los caracteres se heredaba, y el otro simplemente desaparecía, en contra de una idea simplista de

(a) Para una exposición sumaria de los principales resultados y aplicaciones del cálculo de probabilidades, no es necesario proseguir con el estudio de estos teoremas. Para un tratamiento amplio de los mismos y algunas aplicaciones, ver FRECHET (6)

transmisión por herencia, que lleva a la idea de mezcla de los mismos.

Pero las particularidades observadas por MENDEL no terminan con este descubrimiento de la existencia de caracteres d o m i n a n t e s (que en la herencia aparentemente anulan a otros). En una segunda generación, obtenida cruzando entre sí plantas de la primera, en las que aparecía solamente el caracter dominante, volvían a observarse plantas en las que el color de las flores era el dominado o r e c e s i v o y el fenómeno se producía con frecuencias relativas muy aproximadamente constante, correspondiendo 3/4 a las plantas que mostraban el caracter dominate, y 1/4 al recesivo.

MENDEL se limitó a enunciar estos resultados en forma de leyes relativas a la existencia de la dominancia y a las frecuencia relativas de aparición en distintas combinaciones, y su trabajo fué practicamente ignorado.

A fines del siglo XIX y principios del XX otros investigadores reencontraron las leyes de MENDEL, y pronto se dió una explicación relativamente satisfactoria de este curioso fenómeno.

Se observó que las células de las plantas presentaban unos corpúsculos alargados, llamados c r o m o s o m a s por su facilidad en tomar ciertos colorantes empleados en la observación por medio del microscopio, y en algunos casos hasta llegó a aparecer como que esos cromosomas tuvieran la estructura de una doble fila de corpúsculos.

En la reproducción, las células se dividían en dos partes o gametos, y en esa división los cromosomas se dividían también, en forma longitudinal y quedando células de distintas plantas, volviendo a recomponerse los cromosomas.

Si se considera que los cromosomas contienen corpúsculos, o g e n e s , portadores de los caracteres hereditarios, que tienen la propiedad de que cuando se combinan dos gametos, en los cuales hay un gene dominante en uno y un recesivo en el otro , solo aparece en el nuevo ser el caracter dominante, y el recesivo permanece latente,es decir, no se muestra, pero en una generación posterior en la que se combinen dos gametos con recesivos vuelve a aparecer, las leyes de MENDEL pueden explicarse suponiendo, que en la reproducción los gametos portadores de uno u otro tipo de gene se combinan al azar, con i n d e p e n d e n c i a .

En la cruza de dos plantas, una d o m i n a n t e p u r a (cromosomas con los dos genes dominantes), y otra r e c e s i v a (cromosomas con los dos genes recesivos), se combina con un gameto de una planta con un gameto de otra, y en la nueva planta solo puede aparecer el caracter dominante. Pero en la segunda generación los gametos pueden ser portadores de genes dominantes o recesivos, con una probabilida igual a 1/2 para cada uno. La probabilidad de que se combinen,dos recesivos es el producto de sus probabilidades, o sea 1/4, de acuerdo con la observación de Mendel.

III.11.- Para continuar con la aplicación de la noción de sucesos independientes a la

genética, es necesario establecer cierta nomenclatura. El tipo aparente, identificado por el caracter que muestra, se llama f e n o t i p o , y en el caso de plantas con flores de dos colores, hay dos fenotipos. Pero si se atiende a la composión genética (composición del par de genes), existen tres t i p o s genéticos o g e n o t i - p o s que son: el d o m i n a n t e p u r o (ya explicado), el h í b r i d o (con un gene dominante y otro recesivo) y el r e c e s i v o (con los dos genes recesivos).

El fenotipo dominante está formado por los genotipos dominante e híbrido y su probabilidad a partir de una generación de híbridos es 3/4, lo que se demuestra facilmente sumando las distintas probabilidades de los subconjuntos del espacio producto.

Es necesario distinguir las probabilidades de los genes, que se llamarán p y q , de las de los genotipos, que se llamaran u(para los dominantes), 2v (para los híbridos), y w (para los recesivos).

Entre estas probabilidades existen las siguientes relaciones:

(3.22) $$p = u + v \quad ; \quad q = v + w$$

que se demuestran calculando u , 2v y w a partir de p y q (aplicando nuevamente un espacio producto, en cada uno de cuyos factores se distinguen dos sucesos con probabilidades p y q):

(3.23) $$u = p^2 \quad ; \quad 2v = 2pq \quad ; \quad w = q^2$$

con lo que se tiene:

(3.24) $$u + v = p^2 + pq = p(p + q) = p$$

procediendose igualmente para las otras dos relaciones.

III.12.- LEY DE HARDY. Cuando una población con una distribución de probabilidades p y q para un par de genes (dominante y recesivo), se reproduce entre sí, la composición genética permanece constante en la generación siguiente.

Llamando p_1 a la probabilidad del gene dominante en la generación, y q_1 a la del gene recesivo, se tiene evidentemente:

(3.25) $$u_1 = p^2 \quad ; \quad 2v_1 = 2pq \quad ; \quad v_1 = q^2$$

y aplicando (3.22):

(3.26) $$p_1 = u_1 + v_1 = p^2 + pq = p \quad ; \quad q_1 = q^2 + qp = q$$

III.13.- SELECCION NATURAL. Estas leyes de la herencia parecen contradictorias con una teoría de la evolución, pues de acuerdo a la ley de Hardy la composición genética de la población permanece constante, y los caracteres serían permanentes. Pero si ocurre que los individuos recesivos tienen alguna dificultad para sobrevivir, afectandose la posibilidad de su reproducción, es evidente que se producirá una rarefacción progresiva de los genes recesivos en la población, y el caracter correspondiente tenderá

a desaparecer.

Una hipótesis extrema corresponde a la imposibilidad de reproducción de los individuos recesivos, y en este caso es relativamente simple calcular la velocidad de empobrecimiento de la población en fenotipos recesivos, en función del número de generaciones.

En la generación (n+1)-esima, las probabilidades de los genotipos en la población padre (o apta para reproducirse) serán:

(3.27) $$\hat{u}_{n+1} = \frac{u_{n+1}}{1 - w_{n+1}} \quad ; \quad \hat{v}_{n+1} = \frac{v_{n+1}}{1 - w_{n+1}}$$

o sea son las probabilidades condicionales de los genotipos dominante e híbrido, dado el suceso "posibilidad de reproducirse" (de probabilidad $1 - W$), y aplicando (3.25):

(3.28) $$p_{n+1} = \frac{u_{n+1} + v_{n+1}}{1 - w_{n+1}} = \frac{p_n^2 + p_n \cdot q_n}{1 - q_n^2} = \frac{p_n}{1 - q_n^2}$$

Descomponiendo la diferencia de cuadrados del denominador del último término, en el producto de la suma por la diferencia de las bases, y teniendo en cuenta que $p_n = 1 - q_n$, se tiene finalmente:

(3.29) $$p_{n+1} = \frac{1}{1 + q_n}$$

Procediendo en igual forma para q_{n+1}, se llega a:

(3.29) $$q_{n+1} = \frac{q_n}{1 + q_n}$$

de donde resulta:

(3.30) $$\frac{1}{q_{n+1}} = \frac{1}{q_n} + 1$$

Como (3.30) es válida para cualquier n se la puede aplicar para expresar $1/q_n$ prosiguiendo reiteradamente hasta llegar a $1/q$, con lo que se tendrá:

(3.31) $$\frac{1}{q_{n+1}} = n + \frac{1}{q}$$

y por lo tanto:

(3.32) $$q_{n+1} = \frac{q}{1 + n.q} \quad \therefore \quad w_{n+1} = \left(\frac{q}{1 + n.q}\right)^2$$

Si, por ejemplo, $q = 1/2$ (la población original es de híbridos), al cabo de diez generaciones en las que los fenotipos recesivos no se reproducen, la probabilidad de que aparezca un recesivo en la generación siguiente será 1/144, con lo que queda demostrada la rarefacción de los recesivos en este caso especial de selección natural.

Este resultado es aplicado por los genetistas, simplemente impidiendo la reproducción de los recesivos, y de esta manera se eliminan los genes recesivos de una población de híbridos, obteniéndose así una variedad fija con algun caracter deseable, o

sea, los llamados "puros por cruza". El procedimiento se llama "endocría" (en inglés, "inbreeding "), por que depende de que la reproducción se realice siempre dentro de la misma población (a).

III.14.- APLICACION DE LA NOCION DE DEPENDENCIA. Si se desea calcular el número de textos distintos que se pueden escribir a máquina en un hoja de papel con capacidad para 25 renglones de sesenta espacios cada uno, dicho número se obtiene teniendo en cuenta que en cada espacio se puede escribir una letra, o bien hay un espacio en blanco de separación entre palabras. El número total de signos es 28 (alfabeto con ñ, w y mas el espacio en blanco, sin tener en cuenta signos de puntuación, paréntesis, etc., los que pueden escribirse con palabras). En dos espacios hay 28 por 28 combinaciones diferentes, y en el total de los 1.500 espacios de la hoja hay $28^{1.500}$ combinaciones.

Este número depende de considerar como distinto un texto cuando está formado por signos distintos, con independencia de su significado. Con esta definición de texto distinto, el número $28^{1.500}$ mide la c a n t i d a d d e i n f o r m a c i ó n que puede tener una hoja de papel.

Si se considera que el texto tiene que estar escrito en un idioma determinado, el número de textos distintos será mucho menor, pues no todas las distintas combinaciones que pueden formarse tendran un sentido en ese idioma. La cantidad de textos distintos correspondería en este caso a la que teoricamente podría formarse con 28 signos y un número menor de espacios (dado por el primer exponente entero de 28 con el cual se obtiene un número superior al de textos en ese idioma), lo que indica que la capacidad de transmitir información que tiene la hoja de papel se aprovecha mal, debido a la estructura del idioma escrito, que hace necesario ocupar más espacio del teoricamente necesario.

En el caso de la hoja de papel puede no ser muy importante esta observación, que sería mas bien una simple curiosidad. Pero si en vez de una hoja de papel es un teletipo que trasmite 1.500 caracteres en un cierto intervalo de tiempo, se ve inmediatamente que hay una importancia económica en acortar el tiempo que se necesita para trasmitir un cierto texto o lo que es lo mismo, en aumentar el número de textos que pueden trasmitirse en un mismo intervalo de tiempo.

La solución del problema requeriría efectivamente calcular el número mínimo de signos necesario para transmitir todos los mensajes que en escritura ordinaria requieren 1.500 espacios, y encontrar un procedimiento de codificación, que hiciera corresponder a cada mensaje una de las combinaciones distintas que pueden formarse con ese número

(a) En el libro de FELLER (4) se desarrollan otros ejemplos de genética. Para una exposición mas amplia, ver, por ejemplo, R. A. FISHER:"The Genetical Theory of Natural Selection", New York (reimpresion) Dover

menor de signos, acortando al máximo la longitud del texto.

Así planteado, el problema es comprensible, pero la solución dista de ser inmediata. Sin embargo, el matemático norteamericano CLAUDE SHANNON logró encontrarla, haciendo uso de la noción de sucesos dependientes.

En efecto, la primera determinación del número de textos distintos sigue un procedimiento idéntico que el de calcular la probabilidad de un grupo de 1.500 signos, cuando existen 28 signos equiprobables (la probabilidad de cada grupo sería el inverso del número total de textos). Pero en un idioma la equiprobabilidad de los signos no es cierta. La probabilidad de la aparicion de un signo, depende del anterior. Por ejemplo, si se tiene una m, en castellano podrá ser seguido por una vocal, por las consonantes b ó p, o tal vez por un espacio en blanco, correspondiendo una probabilidad nula a los otros signos. La disminución del número de textos distintos se debe a que en lugar de 28 habrá que considerar solamente 8 signos a continuación de una m. Aplicando los métodos especiales para tratar cadenas de sucesos dependientes (capítulo 9) será posible fundar la TEORIA DE LA INFORMACION, como se llama a esta reciente rama de las matemáticas (a).

III.15.- APLICACIONES A LA FISICA. El uso combinado de las nociones de independencia y dependencia estadística, enunciadas y discutidas en el párrafo II.12 del capítulo anterior, permite precisar el significado de las probabilidades de sistemas de partículas calculadas en II.12 como ejemplo de aplicación del cálculo combinatorio, haciendo posible la demostración de teoremas que resuelven problemas físicos importantes.

Como ejemplo, se tratará con mayor detalle algunas cuestiones relativas a la estadística de Maxwell-Boltzmann, hasta llegar a justificar, sin mayores problemas matemáticos, la ley de la radiación del cuerpo negro, descubierta por Planck en 1901.

Suponiendo, como en II.12 , un sistema formado por gran número de partículas de la misma naturaleza física, susceptibles de ser clasificadas en grupos segun una cierta magnitud, como podría ser la energía, se tendrá:

(3.33)
$$r_1 + r_2 + \dots + r_k = n$$
$$e_1 r_1 + e_2 r_2 + \dots + e_k r_k = u$$

expresiones en las que n es el número total de partículas, u la energía total del sistema, y r_j el número de partículas que poseen la energía e_j (si se admite una variación continua de la energía, lo que constituiría una hipótesis adicional, e_j sería una e n e r g í a p r o m e d i o entre dos valores que determinan un intervalo).

Las hipótesis básicas son: (1) Existe una probabilidad p_j bien definida para que una partícula se encuentre en el grupo j; (2) Las partículas no ejercen ningun tipo de acción entre sí, lo que en particular conduce a afirmar que la probabilidad p_j de una

(a) Para una exposición de divulgación, ver GILBAUD: "La Cybernetique" Presses Universitaires de France.

partícula es independiente de que existan o no otras partículas en el mismo grupo.

Considerando que las partículas son distinguibles, es decir, que dos sistemas de partículas que tienen la misma distribución en grupos caracterizada por la primera relación de (3.33) son distintos cuando uno o varios pares de moléculas individualizadas se encuentran permutadas en su posición en dos grupos distintos, a cada configuración o distribución de las partículas corresponderá una probabilidad.

(3.34) $$p = p_1^{r_1} . p_2^{r_2} \ldots p_k^{r_k}$$

La probabilidad (2.15) calculada en el párrafo II.12 corresponde a la hipótesis simplificada de que todas las p_j son iguales entre sí, o sea a $\frac{1}{k}$, con los que los exponentes r_j se suman, obteniendose finalmente la misma expresión hallada mediante el uso de la combinatoria. La probabilidad (3.34) ha sido calculada partiendo de los estados m i c r o s c ó p i c o s, o estados de cada partícula. Pero hay que calcular la probabilidad de los estados m a c r o s c ó p i c o s del sistema, caracterizados por que en cada nivel de energía e_j se encuentran r_j partículas.

Para esta última finalidad, hay que sumar todas las expresiones (3.34) correspondientes a las configuraciones del sistema en las que existen r_j partículas en cada nivel de energía. Como todas las probabilidades (3.34) son iguales entre si, en este caso, basta multiplicar su valor por el número de combinaciones de n elementos, de los cuales r_1 se encuentran en el primer grupo, r_2 en el segundo, etc., obteniéndose:

(3.35) $$P = \frac{n!}{r_1!\ r_2!\ \ldots\ r_k!}\ p_1^{r_1} . p_2^{r_2} \ldots p_k^{r_k}$$

que es la probabilidad buscada del estado macroscópico.

III.16.- Cuando el sistema de partículas posee una energía constante, hay estados macroscópicos que no son posibles, porque el valor de la segunda expresión de (3.33) no sería igual a esa energía constante.

En este caso las probabilidades (3.35) deben modificarse, porque el espacio de probabilidades se reduce a aquellos espacios compatibles con la condición fijada, y l a p r o b a b i l i d a d d e l e s t a d o m a c r o s c ó p i c o t e n d r á e n t o n c e s l a e s t r u c t u r a d e u n a p r o b a b i l i d a d c o n d i c i o n a l, como se la definió mediante la fórmula (3.7) del párrafo II. en donde A es el suceso "poseer la energía dada" y AB "estar en el estado macróscopico B y poseer la energía dada". Como el suceso B está contenido en el A (cada estado macroscópico tiene una energía total bien definida, y no puede observarse si la energía total es otra), P(AB) será igual a P(A), o sea a (3.35), y P(A) será igual a la suma de las probabilidades de todos los estados macroscópicos que tienen la misma energía. Indicando a esa suma con 1/C se tendrá entonces que la probabilidad del estado macros

cópico, dada una cierta energía total U, será entonces, conforme a la definición de probabilidad condicional:

(3.36) $$P = \frac{C_o n!}{r_1!\, r_2! \ldots r_k!} \cdot p_1^{r_1} \cdot p_2^{r_2} \ldots p_k^{r_k}$$

III.17.- DISTRIBUCION CANONICA. De todos los estados macrocópicos que poseen una cierta energía total U dada, interesa caracterizar (o sea determinar efectivamente la distribución de las partículas, mediante el conocimiento de los r_j) aquel cuya probabilidad (3.36) es máxima. La configuración correspondiente se denomina **distribucion canónica**.

Llamando b_j a los valores de los r_j de la hipotética distribución canonica y P_b a su probabilidad, el cociente de la probabilidad P_r de un estado macroscópico genérico, que cumple con la condición de poseer la misma energía total, por la probabilidad de la distribución canónica, será:

(3.37) $$\frac{P_r}{P_b} = \left(\frac{b_1!}{r_1!}\right) \cdot \left(\frac{b_2!}{r_2!}\right) \cdots \left(\frac{b_k!}{r_k!}\right) p_1^{r_1-b_1} \cdot p_2^{r_2-b_2} \ldots p_3^{r_k-b_k}$$

expresión en la que la constante C se ha eliminado, por figurar en el numerador y en el denominador.

Es posible determinar la forma que tienen los b_j, partiendo de la desigualdad:

(3.38) $$\frac{m!}{n!} \leqslant m^{m-n}$$

En efecto, si $m > n$, se tiene

(3.39) $$\frac{m!}{n!} = (n+1)\,(n+2) \ldots m < m^{m-n}$$

Si $m < n$:

(3.39) $$\frac{m!}{n!} = \frac{1}{(m+1)\,(m+2) \ldots n} < \frac{1}{m^{n-m}} = m^{m-n}$$

y si $m=n$ ó $m=n+1$, vale la igualdad.

Con esta desigualdad el cociente (3.38) se acota superiormente:

(3.40) $$\frac{P_r}{P_b} \leqslant \left(\frac{b_1}{P_1}\right)^{b_1-r_1} \cdot \left(\frac{b_2}{P_2}\right)^{b_2-r_2} \cdots \left(\frac{b_k}{P_k}\right)^{b_k-r_k}$$

Pasando a los logaritmos, se tiene

(3.42) $$\log \frac{P_r}{P_b} = \sum (b_j - r_j) \log \frac{b_j}{P_k}$$

Si los

$$b_j = n\, P_j\, e^{y_o\, y_1 e_j}$$

el segundo termino de la desigualdad se transforma en:

(3.43) $$\sum (b_j - r_j) \log \frac{b_j}{P_j} = (\log n + y_o) \sum (b_j - r_j) + y_1 \sum e_j (b_j - r_j)$$

y aplicando las relaciones (3.33), resulta siempre nulo, porque por hipótesis b_j y r_j las satisfacen.

Como el logaritmo de la unidad es cero, si los b_j tienen la expresión arriba indicada, el cociente (3.37) tendrá un valor menor que la unidad, cualquiera que sean los r_j (a menos que coincidan con los b_j), y por lo tanto la distribución canónica está caracterizada por la forma dada de los b_j.

III.18.- LEY DE PLANCK DE LA RADIACION DEL CUERPO NEGRO. Si se considera que el cuerpo negro está formado por un conjunto de partículas emisoras de radiación de frecuencia ν clasificables en niveles de energía e_j proporcionales a la radiación y a los números enteros, comenzando por el cero (hipótesis de Planck), se tiene que los b_j de la distribución canónica, admitiendo que todas las probabilidades p_j son iguales entre si (o sea iguales a $1/k$), se expresaran:

(3.44) $$b_j = \frac{n}{k} e^{y_o + y_1 h\nu (j-1)} = A\, z^{j-1}$$

en donde $A = \frac{n}{k} e^{y_o}$ y $z = e^{y_1 h\nu}$ y h es la constante de Planck.

Las expresiones (3.33) se transformaran en este caso, tomando la forma:

(3.45) $$A(1 + z + z^2 + --- + z^{k-1}) = A\,\frac{z^k - 1}{z - 1}$$

$$A\left[z + 2z^2 + --- + (k-1)\, z^{k-1}\right] = Az.h\nu \frac{d}{dz}\,\frac{z^k - 1}{z-1} = U$$

en las que el paréntesis de la primera se ha sumado segun la fórmula conocida de las progresiones geométricas de razón z y primer término unitario, y al segundo paréntesis se le ha aplicado el sencillo artificio de considerarlo como la derivada de una progresión geométrica de igual razon que la anterior, si previamente se saca factor común z .

El cociente de la segunda expresión de (3.45) por la primera, o sea la energía media, será:

(3.46) $$\frac{U}{nh\nu} = z.\frac{z-1}{z^k - 1} \cdot \frac{d}{dz}\,\frac{z^k - 1}{z-1}$$

Teniendo en cuenta que es posible demostrar que $y_1 = -1/KT$ en donde K es la constante de los gases de Boltzmann, y T la temperatura absoluta (ambos números son positivos), z será un número menor que la unidad, con lo que admitiendo que k es un número lo suficientemente grande como para que se pueda prescindir de la potencia z^k (que tiende al limite cero), el cociente (3.46) se simplifica considerablemente:

(3.47) $$\frac{U}{nh\nu} = \frac{z(1-z)}{(1-z)^2}\;\frac{z}{1-z}$$

de donde se deduce:

(3.48) $$U = nh\nu\,\frac{e^{y_1 h\nu}}{1 - e^{y_1 h\nu}} = nh\nu\,\frac{e^{-\frac{h}{KT}}}{1 - e^{-\frac{h}{KT}}}$$

que es la famosa ley de Planck para la radiación del cuerpo negro, que establece la relación de dependencia entre la energía, la frecuencia y la temperatura, primera aplicación de la hipótesis de la distribución discontinua de la energía (o hipótesis cuan

tica).

Como comentario, cabe observar que la dificultad del problema reside precisamente en la formulación de las hipótesis físicas. La aplicación del cálculo de probabilidades es de tipo completamente elemental, y sólo requiere cierta habilidad algebraica para simplificar las expresiones que se obtienen (a).

III.19.- TEOREMA DE BAYES. Como consecuencia inmediata del teorema del producto, enunciado en su caso más simple mediante la expresión (3.9), la probabilidad condicional de A dado B resulta igual a:

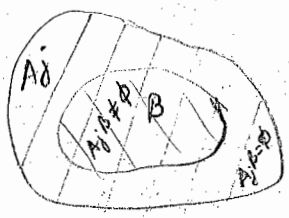
$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} \tag{3.49}$$

que es el caso más simple del teorema de Bayes.

$\sum_{j=1}^{n} P(A_j) = 1$

Si el suceso A forma parte de un sistema completo de sucesos excluyentes A_j, la probabilidad de B que figura en el denominador de (3.8) se puede expresar como la suma de la probabilidades de la realización conjunta de B con cada uno de los A_j (algunas de estas probabilidades pueden ser nulas), y a cada uno de los sumandos se le puede aplicar nuevamente el teorema del producto (3.9):

$$P(B) = \sum_j P(A_j B) = \sum_j P(B/A_j)P(A_j) \tag{3.50}$$



con lo que (3.49) se transforma en

$$P(A_{j_o}/B) = \frac{P(B/A_{j_o})P(A_{j_o})}{\sum_j P(B/A_j)P(A_j)} \tag{3.51}$$

en la que j_o indica un indice determinado de la familia $j(1, \ldots, n)$, y que es el caso general del teorema de Bayes. (bais)

La probabilidad $P(A_j)$ se llama probabilidad "a priori" y la probabilidad condicional calculada segun indica (3.51), probabilidad "a posteriori".

III.20.- Esta denominacion corresponde a un experimento aleatorio muy particular, cuyo esquema es el siguiente: (1) Existe una familia de urnas o bolilleros, que contienen bolillas de dos colores distintos, en proporciones que en general seran diferentes entre sí, y A_j identifica el bolillero j-esimo; (2) El observador conoce la composición de cada bolillero, pero no puede identificarlo antes de hacer una extracción, y se supone que el procedimiento de selección del bolillero previo a dicha extracción es un experimento aleatorio que atribuye a cada uno la probabilidad $P(A_j)$; (3) Efectuada una extraccion, y obtenido el resultado B, se trata de calcular la probabilidad de que el bolillero seleccionado para el experimento sea el A_{j_o}.

El observador, antes de ralizar la extracción, tenía solamente una cierta prob

(a) En estas aplicaciones físicas se ha seguido a VON MISES (20)

bilidad ("a priori" del experimento) de seleccionar el bolillero A_{j_0}. Pero una vez que ha realizado la extracción, puede calcular la probabilidad ("a posteriori" del experimento) de haber seleccionado A_{j_0}, la que en general será distinta de $P(A_{j_0})$. En otros términos, (3.51) permite medir como la presentación del suceso B ha hecho variar el conocimiento que poseía el observador, que se reducía al de las probabilidades "a priori".

Los autores clásicos, que interpretaban la noción de probabilidad como una medida del conocimiento que se tenía acerca de la realización del fenómeno (o "grado de creencia racional" del mismo), consideraron al teorema de Bayes como de gran importancia, y le asignaron una posicion central en el cálculo de probabilidades.

Dentro de esta corriente de ideas, hoy casi totalmente abandonada, las probabilidades "a priori" se denominaban tambien "probabilidad de las causas", y para su conocimiento se aplicaba frecuentemente un "principio de indiferencia", o "principio de razón no suficiente" (segun Santiago Bernouilli), que consistía en aceptar que "cuando no hay razón conocida para predicar de nuestro sujeto una, más bien que otra de varias alternativas, entonces relativamente de tal conocimiento, la afirmación de cada uno de estas alternativas tiene una probabilidad igual"(a). En el esquema anteriormente explicado, si existían n bolilleros, y simplemente es imposible deducir del aspecto exterior (ausencia de marcas, etc.) su composición, la aplicacion de este principio conduce a atribuirles una probabilidad "a priori" a cada uno de ellos igual a $1/n$.

Generalizando este esquema, es claro que el teorema de Bayes tendría una importancia extraordinaria, pues permite fundar un método de inducción, o inferencia para investigar las causas de un fenómeno.

En la teoría moderna, en la que el cálculo de probabilidades es un modelo matemático apto para describir y analizar los fenómenos aleatorios, el teorema de Bayes ha perdido mucha importancia, pues su validez queda restringida al caso en que la elección de las urnas o bolilleros (o de las "causas") es realmente un fenómeno aleatorio en el sentido precisado en el capítulo I, y se conocen las probabilidades respectivas, condiciones ambas no muy frecuentes en la práctica.

Un ejemplo de aplicación válido del teorema de Bayes es el siguiente: un genetista se encuentra ante un individuo (planta o animal), que muestra un caracter dominante. El genetista sabe que el individuo es descendiente de una población de híbridos, y por lo tanto ignora si se trata de un dominante puro, o nuevamente de un híbrido (cualquiera de los dos tipos, como se vió antes, en III.11 puede provenir de una pareja de padres híbridos), pero desea tener una regla de selección para clasificarlo en uno de los dos tipos.

La regla buscada puede obtenerse en este caso por aplicación del teorema de Bayes

(a) J. M. KEYNES (12) pag. 42.

mediante el siguiente procedimiento: lo cruza con un individuo recesivo, y observa si el hijo muestra un caracter dominante o no. En el segundo caso, está seguro de que se trata de un híbrido (si fuera dominante puro, el hijo mostraría seguramente el caracter dominante). En el primer caso no puede afirmar que sea un dominante puro, pero puede calcular la probabilidad de que lo sea, que será mayor que la que se determina partiendo simplemente del hecho que el individuo discutido desciende de una población de híbridos, lo que representa un progreso, sobre todo en el caso en que procediendo adecuadamente puede hacer variar esa probabilidad, haciendola tan cercana a la unidad como crea conveniente.

La probabilidad en cuestión se calcula por aplicación de (3.51). En una población de descendientes de híbridos, la probabilidad de A_1 (dominante puro) es 1/3, y la de A_2 (híbrido) es 2/3. La probabilidad de B condicional de A_1 (siendo el suceso B que el hijo del individuo investigado y de un recesivo muentre el caracter dominante), es igual a la unidad, y la de B condicional de A_2 es un medio. Se tiene:

$$(3.52) \qquad P(A_1/B) = \frac{P(B/A_1)P(A_1)}{P(B/A_1)P(A_1)+P(B/A_2)P(A_2)} = \frac{1.1/3}{1.1/3 + 2/3.1/2} = 1/2$$

con lo que ha mejorado la probabilidad "a priori" igual a 1/3 que el genetista ya conocia. Si este último aplica repetidamente la regla a diversos casos, se equivocará en promedio la mitad de las veces, al decidir que es un dominante puro cuando el hijo de esa cruza especial muestra el caracter dominante, lo que representa un progreso sobre equivocarse los 2/3 de las veces, si se atiene solamente a las probabilidades "a priori".

Si considera, como posiblemente lo hará, que equivocarse la mitad de las veces es un riesgo grande (lo que depende de circunstancias económicas o de cualquier otra clase, incluso prestigio personal, pero extrañas al cálculo de probabilidades, que no tiene unidad para decidir si una probabilidad es grande o pequeña), la regla se puede perfeccionar, en caso que se posible aumentar el número de hijos. Si se pueden observar **n** hijos, si alguno muestra el caracter recesivo el problema está solucionado en contra de la aceptación de la hipótesis de que el padre es dominante puro, como antes, pero si todos muestran el caracter dominante, la probabilidad de B condicional de A_2 es $(1/2)^n$. Entonces (3.52) se transformará en:

$$(3.53) \qquad P(A_1)/B = \frac{1.1/3}{1.1/3 + 2/3.(1/2)^n}$$

la que evidentemente tiende a la unidad cuando n crece indefinidamente (a).

III.21.- Otro caso particular, pero importante, de posible aplicación válida del teorema de Bayes, aparece en la inspección de materiales, cuando es posible admitir que la composición de la producción de un cierto período, o de una partida de magnitud determi

(a) Para más detalles sobre el teorema de Bayes, ver el libro de David (24)

nada que se compra, tiene una probabilidad conocida (por composición se entiende la proporcion de elementos defectuosos sobre el total, conforme a una especificación determinada).

El suceso B es, en este caso, la extraccion de una muestra integrada por n elementos. Las probabilidades condicionales de B, dada una cierta composición, se pueden calcular facilmente (por procedimientos que se explicaran en el capitulo siguiente), y se tienen así todos los elementos para calcular la probabilidad de una cierta composición del total, dado el resultado de una muestra. Establecidas ciertas reglas (cantidad de elementos que integran la muestra, y aceptación o rechazo de la producción o partida segun que en la muestra el número de elementos defectuosos sea menos, igual o mayor n que un cierto número r), mediante el teorema de Bayes será posible calcular la probabilidad de equivocarse, y orientar la modificacion de las reglas para reducir esa probabilidad, en caso necesario.

APENDICE AL CAPITULO III.

III.A.1.- Al comentar el teorema de Bayes, se destacó la diferencia entre la concepción moderna del cálculo de probabilidades, y la que de éste tenían los autores clásicos. La idea de fundamentar axiomaticamente el cálculo de probabilidades como un modelo matemático de los fenómenos aleatorios, en la misma forma en que la geometría lo es del estudio de la posición delos cuerpos en el espacio, es relativamente reciente y se debe a VON MISES, quien la desarrolló extensamente por primera vez en 1919 (a). En este curso se sigue, en forma simplificada, la axiomática formulada por Kolmogoroff (14) en 1933, que se puede expresar en forma más compacta (segun LOEVE (18)):

"Dado un espacio (suceso cierto) y una familia de subconjuntos no vacía, se "cumplen las siguientes propiedades:

1) La familia incluye sus complementarios, uniones e intersecciones en núme"ro finito:

2) Existe una función de conjunto, definida sobre la familia, que es norma"da, no negativa y finitamente aditiva."

De esta manera el cálculo de probabilidades aparece como un caso de aplicación de la teoría de la medida, utilizada en la teoría de funciones de variable real, pero es necesario tener en cuenta que hay problemas del cálculo de probabilidades extraños a los de la teoría de la medida, y que constituyen una problemática especial. Ellos estan dados por las nociónes de DEPENDENCIA E INDEPENDENCIA ya tratadas, por lo que el calculo de probabilidades, así como no se confunde con la combinatoria, tampoco se confunde con la teoría de la medida, sino que usa de ambas como recursos.

Como toda axiomática, el cálculo de probabilidades es aplicable al estudio de cuestiones extrañas a las que motivaron su constitución, como se verá en el capítulo siguiente.

III.A.2.- JERZY NEYMAN (b) llama a esta teoría moderna de las probabilidades, que las interpreta como frecuencias relativas, "teoría clásica" porque realmente lo que interesaba a los interlocutores de PASCAL y FERMAT (algunos de cuyos nombres ha conservado la historia, como el caballero de Meré) era la frecuencia relativa con que se presentaban ciertos resultados en los juegos de azar. Pero desde los comienzos se mezclaron cuestiones de naturaleza más o menos lógica, como las de "grado de conocimiento" o "grado de creencia razonable" ya aludidas anteriormente, y conviene distinguir entre

---

(a) "Grundlagen del Wahrscheinlichkeitrechnung", Mathemtische Zeitschrift, vol 4 (1919) pag. 1.

(b) "Outline of a teory of statistical estimation based on classical theory of probability", Philosophical Transactions of the Royal Society, vol 236 (1937) pag.333.

la "teoría lógica de las probabilidades", y la teoría de las frecuencias relativas, o "teoría clásica" si se acepta la denominación de Neyman.

El campo de aplicación de ambas teorías es notoriamente distinto, porque la teoría clásica permite formular enunciados válidos solamente para conjuntos de sujetos y de observaciones, mientras que nada impide en la teoría lógica hablar de la probabilidad de una única observación.

Sin embargo, la posibilidad de interpretar los enunciados de la teoría clásica en forma operacional bien definida, como se vió al tratar una posible aplicación del teorema de Bayes, interpretando las probabilidades como frecuencias relativas, ha permitido hacer de ella un uso más extendido, y cada vez en constante aumento, que ha planteado nuevos problemas y ha sido motivo del descubrimiento de nuevos teoremas y de capitulos integros de teoría, como la de la inferencia estadística, mientras que la teoría lógica ha quedado practicamente paralizada, y actualmente parece que le quedara como único campo nuevo el de la teoria de la decisión, en ciertos casos todavía no bien esclarecidos(a).

III.A.3.- Además de las diferencias de interpretación y por consiguiente de campos de aplicación, existen ciertas diferencias de contenido formal entre ambas teorías , de las que es buen ejemplo el principio de indiferencia, citado al explicar la interpretación tradicional del teorema de Bayes.

Este principio de indiferencia es peculiar de la teoría lógica (no hay que confundirlo con los sistemas de Laplace, empleados en el capitulo II, porque en estos últimos se parte precisamente de un cierto conocimiento del mecanismo del fenómeno aleatorio para deducir que las alternativas son equiprobables), y aun cuando no todos los partidarios de esa teoría lo aceptan en la forma extrema en que se lo ha usado en el texto, lo cierto es que presenta una dificultad seria para decidir cuando es aplicable y cuando no (ver la extensa discusión que le consagra Keyhes), problema que no preocupó a otros autores, que lo usaron desaprensivamente.

Aplicando sin limitacion el principio de indiferencia, es posible llegar a absurdos notables, como el siguiente: ¿Cual es la probabilidad de vida en Marte? Esta pregunta tiene sentido en la teoría lógica, pero no en la clásica, porque en este último caso no es posible pensar en la frecuencia relativa de vida en Marte.

Si no se sabe nada sobre si en Marte existe una determinada forma de vida (por ejemplo plantas de una cierta especie), se le asigna probabilidad 1/2 a la existencia , y 1/2 a la no existencia. La probabilidad de que haya al menos una forma de vida es la del suceso opuesto a que no haya ninguna, la que se calcula como igual a $(1/2)^n$, o sea

(a) Ver el libro de L. J. SAVAGE: "The Foundations of Statics", New York,J.Wiley,1954.

igual a la probabilidad del suceso que simultaneamente no haya n formas de vida en Marte. Aumentando n suficientemente, la probabilidad

(3.A.1) $$1 - (1/2)^n$$

tiende al límite uno, con lo que se ha establecido que la probabilidad de que haya una forma de vida en Marte es igual a la unidad. Pero es evidente que de esta manera se puede probar que casi cualquier cosa tiene probabilidad unitaria

SEGUNDA PARTE

TEORIA DE LAS VARIABLES ALEATORIAS DISCRETAS

IV. SUCESIONES DE BERNOUILLI Y VARIABLES ALEATORIAS BINOMIALES

IV.1.- En los dos capítulos anteriores ha sido posible explicar, mediante una utilización simple de los teoremas básicos, los fundamentos de algunas de las aplicaciones más importantes del cálculo de probabilidades.

Pero, sin embargo, subsisten las cuestiones planteadas en la introducción acerca de la naturaleza de la convergencia de las frecuencias relativas a las probabilidades, y de la conveniencia de encontrar nuevas constantes asintóticas.

La necesidad de estudiar dichas cuestiones resulta aparente, por ejemplo, en la deducción de la ley de Planck para la radiación del cuerpo negro. ¿Que sentido físico tiene considerar únicamente la distribución canónica, entre todas las que son compatibles con una cantidad dada de energía del sistema?

Para dar una respuesta, es necesario estudiar previamente con mayor detalle el fenómeno aleatorio elemental, aquel que presenta solamente una alternativa entre dos sucesos.

IV.2.- SUCESIONES DE BERNOUILLI. Históricamente, el fenómeno aleatorio elemental se conoce con el nombre de sucesión de Bernouilli, cuya definición es la de una sucesión de observaciones de un fenómeno aleatorio de alternativa simple y probabilidad constante (de otro modo, las observaciones son independientes, ya que el resultado de una no afecta el de la siguiente, al mantenerse constante la probabilidad).

Para Bernouilli y sus contemporáneos y sucesores inmediatos, la materialización de este fenómeno aleatorio consistía en una urna o bolillero, con N bolillas, de las que pN tienen un caracter distintivo (por ejemplo color blanco), y (1-p)N tienen otro caracter distintivo (color negro). De este bolillero se extraen bolillas, conforme a la regla que despues de cada extracción se vuelve la bolilla al bolillero y se agita el mismo.

Intuitivamente se comprende que el resultado de cada extracción es impredecible, (no puede afirmarse si será una bolilla blanca o negra) y por cierto independiente del resultado de la anterior, mientras que la experiencia indica que prolongando suficientemente la sucesión de extracciones, la frecuencia relativa de la aparición de cada color presenta una estabilidad muy notable.

Lo que Bernouilli tenía en mente al hacer esta formulación, era un juego de azar,

con la probabilidad p de ganar, y la probabilidad $q=1-p$ de perder. La serie de n observaciones representaba una partida de n repeticiones del juego, lo que explica la terminología de llamar resultado favorable a uno de los sucesos, generalmente el de probabilidad p, y desfavorable al otro.

Aunque contemporaneamente con Bernouilli, el astrónomo y matemático inglés HALLEY fundó, en 1693, el cálculo de las primas de seguros sobre la asimilación de la mortalidad a un fenómeno aleatorio, existiendo indicaciones precisas de que aún con anterioridad la cuestión había sido considerada de manera más o menos parecida por el famoso estadista holandés JAHN DE WITT (a), y posteriormente se encontró que muchos otros fenómenos tenian un comportamiento similar, la formulacion en términos de extracciones de bolilleros y casos favorables y desfavorables ha perdurado hasta muy recientemente.

IV.3.- En este capítulo se estudiaran solamente las propiedades directas de las sucesiones de Bernouilli, es decir, las que resultan por deducción, dejando para la parte cuarta y última la consideración detallada de los problemas de inferencia sobre el fenómeno aleatorio (por ej., determinar el valor de p), partiendo de las observaciones.

El problema básico es calcular la probabilidad de que, en una serie de n observaciones, aparezca exactamente r veces uno de los sucesos (y por lo tanto $n-r$ veces el otro).

Aplicando el teorema del producto, para el caso de independencia, si la probabilidad en cada observacion del suceso que se presenta r veces en n es p, resulta inmediatamente que la probabilidad de una cierta sucesión de n observaciones con r veces la alternativa en cuestión, es igual a $p^r q^{n-r}$. Cada uno de estas sucesiones representa un suceso excluyente, y la probabilidad de la aparicion de una cualquiera de ellas será, por aplicación del teorema de adición, igual a la suma de las probabilidades. El número de sucesiones distintas será el de combinaciones de n elementos tomados en grupos de r, y como cada una de las sucesiones tiene igual probabilidad, la aplicación del teorema de adición se reduce al producto de dicha probabilidad por el del número de combinaciones:

$$P_n(r) = \frac{n!}{r!(n-r)!} p^r q^{n-r} \tag{4.1}$$

Como verificación de (4.1), se puede sumar sobre todos los r posibles (de cero a n), y como se trata de un sistema completo de sistemas excluyentes, la suma tiene que ser igual a la unidad, lo que resulta inmediatamente por ser cada uno de los sumandos el término respectivo del desarrollo de la potencia n-ésima del binomio $(p+q)=1$.

$$\sum_{r=0}^{n} \frac{n!}{r!(n-r)!} p^r q^{n-r} = (p+q)^n = 1 \tag{4.2}$$

(a) Ver TODHUNTER (23), cap. V.

IV.4.- VARIABLES ALEATORIAS DISCRETAS. Mediante (4.1) es posible asignar una cierta probabilidad a cada uno de los valores de una variable, que es discreta por estar restringida a tomar solamente valores enteros, en este caso los $n+1$ valores sucesivos desde 0 hasta n. Este es, históricamente, el primer ejemplo conocido de una v a r i a b l e a l e a t o r i a (o estocástica) (a).

La definición rigurosa de variable aleatoria exige tener presente que, en realidad es una generalización del concepto de función, porque los valores numéricos tienen como argumento un suceso. Estas funciones generalizadas son de ocurrencia frecuente (por ejemplo, el área de las figuras planas, o la longitud de un arco de curva), y se las denomina funcionales. Mas precisamente, una v a r i a b l e a l e a t o r i a e s u n f u n c i o n a l d e f i n i d o s o b r e u n e s p a c i o d e p r o b a b i l i d a d e s, y el conjunto de sus argumentos es el conjunto de sucesos, o sea la familia de subconjuntos que los representa en el espacio de probabilidades.

IV.5.- VARIABLES ALEATORIAS BINOMIALES. Las variables aleatorias en las que la correspondencia entre sus valores y las probabilidades, o l e y d e p r o b a b i l i d a d e s, tiene la forma dada en (4.1) se denominan, por la analogía entre la ley y el término genérico del desarrollo de la potencia de un binomio, v a r i a b l e s a l e a t o r i a s b i n o m i a l e s, locución por la que se reemplazará en lo sucesivo el esquema del bolillero con su regla de extracción.

El estudio de estas variables no queda terminado con la determinación de su ley de probabilidad, pues del comportamiento de esta última es posible deducir ciertas propiedades importantes, comunes a todas ellas.

Formando el cociente entre las probabilidades correspondientes a dos valores sucesivos de la variable aleatoria, se tiene:

$$\frac{P_n(r+1)}{P_n(r)} = \frac{r!(n-r)!}{n!} \cdot \frac{n!}{(r+1)!(n-r-1)!} \cdot \frac{p^{r+1} \cdot q^{n-r-1}}{p^r \cdot q^{n-r}} \tag{4.3}$$

y desarrollando los factoriales en producto de números enteros y simplificando:

$$\frac{P_n(r+1)}{P_n(r)} = \frac{n-r}{r+1} \cdot \frac{p}{q} \tag{4.4}$$

Este cociente es mayor o menor que la unidad según que r sea menor o mayor que $np-q$, lo que puede verse facilmente:

$$p(n-r) \gtrless (r+1)q \qquad np-q \gtrless r(p+q) = r \tag{4.5}$$

La diferencia $np-q$ es, en general, un número no entero. Suponiendo en este caso que r_1 sea el único entero comprendido en el intervalo $np-q$ y $np-q+1$, las probabilidades (4.1) crecen desde $r=0$ hasta $r=r_1$, y disminuyen a partir de $r=r_1$ hasta $r=n$.

(a) Tambien llamada a veces v a r i a b l e c a s u a l, por traducción directa del italiano "variabile casuale".

La ley de probabilidades (4.1) define por lo tanto una función de r que es monótona creciente desde 0 hasta alcanzar su único máximo en r_1, luego del cual es monótona decreciente.

En el caso especial en que $np-q$ es un número entero, para $r = np-q$ el cociente (4.3) es igual a la unidad. Las probabilidades de r y de $r+1$ seran iguales entre sí, y aplicando el razonamiento anterior se tiene que (4.1) es una función monótona creciente desde 0 hasta $r = np-q$, permanece constante hasta $np-q+1$, y luego es monótona decreciente.

IV.6.- La interpretación concreta de (4.1) y del caracter simple de su comportamiento en función de r, es la siguiente: dado un conjunto de N series de observaciones, integrada cada serie por n observaciones, la frecuencia relativa del suceso con probabilidad p puede tomar $n+1$ valores, al variar r de 0 hasta n. Para cada uno de estos valores, (4.1) define una probabilidad que para N suficientemente grande es aproximada por la frecuencia relativa (número de series de observaciones en que se presenta el valor r/n, sobre el total N) con que r/n aparece al repetir la serie de observaciones.

La probabilidad máxima corresponde a un valor de r/n que es cercano a p, y cuanto más se aleja la frecuencia relativa de ese valor, tanto más pequeña es la probabilidad.

La aplicación sistemática de la axiomática del Capítulo II, y de los teoremas de ella derivados, permite establecer un comportamiento de las frecuencias relativas que es coincidente con el que puede observarse de presentar oscilaciones en torno a un cierto valor. Esto es una indicación de la consistencia entre el modelo matemático adoptado y los fenómenos aleatorios que se trata de estudiar con él, y además permite precisar la naturaleza de las fluctuaciones, que se caracterizan por obedecer a una ley de probabilidades bien determinada.

IV.7.- Como ejemplo, se estudiará el comportamiento de una moneda supuesta equilibrada (probabilidad de cada una de las figuras igual a 1/2), en series de 10 tiros.

El valor de r al cual corresponderá la probabilidad máxima es el entero siguiente a $np-q = 5-1/2 = 4,5$ o sea $r_1 = 5$, que en este caso coincide con el valor teórico que podría esperarse si no hubiera oscilaciones en la frecuencia relativa.

La probabilidad de ese valor $r_1 = 5$ será

$$P_{10}(5) = \frac{10!}{5!5!}\left(\frac{1}{2}\right)^{10} = \frac{2.3.2.7.3.}{1024} = 0,246 \tag{4.6}$$

La probabilidad de los valores vecinos, $r = 4$ y $r = 6$ serán iguales entre sí, porque al ser $p = q = 1/2$ (4.1) resulta una función simétrica:

(4.7) $$P_{10}(4) = P_{10}(6) = \frac{10!}{6!\,4!}\left(\frac{1}{2}\right)^{10} = 0{,}205$$

La probabilidad de que r tome un valor mayor que 6 o menor que 4 será aproximamente, teniendo en cuenta que $\sum_{r=4}^{6} P_{10}(r) = 0{,}656$:

(4.8) $$P_{10}\left[(r<4)+(r>6)\right] = 1 - 0{,}656 = 0{,}434$$

y la de los valores extremos $r=0$ y $r=10$

(4.9) $$P_{10}(0) = P_{10}(10) = \left(\frac{1}{2}\right)^{10} = \frac{1}{1.024} \cong 0{,}001$$

Conforme a lo que se había deducido por el estudio de la ley de probabilidades, la probabilidad máxima corresponde a una frecuencia relativa cercana (en este caso coincidente) del valor teórico. Los valores vecinos tienen también probabilidades relativamente elevadas, pero los valores alejados poseen probabilidades muy pequeñas, y en una repetición de la serie de $n=10$ tiros con una moneda equilibrada, se observaran muy rara vez.

IV.8.- Por aplicación directa de la ley binomial, es posible resolver una variedad de problemas correspondientes a diversas materializaciones del fenómeno estudiado.

Por ejemplo, supuesta una tasa de mortalidad igual a 0,03 ¿Cual es la probabilidad de que en una población de 100 personas que poseen la edad a que corresponde esa tasa de mortalidad, se produzcan en el año no más de 10 fallecimientos?

La probabilidad buscada es igual a la suma de los términos de la ley binomial, en la que $n=100$, $p=0{,}03$, $q=0{,}97$ y r toma los valores desde 0 hasta 10:

(4.10) $$P = \sum_{r=0}^{10} \frac{100!}{r!(100-r)!}\, 0{,}03^{r} \cdot 0{,}97^{100-r} = 0.999875$$

Esta probabilidad es tan cercana a la unidad, que en el caso de que la observacion del fenómeno se repita a lo largo de varios años, en todos ellos se observará un número de fallecimientos menor de diez.

Si la pregunta se modifica, planteando la cuestión: en una serie de veinte años en que se observan grupos de 100 personas con tasa de mortalidad de 0,03 ¿en cuantos años no se observará ningun fallecimiento? Habrá que calcular un solo término de la ley binomial, el que resulta igual a 0,047553, o sea bastante próximo a 0,05 . Por lo tanto, en veinte años se observará que no hay ningun fallecimiento en $20.0{,}05=1$ año.

Si en vez de ser un grupo de cien personas con una edad determinada y una tasa de mortalidad de 0,03 , se trata de cien piezas de máquina en un proceso industrial que tiene un nivel de calidad (porcentaje de elementos defectuosos conforme a cierta especificación) igual a 0,03, y se supone además que las piezas han sido extraídas de la producción de acuerdo con un procedimiento que autoriza a establecer la independencia de las observaciones, lo que es fundamental, las preguntas anteriores se transforman en las

dos siguientes: ¿Cual es la probabilidad de que una muestra de 100 elementos tenga un número de elementos defectuosos comprendidos entre 0 y 10 ? y ¿Cual es la probabilidad de que una muestra de cien elementos no tenga ningun elemento defectuoso? Los valores numéricos son los mismos.

Es importante señalar que toda regla de aceptación de partidas mediante muestra debe ser discutida conociendo las probabilidades del tipo de las calculadas a fin de establecer la frecuencia relativa de los posibles errores. Pero esta discusión es compleja, y corresponde a la inferencia estadística.

Finalmente, si en un censo de población se hubiera tabulado los grupos familiares la pregunta ¿Cual es la frecuencia relativa de las familias de cinco hijos, con dos hijos varones? podría contestarse sin estudiar los resultados del censo, sino simplemente calculando:

(4.11) $$P = \frac{5!}{2!3!}\, 0{,}51^2 \cdot 0{,}49^3$$

IV.9.- A pesar de la simplicidad de las aplicaciones de la ley binomial, en la práctica se choca con la grave dificultad de los cálculos numéricos. Cuando n es pequeño, y las probabilidades son números fraccionarios, cuyas potencias son facilmente calculables, como en el ejemplo de la moneda, el problema no es serio. Pero si n es grande y las probabilidades no son números sencillos, la cuestión cambia. Los factoriales crecen con extraordinaria rapidez (el factorial de 10 tiene 7 cifras, el de 20 es un número de 19 cifras y el 100 tiene 158 cifras), y hay que usar logaritmos para calcular las potencias de probabilidades.

La ley binomial ha sido tabulada, para ciertos valores de n . Por ejemplo, la tabla de ROMIG (31) da los valores (4.1), para n entero y múltiplo de cinco, desde $n=50$ hasta $n=100$, y p desde 0,01 hasta 0,50 (de centésimo en centésimo). A fin de facilitar las aplicaciones, la misma tabla da también los valores acumulados de los terminos de la ley binomial, incluyendo los valores de r para los que las probabilidades correspondientes no son inferiores a la aproximación con que ha sido calculada (un millonésimo).

Para valores de n elevados, y para estudiar nuevas propiedades de la ley binomial, especialmente cuando se requiere sumar un número arbitrario de términos, la forma (4.1) la hace de muy dificil manejo, y es necesario utilizar otros métodos para describir la ley de probabilidades.

IV.10.- Un método de descripcion muy adecuado para estudiar las propiedades de la ley de probabilidades, es el de los momentos, inspirado en los métodos empleados en mecáni

ca para describir un sistema de puntos materiales.

El método consiste en interpretar las probabilidades asignadas por (4.1) a cada valor de r, como masas de puntos materiales, dispuestos sobre una línea recta a una distancia igual a r de un cierto origen.

El momento de orden j se define como:

(4.12) $$\mu_j = \sum_{r=0}^{n} P_n(r) r^j$$

Es fácil ver que si se conoce el campo de definición de la variable aleatoria (en el caso los $n+1$ enteros desde 0 hasta n), las probabilidades correspondientes quedan definidas como la solución única de un sistema de ecuaciones lineales cuyos términos independientes son $n+1$ momentos (desde el de orden cero, que es simplemente la suma de las probabilidades, hasta el de orden n) (a):

(4.13) $$\begin{array}{c} P_n(0) + P_n(1) + \ldots + P_n(n) = 1 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ P_n(1).1 + P_n(2)2^n + \ldots + P_n(n)n^n = \mu_n \end{array}$$

La forma especial de los coeficientes del sistema de ecuaciones dado asegura que su determinante nunca es nulo.

La utilidad de esta forma de determinar la ley de probabilidades parece ser ilusoria, porque en la definición de cada momento intervienen las probabilidades que precisamente se desea determinar. Pero tanto en el caso de la ley binomial como de otras leyes usuales que se estudiarán más adelante, el valor de los momentos puede deducirse de la forma del término genérico de la ley, sin necesidad de calcularlo explícitamente.

Subsiste, en cambio, la dificultad de que no es sencillo resolver un sistema de ecuaciones lineales. El número de operaciones requerido crece rápidamente al crecer n, lo que puede significar una imposibilidad práctica, a menos que se disponga de una computadora electrónica.

El método de los momentos no es, por lo tanto, adecuado para calcular la probabilidad de un cierto valor de la variable aleatoria, lo que en el caso de la binomial puede hacerse mediante otros procedimientos que se estudiarán en los capítulos VI y VII.

Pero en cambio, tiene la importante ventaja de que limitando el número de momentos a considerar se tiene una descripción simple de la ley, que necesariamente será incompleta (porque no intervienen todos los momentos), pero suficiente para establecer propiedades importantes.

IV.11.- Los momentos que ha encontrado más convenientes de utilizar son los de primer y

(a) Disponiendo de un número suficiente de momentos, es posible determinar también el campo de definición de la variable. Para una demostración simple, ver M.Watanabe:"Extensions of Theorems of Bernouilli, Poison and Tchebicheff", Tohoku Mat. Journal, Vol. 12 (1917) pag. 24.

segundo orden.

Conforme a la definición (4.12), el momento de primer orden será:

(4.14) $$\mu_1 = \sum_{r=0}^{n} P_n(r)r = \sum_{r=0}^{n} \frac{n!}{r!(n-r)!} p^r q^{n-r} . r$$

Para calcularlo efectivamente, hay que tener en cuenta que el primer término es nulo, porque $r=0$. Eliminandolo, r se puede simplificar en todos los demás términos con el factorial que figura en el denominador. Ademas , y por la misma razon (r no nulo) se podra sacar factor común p , conjuntamente con n , resultando:

(4.15) $$\mu_1 = np \sum_{r=1}^{n} \frac{(n-1)!}{(r-1)!(n-r)!} p^{r-1} q^{n-r}$$

Efectuando el cambio de variables $r-1=r'$, la suma del segundo miembro es simplemente la de todas las probabilidades de una ley binomial de orden n-1 , porque $n-1-(r-1) = n-r$, que es igual a la unidad, por tratarse de un sistema completo de sucesos excluyentes, con lo que se tiene finalmente:

(4.16) $$\mu_1 = np$$

El momento de primer orden se acostumbra a designar con la letra m , y por razones que se veran mas adelante, se suele denominar v a l o r p r o b a b l e, y tambien e s p e r a n z a m a t e m á t i c a de la variable aleatoria.

El momento de segundo orden, aplicando el mismo tipo de simplificaciones que para el de primero, resulta ser:

(4.17) $$\mu_2 = \sum_{r=0}^{n} P_n(r)r^2 = np \sum_{r=1}^{n} \frac{r\,(n-1)!}{(r-1)!(n-r)!} p^{r-1} q^{n-r} = np \left(\sum_{r=2}^{n} \frac{(n-1)!}{(r-2)!\,(n-r)!} p^{r-1} q^{n-r} + \right.$$

$$\left. + \sum_{r=1}^{n} \frac{(n-1)!}{(r-1)!\,(n-r)!} p^{r-1} q^{n-r} \right) = np\left[(n-1)p+1\right] = n^2p^2 + npq$$

Si efectúa un cambio del origen de referencia, el momento de segundo orden varia rá su valor, y se puede probar que pasará por un mínimo cuando el origen se lleva a una abcisa r por (r-k), donde k es una constante arbitraria que indica el desplazamiento positivo o negativo del origen :

(4.18) $$\mu'_2 = \sum_{r=0}^{n} P(r)(r-k)^2 = \sum_{r=0}^{n} P_n(r)r^2 - 2k \sum_{r=0}^{n} P_n(r)r + k^2 = \mu_2 - 2km + k^2$$

sumando y restando m^2:

(4.19) $$\mu'_2 = \mu_2 + (m-k)^2 - m^2$$

Cualquiera que sea el signo de la diferencia (m-k), su cuadrado es positivo, y como los otros dos términos del segundo miembro son constantes, el valor del momento será mínimo cuando $m-k=0$

La transformación correspondiente al traslado del origen de la variable aleatoria a su primer momento se llama c e n t r a r la variable. Los momentos referidos al

nuevo origen se llaman m o m e n t o s c e n t r a d o s , y es inmediato probar que el momento de primer orden de una variable centrada siempre es nulo. El momento centrado de segundo orden, mínimo entre todos los momentos de ese orden que se pueden formar desplazando el origen, recibe el nombre de v a r i a n z a, y se designa comunmente con σ^2.

En el caso de la ley binomial, teniendo en cuenta que por aplicación de (4.19) la relación entre la varianza y el momento de segundo orden al origen calculado en (4.17) es:

(4.20) $$\sigma^2 = \mu_2 - m^2$$

la varianza es directamente igual a

(4.21) $$\sigma^2 = npq$$

IV.12.- El momento de primer orden de la variable binomial constituye una m e d i d a d e p o s i c i o n , en el sentido de que identifica los valores en torno a los cuales se concentra la probabilidad, ya que como se vió en (IV. 5), la ley de probabilidad (4.1) es una función definida para los valores enteros de r , que pasa por un máximo en el entero comprendido entre np y np+1 , si np no es entero, y tiene dos valores máximos iguales a np y np+1 si np es entero.

La varianza completa esta descripción sumaria de la ley, posibilitando la introducción de una m e d i d a d e d i s p e r s i o n. Si se llama x a la v a r i a b l e a l e a t o r i a c e n t r a d a , la sumatoria que define la varianza puede descomponerse en dos partes, segun que el valor absoluto x de la variable sea inferior a t veces σ, o igual o mayor que esa magnitud:

(4.22) $$\sigma^2 = \sum_{|x|<t\sigma} P_n(x)x^2 + \sum_{|x|\geq t\sigma} P_n(x)x^2$$

Las dos sumatorias del segundo miembro son positivas, porque cada uno de sus términos (producto de una probabilidad por un cuadrado) es también positivo, por lo que si se suprime la primera se tiene una desigualdad de sentido definido:

(4.22) $$\sigma^2 \geq \sum_{|x|\geq t\sigma} P_n(x)x^2$$

Si en el segundo miembro se reemplaza x^2 por $(t\sigma)^2$, el signo de la desigualdad se mantiene:

(4.23) $$\sigma^2 \geq t^2\sigma^2 \sum_{|x|\geq t\sigma} P_n(x)$$

Simplificando la varianza:

(4.24) $$1 \geq t^2 \sum_{|x|\geq t\sigma} P_n(x)$$

resulta que la p r o b a b i l i d a d d e q u e l a v a r i a b l e a l e a t o r i a c e n t r a d a a d o p t e v a l o r e s m a y o r e s

que t veces σ, es menor que $1/t^2$:

(4.25) $$\sum_{|x| \geq t\sigma} P_n(x) \leq \frac{1}{t^2}$$

que es la célebre desigualdad de TCHEBICHEFF (a).

El papel que ~~se~~ desempeña esta desigualdad es reemplazar el dificil problema de sumar términos de la ley binomial, introduciendo una cota superior a los valores que esas sumas pueden tomar, cuando la variable aleatoria centrada supera una magnitud fijada en función de la raíz cuadrada de la varianza. Debido a que de esta forma se puede acotar la dispersión en probabilidad de la variable, dicha raiz cuadrada (considerada siempre con signo positivo) se denomina usualmente como d i s p e r s i ó n.

IV.13.- La descripcion sumaria del comportamiento de la variable aleatoria binomial, mediante una medida de posición y otra de dispersión, se puede generalizar a otras variables aleatorias. En este capitulo, se limitará esta generalización a la variable aleatoria f r e c u e n c i a r e l a t i v a, en las sucesiones de Bernouilli.

Reemplazando r por r/n en (4.14) y (4.17), se tiene inmediatamente que:

(4.26) $$m = p \qquad \sigma^2 = \frac{pq}{n}$$

Mientras que en el caso de la variable aleatoria binomial, al crecer n aumenta indefinidamente su campo de definición, y la única conclusión que nos permite alcanzar la desigualdad de TCHEBICHEFF es que la dispersión crece tambien indefinidamente, pero con mas lentitud, dado que la σ es proporcional a la raiz cuadrada de n, en la frecuencia relativa es posible llegar a un resultado notablemente mas preciso: l a p r o b a b i l i d a d d e q u e e l v a l o r a b s o l u t o d e l a d i f e r e n c i a e n t r e l a f r e c u e n c i a r e l a t i v a y s u p r i m e r m o m e n t o s e a m a y o r q u e u n c i e r t o ε p o s i t i v o y a r b i t r a r i o, t i e n d e a l l i m i t e c e r o c u a n d o n c r e c e i n d e f i n i d a m e n t e (teorema de BERNOUILLI).

Haciendo $\varepsilon = \sigma t$, de donde $t = \varepsilon/\sigma$, la aplicacion de la desigualdad de Tchebicheff conduce inmediatamente a:

(4.27) $$P\left(\left|\frac{x}{t} - p\right| > \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2} = \frac{pq}{n\varepsilon^2}$$

El producto pq no puede superar un cuarto, como resulta de que siendo $p+q=1$, se tiene que $pq = (1/2 + a)(1/2 - a) = 1/4 - a^2$, por lo que si se fija un δ positivo y arbitrario, basta hacer crecer suficientemente n para que el segundo miembro sea menor que δ, u n i f o r m e m e n t e e n p y q, con lo que se ha demostrado el teorema de Bernouilli.

(a) "Sur les valeurs Moyennes" Journal de Mat. Pures et Appliqués. vol.12(1867) pag.177.

El teorema de Bernouilli se publicó en 1713, siglo y medio antes del descubrimien por Tchebicheff de la desigualdad que lleva su nombre, y su demostración original resultó sumamente enojosa, debido a la dificultad de sumar términos de la ley binomial(a)

IV.14.- Como el primer momento de la variable aleatoria frecuencia relativa, es precisamente la probabilidad p del suceso que se presentó r veces en n observaciones, el teorema de Bernouilli permite precisar el comportamiento de las frecuencias relativas, de presentar una estabilidad creciente al aumentar n, y la expresión de que las frecuencias relativas tienen por límite a la probabilidad, a las que se aludió en la introducción, en el sentido de que la probabilidad de un desvío entre la frecuencia relativa y la probabilidad tiene a su vez una probabilidad que tiende a cero cuando n crece indefinidamente. En otros términos, una vez fijado tanto el ε como el δ, aplicando (4.27) es posible determinar el n mínimo necesario para que si la serie de n observaciones se repite N veces (N suficientemente grande), la frecuencia relativa de las series en las que el valor absoluto de la diferencia entre r/n y p es mayor que ε, será menor que δ.

Es preciso tener en cuenta que (4.27) no significa la convergencia de las frecuencias relativas al número p en el sentido ordinario del analisis. Para que se verifique esta última convergencia, se requiere determinar un n_1 tal que para todo $n > n_1$ se verifique la desigualdad del primer término de (4.27), pero el teorema de Bernouilli no permite llegar a esa conclusión.

Dado un conjunto de N series de observaciones, cada una de las cuales tiene una longitud n_2 mayor que el n_1 necesario para que segun (4.27) la probabilidad de la diferencia sea menor que δ, la interpretacion dada anteriormente permite enunciar que la diferencia entre la frecuencia relativa y la probabilidad será menor que ε al menos en $(1-\delta)N$ de las series, pudiendo ser mayor en la otras δN. Si las series de observaciones se alargan a $n_3 > n_2$, nuevamente se llega a la misma interpretación, pero el teorema de Bernouilli no permite afirmar que en cada una de la $(1-\delta)N$ series de observaciones que para n_2 cumplian la desigualdad la continuan cumpliendo en n_3. Lo único que autoriza a enunciar será que las sucesiones excepcionales (que no cumplen la desigualdad) seran en número inferior a δN, pero una sucesión que la cumplía en n_2 puede no cumplirla en n_3 y recíprocamente.

(a) USPENSKY (22), cap.IV, da una demostración similar a la original de Bernouilli, que puede consultarse para comparar extensiones.

En realidad este tipo de convergencia de las frecuencias relaitivas a las probabilidades, es una ampliacion del concepto de convergencia ordinaria, por lo que Cantelli (a) la llamó convergencia en probabilidad, concepto que cuando se asimila la probabilidad a la noción de medida utilizada en la teoria de funciones, resulta ser equivalente al de convergencia en medida.

Es muy natural preguntarse si es posible estudiar en que casos se verifica realmente la convergencia en el sentido del análisis, de donde resulta la necesidad de estudiar la probabilidad de la convergencia, expresión que conjuntamente con la de convergencia en probabilidad podría parecer un simple juego de palabras, pero que no lo es, y que da un lugar a un importante conjunto de problemas matematicos, cuya consideracion fué iniciada por Borel (b), y que escapan a las posibilidades de este curso, por requerir conocimientos de teoría de funciones que no figuran entre sus prerrequisitos (c).

IV.15.- Es necesario señalar que el teorema de Bernouilli no constituye propiamente una demostracion de que las frecuencias relativas convergen a las probabilidades, ni mucho menos de la estabilidad repetidamente observada. Estas propiedades son experimentales, y resultan implícitas en las propiedades mas simples que se eligieron para formular la axiomática del capítulo II.

El teorema de Bernouilli es simplemente una prueba de la consistencia del modelo matemático introducido, con respecto al conjunto de fenomenos que se desea estudiar (ver, por ejemplo, la discusión de CASTELNUOVO (1)), y simplemente permite formular esas propiedades con la terminología del modelo, obteniendo mayor claridad de expresión.

A este respecto, es interesante observar que desde el principio del desarrollo del cálculo de probabilidades no se cayó en la confusión de creer que se había demostrado la tendencia a un límite de las frecuencias relativas. Segun TODHUNTER ((23) cap VII) el propio Bernouilli entendió que la mayor importancia de su teorema residia en dar un método para aproximar p cuando no se conocía su valor, o sea en constituir lo que hoy se llamaría un método de inferencia estadística.

En efecto, fijados ε y δ, es posible determinar el n necesario para que la frecuencia relativa se aproxime con probabilidad mayor que δ:

(4.28) $$\frac{1}{4\varepsilon^2} = \delta$$

(a) "La tendencia ad un limite nel senzo del Calcolo delle Probabilita", Rendiconti del Circolo Matematico di Palermo, vol.41(1916)pag.191.
(b) "Les probabilites denombrables" Rend. Circ. Mat. Palermo, vol. 27, (1909)pag.247.
(c) En el libro de FELLER (4) se puede ver un tratamiento relativamente elemental de estas cuestiones.

Si $\varepsilon=0,01$ y $\delta=0,01$, n resulta igual 250.000. Este número resulta muy elevado y es considerablemente mayor que el mínimo indispensable, debido precisamente a la naturaleza simple de la acotación que da la desigualdad de Tchebicheff, que es demasiado imprecisa para una aplicación de este tipo, que requiere otros métodos mas exactos, no solamente desde el punto de vista numérico, sino también de su interpretacion operacional, lo que se tratará en la parte IV.

IV.16.- APLICACION A LA DEDUCCION DEL TEOREMA DE WEIERSTRASS. Los resultados del calculo de probabilidades pueden utilizarse en campos muy distintos del de los fenómenos aleatorios, siendo un buen ejemplo la demostracion elemental del importante teorema: toda funcion continua en un intervalo cerrado puede ser aproximada uniformente por polinomios.

Sea el polinomio, en el que x está comprendido en el intervalo (0,1):

$$P_n(x) = \sum_r f\left(\frac{r}{n}\right) \binom{n}{r} (1-x)^{n-r} x^r \tag{4.29}$$

Es evidente que para todo x comprendido en (0,1), se verifica

$$f(x) = \sum_r f(x) \binom{n}{r} x^r (1-x)^{n-r} \tag{4.30}$$

Si f(x) es continua en el intervalo cerrado (0,1), se verifica que es acotada en dicho intervalo:

$$f(x) \leqslant M \tag{4.31}$$

y que la continuidad es uniforme:

$$|f(x_1) - f(x_2)| < \varepsilon \qquad \text{si} \quad |x_1 - x_2| < \delta \tag{4.32}$$

Fijados el ε y el δ, basta tomar $n > 1/\delta$ para que dado x exista un r/n tal que

$$\left| x - \frac{r}{n} \right| < \delta \tag{4.33}$$

El valor absoluto de la diferencia entre f(x) y el polinomio, teniendo en cuenta (4.29) y (4.30), será:

$$|f(x) - P_n(x)| \leqslant \sum_r \left|f(x) - f\left(\frac{r}{n}\right)\right| \binom{n}{r} x^r (1-x)^{n-r} \tag{4.34}$$

Descomponiendo la sumatoria del segundo miembro en dos partes, segun que $\left|x-\frac{r}{n}\right|<\delta$ o $\left|x-\frac{r}{n}\right| \geqslant \delta$

$$|f(x)-P_n(x)| \leqslant \sum_{|x-r/n|<\delta} \left|f(x)-f\left(\frac{r}{n}\right)\right| \binom{n}{r} x^r (1-x)^{n-r} + \sum_{|x-r/n|\geqslant\delta} \left|f(x)-f\left(\frac{r}{n}\right)\right| \binom{n}{r} x^r (1-x)^{n-r} \tag{4.35}$$

Por la continuidad uniforme $\left|f(x)-f\left(\frac{r}{n}\right)\right|<\varepsilon$ si $\left|x-\frac{r}{n}\right|<\delta$ y por lo tanto se verifica que la primer sumatoria es menor que ε, ya que los terminos del desarrollo del binomio son todos positivos, y la suma de un número arbitrario de ellos no puede superar la unidad.

(4.36) $$\sum_{|x-\frac{r}{n}|<\delta} \left| f(x) - f\left(\frac{r}{n}\right) \right| \binom{n}{r} x^r (1-x)^{n-r} \leqslant \varepsilon$$

En la segunda sumatoria, la acotacion de f(x) asegura que $\left| f(x) - f\left(\frac{r}{n}\right) \right| \leqslant 2M$, y por lo tanto:

(4.36) $$\sum_{|x-\frac{r}{n}|\geqslant\delta} \left| f(x) - f\left(\frac{r}{n}\right) \right| \binom{n}{r} x^r (1-x)^{n-r} \leqslant \varepsilon$$

si n es lo suficientemente grande para que por aplicación del teorema de Bernouilli la suma de los términos del binomio sea menor que $\varepsilon/2M$.

Teniendo en cuenta (4.35) y (4.36):

(4.37) $$\left| f(x) - P_n(x) \right| < 2\,\varepsilon$$

que se verifica con independencia de x, que desempeña el papel de p en el teorema de Bernouilli, en cuya demostración se vió que la acotacion (4.36) era independiente de p.

Si la funcion está definida en un intervalo (a,b), basta hacer un cambio de variables, para llevarla a un intervalo de longitud unitaria, con lo que queda demostrado el teorema , con toda generalidad.

Existen otros problemas, como los relativos al estudio del desarrollo de los números reales en un sistema de numeración de base 10, que tampoco son fenómenos aleatorios pero en los cuales hay una estrecha analogía con los resultados del cálculo de probabilidades. Considerando que la aparicion de un dígito tiene una probabilidad 1/10, es posible demostrar importantes teoremas (a) de la teoría de los números.

(a) BOREL, memoria citada en pag.60.

C A P I T U L O V.

A D I C I O N D E V A R I A B L E S A L E A T O R I A S.-

LEYES DE LOS GRANDES NUMEROS

V.1.- Al estudiar en el capítulo anterior la variable aleatoria binomial, se encontró que al crecer n la probabilidad se concentraba en la vecindad del primer momento o v a l o r p r o b a b l e, autorizando así a formular predicciones sobre el fenómeno aleatorio, en un sentido precisado por el teorema de Bernouilli.

Es evidente la conveniencia de poder generalizar estos resultados a fenómenos aleatorios más complejos, caracterizados por otras variables aleatorias. Por ejemplo, en la estadística de Maxwell-Boltzmann, habría que probar que al crecer n (en el caso , número de partículas), la probabilidad se concentra en torno a un valor de la variable aleatoria asociada, que corresponde a la distribución canónica. Como en los sistemas físicos que se estudian con esa estadística (cuerpos gaseosos, cuerpo negro) el número de partículas es muy grande, la probabilidad se concentra en torno del estado macroscópico correspondiente a la distribución canónica, y desde el punto de vista de estudiar propiedades del sistema resulta justificado limitarse a esa distribución, que corresponde al estado que se observará con una frecuencia tan cercana a la unidad, que puede considerarse equivalente a la certeza.

Pero, en vez de estudiar variables aleatorias particulares, en vista de determinadas aplicaciones, interesa en este curso establecer métodos generales, por lo que dejando de lado las estadísticas físicas , introducidas sólo a título ilustrativo, se extenderá en este capítulo el teorema de Bernouilli, mediante la introducción de la operación de suma de variables aleatorias.

V.2.- ADICION DE VARIABLES ALEATORIAS DISCRETAS. Sea dos variables aleatorias (en el sentido antes definido de funcionales sobre espacios de probabilidades) x e y, y dado un suceso del espacio producto, se le asocia la suma de las variables aleatorias. El resultado será una nueva variable aleatoria, definida sobre el espacio producto, y el problema general de la adición consiste en deducir su ley de probabilidades a partir de la leyes de las variables sumandos.

Un ejemplo simple de variable aleatoria suma de otras, está dado por la variable binomial, que puede interpretarse como la suma de n variables elementales, que toman cada una los valores 0 y 1, con probabilidades q y p.

La ley de probabilidades (4.1) se obtuvo directamente sobre el espacio producto ,

porque debido a que las variables sumandos son independientes entre sí (en el sentido de que la probabilidad con que una de las variables sumandos toma un cierto valor es independiente de los valores que tomen las otras), fué posible definir en forma simple en dicho espacio producto un sistema de sucesos de Laplace.

V.3.- Dadas las ventajas que presenta la descripción de una ley de probabilidades mediante el valor probable y la varianza, en vez de tratar el problema general de la determinación de la ley de la variable suma a partir de las leyes de las variables sumandos, interesa establecer qué relación existe entre los momentos de primer y segundo órden, de la variable suma, y los respectivos de las variables sumandos.

La definición de valor probable de la variable suma es:

(5.1) $$m = \sum_k z_k P(z_k)$$

donde $P(z_k)$ es la probabilidad de que la suma de x e y tome el valor z_k. Esta probabilidad se descompone en la suma de las probabilidades de los sucesos excluyentes constituídos cada uno por un par de valores x_i e y_j tales que $x_i + y_j = z_k$ y por lo tanto (5.1) se puede expresar:

(5.2) $$m = \sum_i \sum_j (x_i + y_j) P(x_i y_j)$$

Aplicando la propiedad distributiva de la suma se tiene:

(5.3) $$m = \sum_i x_i \sum_j P(x_i y_j) + \sum_j y_j \sum_i P(x_i y_j)$$

y como la suma $\sum_j P(x_i y_j) = P(x_i)$ y $\sum_i P(x_i y_j) = P(y_j)$ el resultado final es:

(5.4) $$m = \sum_i x_i P(x_i) + \sum_j y_j P(y_j) = m_1 + m_2$$

o sea que el valor probable de la suma de dos variables aleatorias es igual a la suma de los valores probables de las variables sumandos.

Conviene discutir más detalladamente el método de demostración de este importante teorema, porque se funda en la aplicación de nociones que, como se señaló oportunamente en el párrafo (III.A.1), son típicas del cálculo de probabilidades, como las de dependencia e independencia, y su utilización mediante el método tambien típico de la construcción de un espacio producto.

Representando las variables x e y sobre dos ejes coordenados (Fig.1), un par de valores (x,y) identifica un punto de un reticulado plano, y los puntos en los que $x_i + y_j = z_k$ se encuentran sobre una recta identificada por el valor de z_k. A cada uno de los puntos así determinados, corresponde una probabilidad, que por el teore

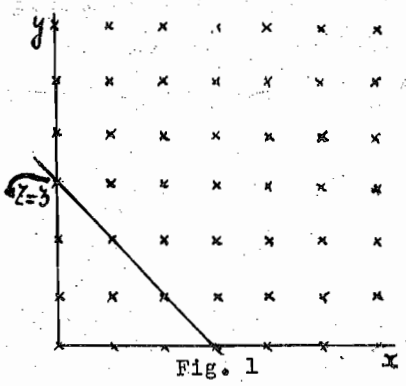


Fig. 1

ma del producto (3.9) será:

(5.5) $P(x_i y_j) = P(x_i)P(y_j/x_i) = P(y_j)P(x_i/y_j)$

La fórmula anterior define en cada uno de los puntos del retículado una probabilidad que se designa como distribución conjunta de x e y. La suma de (5.1) equivale a fijar un valor de z, al que se le multiplica por la suma de las probabilidades de los puntos ubicados sobre la recta oblicua correspondiente a ese valor de z, operación en general difícil. Al aplicar la propiedad distributiva en (5.3), se sigue un orden más racional. En la primera sumatoria se fija primero x_i y se suman las probabilidades de los puntos ubicados sobre la vertical correspondiente, y luego se suma segun x_i. En la segunda se realiza una operación similar, sólo que sobre las horizontales.

Fijado un valor de x_i, los distintos valores de y determinan una familia completa de sucesos excluyentes, tal que la suma de las probabilidades (5.5) es igual a $p(x_i)$, porque fijando i y sumando respecto de j se tiene:

(5.6) $$\sum_j P(x_i y_j) = P(x_i)\sum_j P(y_j/x_i) = P(x_i)$$

dado que la suma de todas las probabilidades condicionales es igual a la unidad. Ya al enunciar el teorema de Bayes, en (3.51) se aplicó el mismo procedimiento para expresar la probabilidad que figura en el denominador.

Es importante señalar que el teorema de adición de los valores probables (5.4), vale tanto para variables aleatorias dependientes como independientes, desde que la propiedad fundamental de que la suma de las probabilidades de (5.5) con respecto a un índice mientras el otro permanece constante es igual a la probabilidad del valor de la variable cuyo índice permanece fijo, tampoco depende de la dependencia o independencia.

V.4.- El teorema de adición de los valores probables es cierto para cualquier número finito de sumandos, lo que se demuestra por inducción completa, utilizando la propiedad asociativa de la suma:

(5.7) $$x^1 + x^2 + \ldots + x^{n-1} + x^n = z^{n-1} + x^n$$

en donde el índice superior indica distintas variables aleatorias, y z^{n-1} es la suma de las n-1 primeras variables aleatorias, y como el teorema es cierto para $n = 2$, es cierto para todo n.

V.5.- Si se multiplica todos los valores de una variable aleatoria por una constante el valor probable resulta multiplicado por la misma constante :

(5.8) $$\sum P(x)kx = k\sum P(x)x$$

y teniendo en cuenta el teorema de adición anteriormente demostrado, resulta conveniente introducir un operador simbólico E, que significa multiplicar cada valor por su probabilidad y sumar, que tendrá la propiedad de ser un operador lineal

(5.9) $$E(ax+by)=aE(x)+bE(y)$$

Si a una variable aleatoria se le añade una constante aditiva, la aplicación del operador E conduce a:

(5.10) $$E(ax+c)=aE(x)+c$$

porque como la constante c aparece en todos los términos de la suma, es factor comun de la suma de todas las probabilidades, que es igual a la unidad.

V.6.- Mediante el uso del operador E es fácil demostrar el teorema de adición de las varianzas, aplicandolo al cuadrado de la suma de dos variables aleatorias, que se suponen centradas:

(5.11) $$E\left[(x+y)^2\right]=E(x^2+y^2+2xy)=E(x^2)+E(y^2)+2E(xy)$$

Si las variables aleatorias son independientes entre sí, el término correspondiente al doble producto se puede evaluar sin dificultades, pues consiste en multiplicar cada uno de los valores posibles del producto por su probabilidad, que en este caso es el producto de las probabilidades (por la hipótesis de independencia), y sumar. En esta suma, se puede sumar primero con respecto a una variable y después con respecto a la otra, como en (5.3), resultando:

(5.12) $$E(xy)=E(x)E(y)$$

pero como las variables son centradas, tanto E(x) como E(y) son nulas, con lo que (5.11) se reduce a:

(5.12) $$E(x+y)^2=E(x^2)+E(y^2)$$

o sea que, en el caso de independencia, las varianzas son aditivas, teorema que se generaliza a cualquier número (finito) de sumandos.

A veces se utiliza el operador D, que simboliza centrar la variable, elevarla al cuadrado, multiplicarla por las probabilidades y sumarla, pero hay que tener en cuenta que este operador, si bien es aditivo en el caso de independencia, no es lineal, porque si una variable aleatoria se multiplica por la constante, al calcular la varianza esta aparece multiplicada por la constante elevada al cuadrado.

V.7.- Como aplicación que conduce a resultados nuevos, se calculará el valor probable y la varianza de una variable aleatoria vinculada a la binomial, pero más complicada.

Sea el caso del paseo aleatorio lineal, en el que una partícula se encuentra, en el origen del tiempo, en el origen de coordenadas, y en cada unidad de tiempo recibe un impulso aleatorio, que la desplaza una unidad de longitud so -

bre una recta, con probabilidad p de hacerlo en el sentido positivo, y q en el sentido negativo.

Después de transcurrido un tiempo n, la partícula se encontrará a una distancia del origen igual a la diferencia entre el número de impulsos positivos y negativos que haya recibido, y es claro que esta distancia será una variable aleatoria, que depende de una ley de probabilidades que se determina partiendo de la binomial.

Si se llama s a la distancia, l al número de impulsos positivos y m al de impulsos negativos, se tendrá:

(5.14)
$$n = l + m$$
$$s = l - m$$

por lo que a cada par de valores de n y de s corresponderá otro par de valores de l y m bien determinado:

(5.15)
$$l = \frac{n+s}{2} \quad ; \quad m = \frac{n-s}{2}$$

Aplicando la ley binomial, la probabilidad de hallarse a la distancia s en el tiempo n será:

(5.16)
$$P_n(s) = \frac{n!}{\left(\frac{n+s}{2}\right)! \left(\frac{n-s}{2}\right)!} p^{\frac{n+s}{2}} q^{\frac{n-s}{2}}$$

Los valores de s estaran comprendidos entre n y $-n$, pero debido a la forma particular de las relaciones (5.15), que obligan a que la suma y la diferencia de n y de s sean pares (porque l y m son necesariamente enteros), si n es par, s puede tomar solamente valores pares, si n es impar, s puede tomar solamente valores impares, por lo que en realidad s puede tomar solamente $n+1$ de los $2n+1$ aparentemente posibles.

El cálculo directo del valor probable y de la varianza en (5.16) sería muy complicado, pero es posible aplicar los teoremas de adición teniendo en cuenta que s es una variable aleatoria suma de otras elementales que pueden tomar cada una de ellas los valores $+1$ y -1 con probabilidades p y q, y son independientes entre sí.

El valor probable y la varianza de cada una de las variables sumandos serán:

(5.17)
$$m = p-q \quad ; \quad \sigma^2 = (1-p+q)^2 p + (-1-p+q)^2 q = 4q^2 p + 4p^2 q = 4pq$$

y los de la variable s:

(5.18)
$$m = n(p-q) \quad ; \quad \sigma^2 = 4npq$$

con lo que podría aplicarse la desigualdad de Tchebicheff para obtener acotaciones en probabilidad del valor absoluto de la diferencia entre $(p-q)$ y s.

V.9.- TEOREMA DE BERNOUILLI GENERALIZADO. El teorema de Bernouilli permitía demostrar la convergencia en probabilidad de las frecuencias relativas, en el caso de la ley binomial, al valor p, dado la forma especial que tenía la varianza de la frecuencia relativa, como una fracción en la que el numerador es constante y en el denominador apa

rece n , con lo que la aplicación directa de la desigualdad de Tchebicheff permitía establecer que la probabilidad de que el valor absoluto de la diferencia entre la frecuencia relativa y p superara una cierta magnitud arbitraria, tenía por límite cero al crecer n indefinidamente.

Teniendo en cuenta que la variable binomial es una suma de variables aleatorias idénticas, con igual campo de existencia e igual ley de probabilidad, el teorema de Bernouilli aparece como caso especial de un teorema más general relativo a la suma de variables aleatorias: La suma promediada de variables aleatorias independientes iguales converge en probabilidad al valor probable de la variable sumando.

La demostración es inmediata. La varianza de la variable suma de n variables iguales es igual a n veces la varianza de la variable sumando. La variable aleatoria promedio resulta de dividir cada sumando por n, y la varianza correspondiente aparecerá dividida por n^2 , con lo que varianza del promedio será la del sumando genérico dividido por n , por lo que la aplicación de la desigualdad de Tchebicheff conduce a una expresión en la que la acotación de la probabilidad tiende a cero al crecer n indefinidamente. Por otra parte, el valor probable del promedio, será igual al valor probable de la suma, o sea n veces el valor probable del sumando, dividido por n , o sea finalmente, el mismo valor probable de los sumandos.

Este teorema es una consecuencia tan directa del de Bernouilli, que en muchos textos no sólo no se lo demuestra, sino que tampoco se lo enuncia.

Sin embargo, conviene aislarlo y demostrarlo, porque la convergencia en probabilidad del promedio de las observaciones independientes de una variable aleatoria a su valor probable, es el primer resultado nuevo, no contenido en la caracterización previa de un fenómeno aleatorio, que se ha obtenido.

Además, establece la forma que ha resultado más fructífera para el descubrimiento de nuevas constantes asintóticas, que como se señaló en la introducción, es una de las finalidades del cálculo de probabilidades.

V.10.- La aplicación del teorema de Bernouilli generalizado al paseo al azar lineal, conduce a obtener en forma sencilla resultados no aparentes cuando se dispone solamente de la ley (5.16).

Tomando un ε igual al valor absoluto $|p-q|$, habrá un n a partir del cual, la suma promediada tendrá el mismo signo que (p-q), con una probabilidad tan cercana a la unidad como se quiera. Como el signo del promedio es el de la suma, y esta a su

vez es la abcisa de la partícula, resulta que después de transcurrido un tiempo n suficientemente grande, la partícula se encontrará en una posición de abcisa con igual signo que (p-q) y probabilidad mayor que $1-\delta$.

Si en el origen del tiempo se encuentra en el origen de coordenadas una población de N partículas, recordando la interpretación dada al teorema de Bernouilli en (IV.14), después de transcurrido un tiempo n suficientemente grande, una fracción $(1-\delta)N$ del total de la población, se encontrará en la semirecta de signo igual al del valor probable (p-q).

La incertidumbre aparentemente total de llegar a alguna conclusión sobre la posible posición de la población de partículas, que resulta de la descripción del mecanismo del paseo al azar lineal, se ha transformado en un enunciado definido, susceptible de ser comprobado en la experiencia.

Este esquema simple, cuando se lo generaliza a tres dimensiones, da las bases para una teoría del conocido fenómeno del movimiento browniano.

V.11.- La extensión natural del teorema de Bernouilli y su generalización, es a la suma de variables aleatorias no idénticas entre sí. El primer resultado en esta dirección es el teorema de Poisson (descubierto en 1832): La probabilidad de que el valor absoluto de la diferencia entre la suma promediada de variables aleatorias independientes entre sí, cada una de las cuales toma unicamente los valores 0 y 1 con probabilidades q_i y p_i, y el promedio de los valores probables, sea superior que un ε positivo y arbitrario, tiende a 0 al crecer n.

Llamando r_i a cada una de las variables, y p al promedio de los valores probables $\sum p_i/n$, la aplicación de la desigualdad de Tchebicheff a la suma promediada es, teniendo en cuenta que $p_i q_i < 1/4$ y suma menor que $n/4$:

$$P\left(\left|\frac{\sum r_i}{n} - \bar{p}\right| > \varepsilon\right) < \frac{\sum p_i q_i}{n^2 \varepsilon^2} < \frac{1}{4n\varepsilon^2} \tag{5.19}$$

de donde resulta que fijado un δ arbitrario, basta tomar un n suficientemente grande para que la probabilidad de (5.19) sea menor que dicho δ.

Este teorema, históricamente importante, es conocido con el nombre de l e y d e l o s g r a n d e s n ú m e r o s.

V.12.- Tchebicheff extendió el resultado a variables aleatorias más generales, demostrando un teorema en el que, como en el de Bernouilli generalizado, no se hace hipótesis sobre el campo de variación de las variables aleatorias: La probabilidad de que el valor absoluto de la diferencia entre la suma promediada de variables aleatorias tales que la varianza de su suma tiene un orden de crecimiento menor que de n^2 (siendo n el número de variables sumandos) y el promedio de los valores probables, supere un ε positivo y arbitrario, tiende a 0 cuando n crece indefinidamente.

Al crecer indefinidamente el número de sumandos, la varianza de la suma tiende a un límite que es la suma de la serie de las varianzas. Como esta serie es de términos positivos (las varianzas son sumas de productos de probabilidades por cuadrados), sólo puede ser convergente o divergente, y el teorema de Tchebicheff establece una condición suficiente, a menudo fácil de verificar, que comprende el caso de convergencia y hasta un cierto tipo de divergencia.

La demostración se funda en que la varianza del promedio es la de la suma dividida por n^2. Aplicando nuevamente la desigualdad del mismo Tchebicheff:

$$P\left(\left|\frac{\sum r_i}{n} - \bar{m}\right| > \varepsilon\right) < \frac{\sum \sigma_i^2}{n^2 \varepsilon^2} \tag{5.20}$$

en donde r_i es la variable sumando genérica, m el promedio $\frac{\sum m_i}{n}$, y σ_i^2 la varianza del sumando. Por la hipótesis establecida, fijado un δ arbitrario, basta elegir un n suficientemente grande para que la probabilidad (5.20) sea menor que dicho δ.

Este teorema comprende a los dos anteriores, como es facil verificar La diferencia entre el enunciado del teorema de Bernouilli generalizado, en el que figura la convergencia en probabilidad, y los de los teoremas de Poisson y Tchebicheff, se debe a que en los dos últimos casos el promedio de los valores esperados no es independiente de n, como ocurre en el primero.

V.13.- Como ejemplo combinado de aplicación del teorema de Bernouilli generalizado y el de Tchebicheff, se tomará un juego de azar, en el que el jugador tiene una probabilidad p de ganar y q de perder. Si se adopta como variable aleatoria +1 para el caso de que el jugador gane, y -1 si pierde, se está en una nueva interpretación

concreta del modelo establecido para el paseo al azar lineal, lo que no es sorprendente, porque como se señaló ya en el cap. II , las nociones elementales en que se funda el cálculo de probabilidades no agotan la descripción concreta con que se identifica en la experiencia a un cierto fenómeno aleatorio, y de cada modelo siempre son posibles diversas interpretaciones.

Si el juego en cuestión es la ruleta, y el jugador juega a color, se tendrá $p = 18/37$ y $q = 19/37$. El valor probable y la varianza son, de acuerdo con (5.17):

(5.21) $$m = p - q = \frac{-1}{37} \quad ; \quad \sigma^2 = 4pq = \frac{1369}{1369}$$

Después de un cierto número elevado de partidas, el promedio, y por lo tanto la suma, tendrá el mismo signo que el valor probable, y el jugador habrá perdido un número de partidas mayor del que ha ganado, conforme a la discusión del paseo al azar lineal.

De acuerdo a las reglas corrientes del juego de ruleta, si el jugador efectúa antes de una jugada una apuesta de magnitud a , percibe como premio en el caso de ganar una suma igual a $2a$. Su ganancia neta será por lo tanto a (hay que restar la apuesta). En el caso de perder, habrá pagado el importe de la apuesta sin recibir compensación, y su ganancia neta será $-a$.

La nueva variable aleatoria ganancia neta, tomará los valores a y $-a$ con probabilidades p y q , y como es fácil verificar, su valor probable y su varianza son:

(5.22) $$m = a(p - q) = \frac{-a}{37} \quad ; \quad \sigma^2 = 4a^2pq = a^2\frac{1368}{1369}$$

Nuevamente, después de transcurrir un número elevado de partidas, la suma promediada, y por lo tanto la suma (que es el resultado financiero del juego par el jugador), tendrá el signo del valor probable si se cumple la hipótesis del teorema de Bernouilli generalizado, o sea que la apuesta permanece constante, y el jugador perderá con una certeza mayor al crecer el número de jugadas. Dada la convergencia del promedio al valor probable que establece el teorema de Bernouilli generalizado, en este caso dicho valor probable tiene el significado de la ganancia media que puede esperar el jugador si repite muchas veces la partida, lo que justifica el nombre de esperanza matemática, más usual que el de valor probable.

Como la seguridad de la pérdida está establecida para un número indefinidamente creciente de partidas, todo lo demostrado no es contradictorio con el hecho de que haya jugadores que tengan una ganancia neta positiva, si se retiran inmediatamente despues de ganar, en vez de permanecer jugando.

Si el jugador utiliza una política de apuestas variable, habría que utilizar el teorema de Tchebicheff en vez del de Bernouilli generalizado, y la conclusión anterior

podría no ser válida, sí precisamente la política de variación de la apuesta es tal que asegure la divergencia de la serie de las varianzas de un orden mayor que el crecimiento de n^2.

Los casinos propietarios de ruletas por lo general se aseguran contra esta posibilidad introduciendo la regla de limitación de la apuesta, con la que se tiene:

(5.23) $$P\left(\left|\frac{\sum r_i}{n} - \bar{m}\right| > \varepsilon\right) < \frac{\sum \sigma_i^2}{\varepsilon^2 n^2} < \frac{K\sigma^2}{\varepsilon^2 n}$$

en donde K es el límite superior de la apuesta, y nuevamente se obtiene que la ganancia neta del jugador es negativa con una certeza tanto mayor cuanto mayor sea el número de partidas.

Si no se introduce ninguna limitación a la apuesta, la cuestión se complica notablemente, y hay que hacer intervenir el capital de los jugadores.

V.13.- El teorema de Bernouilli y los demostrados en este capítulo son los casos más simples de un conjunto de teoremas, conocidos con el nombre de teoremas límites del cálculo de probabilidades, y reciben la designación genérica de l e y e s d e l o s g r a n d e s n ú m e r o s.

Con los ejemplos dados, es posible comprender la afirmación de GNEDENKO y KOLMOGOROFF (10) segun la cual sin los teoremas límites el concepto de probabilidad sería ininteligible, ya que su valor epistemológico está fundado en que en los fenómenos aleatorios "en gran escala" (como en el caso del paseo al azar lineal de una población de partículas) la acción colectiva crea regularidades estrictas, no aleatorias, en el sentido de que pueden llegar a ser arbitrariamente precisas con el crecimiento del número de repeticiones.

C A P I T U L O V I

T E O R E M A D E D E - M O I V R E Y L A P L A C E.

VI.1.- Si bien el teorema de Bernouilli, y su extensión mediante las leyes de los grandes números a las variables aleatorias sumas de otras, permite llegar a conclusiones importantes, es evidente la conveniencia de completar esas conclusiones mediante el estudio de la rapidez de la convergencia a las nuevas constantes asintóticas descubiertas (los momentos de primer orden).

Lo mismo que en el caso de las leyes de los grandes números, el desarrollo del problema comenzó históricamente con el estudio de la ley binomial, para extenderse luego a las sumas de variables aleatorias.

Tanto por ese motivo como porque el caso particular de la ley binomial es muy importante, y permite aplicaciones muy amplias, lo que justifica un tratamiento especial, se seguirá en este curso el mismo orden.

VI.2.- La aplicación de la desigualdad de Tchebicheff permitió en el capítulo IV, estudiar en una forma sencilla el comportamiento de la ley binomial al crecer n, estableciendo que la dispersión era proporcional nada más que a la raíz cuadrada de n, lo que creaba, r e l a t i v a m e n t e a n, una concentración de la probabilidad, que al reemplazar la variable aleatoria original por la frecuencia relativa se convertía en una convergencia al primer momento, de tipo especial.

Pero la acotación superior que establece la desigualdad de Tchebicheff es aplicable solamente a las sumas de las probabilidades correspondientes a valores de la variable aleatoria centrada cuyo valor absoluto es mayor que un cierto número fijado en función de la dispersión.

En cambio, un estudio más preciso requiere expresar en función de n las sumas de probabilidades correspondientes a un intervalo arbitrario de la variable aleatoria centrada, o d e s v í o , como tambien se la llama.

La suma directa de un número arbitrario de términos de la ley (4.1) es un problema difícil, y ya DE MOIVRE, en 1732, halló conveniente reemplazar (4.1) por una expresión asintótica.

VI.3.- TEOREMA DE DE-MOIVRE. Una versión simplificada del teorema de DE MOIVRE se obtiene reemplazando los factoriales que aparecen en (4.1), por su expresión segun la fórmula de Stirling:

(6.1) $$n! = n^n e^{-n} \sqrt{2\pi n}$$

con lo que (4.1) se transforma(a):

(6.2) $$P_n(r) = \frac{n^n e^{-n}\sqrt{2\pi n}}{r^r e^{-r}\sqrt{2\pi n}\,(n-r)^{n-r} e^{-(n-r)}\sqrt{2\pi(n-r)}}\, p^r q^{n-r}$$

Los exponenciales se simplifican directamente, así como dos de las raíces cuadradas de 2 , con lo que queda una expresión en la que figuran las potencias de n, r y n-r, y una raíz, además de las probabilidades:

(6.3) $$P_n(r) = \frac{n^{n+\frac{1}{2}}}{r^{r+\frac{1}{2}}\,(n-r)^{n-r+\frac{1}{2}}}\;\frac{1}{\sqrt{2\pi}}\, p^r q^{n-r}$$

A fin de reunir las potencias del coeficiente con las de las probabilidades, se multiplica y se divide por la raíz cuadrada de npq:

(6.4) $$P_n(r) = \frac{1}{\sqrt{2\pi npq}}\left(\frac{np}{q}\right)^{r+\frac{1}{2}}\left(\frac{nq}{n-r}\right)^{n-r+\frac{1}{2}}$$

Reemplazando las potencias por un exponencial, se tiene:

(6.5) $$P_n(r) = \frac{1}{\sqrt{2\pi npq}}\, e^{-z}$$

$$z = (r+\tfrac{1}{2})\log\left(\frac{r}{rp}\right) + (n-r+\tfrac{1}{2})\log\left(\frac{n-r}{nq}\right)$$

en donde debido al signo negativo de z, se han invertido las fracciones cuyos logaritmos aparecen en la segunda expresión (6.5).

A fin de poder introducir el desarrollo en serie del logaritmo en una forma que haga después simple el pasaje al límite para n indefinidamente creciente, conviene introducir el siguiente cambio de variables:

(6.6) $$t = \frac{r - np}{\sqrt{npq}} \quad \therefore \quad r = np + t\sqrt{npq}$$

cuya significación se analizará después, y con el cual (6.5) se transforma en

(6.7) $$z = (np + t\sqrt{npq} + \tfrac{1}{2})\log\left(1 + \frac{tq}{\sqrt{npq}}\right) \quad (nq - t\sqrt{npq} + \tfrac{1}{2})\log\left(1 - \frac{tp}{\sqrt{npq}}\right)$$

Si se acotan los valores de t, limitando su variación entre t_1 y t_2 ($t_1 < t < t_2$), para un n suficientemente grande, el segundo término de los binomios de los que se toman los logaritmos será menor en valor absoluto que la unidad, y se podrá utilizar el desarrollo:

(6.8) $$\log(1 + x) = x - \tfrac{1}{2}x^2 + R$$

Aplicando a (6.7), después de realizar las operaciones indicadas y prescindiendo de los términos en los que aparece n en el denominador (lo que permite prescindir

---

(a) Si no se introduce un resto, el significado preciso de (6.1) es solamente $\lim_{n\to\infty} \frac{n!}{n^n e^{-n}\sqrt{2\pi n}} = 1$. La diferencia entre el factorial y su expresión aproximada supera todo límite, pero el valor relativo de esta diferencia tiende a cero, cuando n crece indefinidamenmente. Este comportamiento no afecta la validez de la deducción que sigue, porque se trata de evaluar el límite de cocientes de factoriales, y no de aproximar el valor de un factorial.

tambien de los restos del desarrollo en serie, se tiene:

$$z = (np + t\sqrt{npq} + \frac{1}{2})\left(\frac{tq}{\sqrt{npq}} - \frac{1}{2}\frac{t^2 q^2}{npq}\right) + (nq - t\sqrt{npq} + \frac{1}{2})\left(-\frac{tp}{\sqrt{npq}} - \frac{1}{2}\frac{t^2 p^2}{\sqrt{npq}}\right) =$$

(6.9)

$$= t^2(q+p) - \frac{1}{2}t^2(q+p) = \frac{t^2}{2}$$

expresión en realidad válida sólo para el límite correspondiente a n indefinidamente aproximada:

(6.10) $$P_n(r) = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{t^2}{2}} = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(r-np)^2}{2npq}}$$

que es el teorema de DE MOIVRE.

VI.4.- Es conveniente discutir el procedimiento de deducción utilizando una representación gráfica, en la cual los valores de (4.1) se llevan como ordenadas de los puntos cuya abcisa son los valores de la variable aleatoria. Uniendo dichas ordenadas por rectas, se obtienen una poligonal, en la que sólo los vértices tienen el significado de probabilidades.

Al crecer n indefinidamente, la poligonal se extiende, y las ordenadas de los vértices disminuyen (tratandose de un sistema completo de sucesos excluyentes la suma tiene que ser igual a la unidad, y al crecer n crece tambien el número de sumandos) y el límite de cada una de estas ordenadas es nulo (fijado r, el valor absoluto del desvío $x = r - pn$ crece con n, y como la probabilidad de r está acotada superiormente, segun la desigualdad de Tchebicheff, por $1/t^2$, y t crece como la raíz cuadrada de n, de manera que la cota es también variable con n, de tal manera que tiende a cero).

Por lo tanto, si se pasa al límite para un r fijo, el resultado es trivial y no añade nada nuevo. Pero si se efectúa el cambio de variables, y se pasa al límite para un t fijo, entonces se obtiene el resultado nuevo (6.9), que no es incompatible con la conclusión anterior, porque en (6.10) aparece la raíz cuadrada de n en el denominador.

El cambio de variables tienen por consecuencia que en el nuevo diagrama el proceso de achatamiento de la poligonal se realiza aproximandose a la exponencial $\frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}$ conocida con el nombre de l e y n o r m a l, con lo que se tiene una expresion asintótica de (4.1) independiente de los factoriales y de las potencias de las probabilidades.

El cambio de variables (6.6) consiste en el centrado de la variable aleatoria, y su n o r m a c i ó n por la dispersión, acostumbrandose a llamar v a r i a b l e r e d u c i d a a la variable c e n t r a d a y n o r m a d a (y d e s v í o

r e d u c i d o al desvío normado por la dispersión).

VI.5.- FORMULA DE LAPLACE. Como la ley normal está tabulada, el problema de hallar la suma de un número arbitrario de términos de la fórmula (4.1), se reduciría a calcular lo desvíos reducidos t, leer los valores de la normal en la tabla, sumarlos y dividir la suma por el factor $\sqrt{npq}$, lo que es ya un progreso considerable sobre el cálculo de cada uno de los sumandos, haciendo uso de una tabla de logaritmos de factoriales y de una tabla de logaritmos ordinaria para las potencias de las probabilidades.

Pero Laplace encontró, en 1801, un procedimiento aún más sencillo y de considerable importancia para resolver el problema de la suma.

En lugar de considerar una representación gráfica en la que las ordenadas representan las probabilidades, para deducir la fórmula de Laplace hay que utilizar otra representación en la que las probabilidades corresponden a las áreas, por lo que a cada valor de r se le asignará un rectángulo de altura igual a la probabilidad y de espesor unitario.

Al hacer el cambio de variables, como las abcisas se dividen por $\sqrt{npq}$, las ordenadas hay que multiplicarlas por ese mismo número para que el área del diagrama continúe representando la probabilidad, con lo que teniendo en cuenta (6.10), en el límite para n indefinidamente creciente, la probabilidad de que la variable aleatoria reducida se encuentre dentro de un cierto intervalo de t será la superficie subtendida por la exponencial $\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ o sea la integral de la ley normal:

$$P(t_1 < \frac{r-np}{\sqrt{npq}} < t_2) = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{t^2}{2}} dt \tag{6.11}$$

que tambien está tabulada, y con la cual se resuelve sencillamente el problema de hallar la suma de un número arbitrario de términos de (4.1), determinando el intervalo correspondiente de la variable reducida t.

VI.6.- COMPARACION ENTRE LA APROXIMACION DE LA LEY NORMAL Y DE LA DESIGUALDAD DE TCHEBICHEFF. Como el teorema de De Moivre es un teorema de convergencia, en el sentido ordinario del análisis, de la ley de probabilidad a la función normal, es previsible que dé mejores resultados que la desigualdad de Tchebicheff, que se limita a establecer una cota superior.

Dando valores al desvío reducido t, es suficientemente ilustrativa la siguiente tabla:

	P	
	$P(\lvert t\rvert > k)$	
k	Tchebicheff $P(\lvert t\rvert > k) \leq \frac{1}{k^2}$	Normal $P(\lvert t\rvert > k) = 1 - \int_{-k}^{k} e^{-\frac{t^2}{2}} dt$
1	1	0,31732

2	1/4	0,04550
3	1/9	0,00264
4	1/16	0,00006

Para un desvío reducido igual o menor que la unidad, la desigualdad de Tchebicheff en realidad no acota la probabilidad de superarlo, porque cualquier probabilidad no puede superar a la unidad. Pero precisamente en ese intervalo es donde interesa con frecuencia estudiar las fluctuaciones de la variable binomial, porque es el que incluye los valores de mayor probabilidad, y la ley normal resulta insubstituible.

A medida que crece el valor absoluto del desvío reducido, la acotación de la desigualdad de Tchebicheff disminuye, con la inversa del cuadrado, pero la normal lo hace más rapidamente, razón por la cual, las tablas usuales añadidas como apéndice de los libros de probabilidades o de estadística no suelen extenderse más que a $t=3$.

En compensación, la desigualdad de Tchebicheff no sólo es utilizable con comodidad para demostraciones de convergencia, como en el teorema de Bernouilli, sino que también es aplicable a cualquier variable aleatoria discreta y finita, como las estudiadas hasta ahora. La normal, en cambio es directamente aplicable a la ley binomial, y su extensión a otras variables aleatorias es posible solamente en casos bastante amplios, pero con todo limitados, como se verá en el capítulo VIII.

VI.7.- USO DE LAS TABLAS.- CORRECCION DE BERNSTEIN. El uso correcto de una fórmula asintótica, requiere una estimación de los errores que se pueden cometer, o en otros términos, determinar una expresión para el resto, si se quiere emplearla para obtener valores aproximados.

El método empleado en VI.3 para deducir el teorema de De Moivre, y su complemento la fórmula de Laplace, que es en esencia el empleado por De Moivre, no se presta para esa finalidad. Laplace, en realidad, utilizó otro procedimiento, que modificado por USPENSKY permite obtener una fórmula para el resto, que no es de uso cómodo (ver el Cap. VII., sección 11, fórmula 23 de (22)).

SERGE BERNSTEIN halló que para valores pequeños de h, la fórmula (6.11) da valores con excelente aproximación cuando se modifican los límites, añadiendo o restando 1/2 segun que cada límite sea superior que np, o menor que ese valor.

La demostración de Bernstein ha sido revisada con considerable detalle por FELLER (a), a fin de obtener la mejor aproximación de primer orden, pero para las aplicaciones corrientes la correccion de Bernstein es más que suficiente, para precaverse de cometer errores de significación.

Naturalmente, también hay que tener en cuenta ciertas circunstancias que se des -

(a) "On the normal approximation to the Binomial Distribucion", Ann. Math. Stat. vol. 16 (1945), pag. 319.

prenden de una comparación entre la ley normal y el comportamiento de la ley binomial estudiado con detalle en el capítulo IV.

En primer lugar, la ley binomial asigna probabilidades a un número finito de valores de la variable, mientras que la normal es asintótica a $-\infty$ y a $+\infty$, por lo que su integral, igual a la suma de todas las probabilidades, vale 1 en el intervalo $-\infty, +\infty$. Por lo tanto, dado un n, atribuye probabilidades a valores alejados de np a los que en realidad correspondería una probabilidad nula, por lo que (6.10) y (6.11) no seran aplicables, respectivamente, para valores o para intervalos comprendidos entre valores muy alejados de np. Para este caso especial, FELLER (4), aconseja el uso de otras expresiones, algo más complicadas.

En segundo lugar, (6.10) es simétrica, mientras que la ley binomial lo es solamente si $p = q = 1/2$, y como admite su único máximo en la vecindad de np, será tanto más asimétrica cuanto más se aproxime p a cero o a la unidad, porque en ambos casos el máximo se ubicará cerca de un extremo del campo de definición. La convergencia se verifica siempre, pero será mucho más lenta, y la aplicación de (6.10) y (6.11) requerirá valores elevados de u.

VI.8.- Algunos ejemplos numéricos permitiran aclarar la discusión anterior.

Sea por ejemplo, calcular la probabilidad de que para $n = 10$ y $q = p = 1/2$, los valores de la variable se encuentren contenidos en el intervalo $4 \leqslant r \leqslant 6$.

Calculando la variable reducida t, con las correcciones de Bernstein correspondientes (en el primer límite se resta porque es menor que $np = 0,5$ y en el segundo se suma porque es mayor).

$$t_1 = \frac{4-0,5-5}{\sqrt{10.0,5.0,5}} = -0,951 \quad ; \quad t_2 = \frac{6+0,5-5}{\sqrt{10.0,5.0,5}} = 0,951$$

$$P_{10}(4 \leqslant t \leqslant 6) = \frac{1}{\sqrt{2\pi}} \int_{-0,951}^{+0,951} e^{-\frac{t^2}{2}} dt = 0,6579 \tag{6.12}$$

que se compara favorablemente con el valor exacto hasta el mismo orden decimal, que es 0,6563. A pesar de que n es solamente igual a 10, por lo que habría dudas para la aplicación de (6.11), que es válida solamente en el límite para n indefinidamente creciente, utilizando la corrección de Bernstein, la diferencia es inferior a dos milésimos.

Si no se hubiera empleado la corrección de Bernstein el resultado hubiera sido

$$t_1 = \frac{4-5}{\sqrt{10.0,5.0,5}} = -1,265 \quad ; \quad t_2 = 1,265$$

$$P_{10}(4 \leqslant t \leqslant 6) = \frac{1}{\sqrt{2\pi}} \int_{-1,265}^{+1,265} e^{-\frac{t^2}{2}} dt = 0,7957 \tag{6.13}$$

con un error de 0,1384 por exceso con respecto al valor exacto, o sea casi del 25%, en contra de menos del 2% cuando se empleó dicha corrección.

Es evidente que se han obtenido resultados tan satisfactorios a pesar del reducido valor de n, porque se trata del caso más favorable, con $p = q = 1/2$.

Como regla general, no conviene usar (6.11), aún con la corrección de Bernstein, si $p = 0,2$ o menos (y por lo tanto también cuando $p = 0,8$ o más) y n es pequeño, pero es complicado dar reglas fijas, porque la magnitud del error relativo depende del intervalo $(t_1 . t_2)$. En el capítulo VII se considerará este problema de otra manera.

La fórmula (6.11) también permite calcular directamente las ordenadas, en lugar de (6.10) simplemente eligiendo un intervalo adecuado. Sea, por ejemplo, calcular $P_{10}(4)$, cuando $p = q = 1/2$. Tomando un intervalo de t tal que la variable corregida $r - 1/2$ (en este caso la corrección es subtractiva para los dos límites porque ambos son inferiores a $np = 0,5$) comprenda a $r = 4$, se tiene:

$$t_1 = \frac{4-0,5-5}{\sqrt{10.0,5.0,5}} = -0,951 \quad ; \quad t_2 = \frac{5-0,5-5}{\sqrt{10.0,5.0,5}} = -0,316$$

$$P_{10}(4) = \frac{1}{\sqrt{2\pi}} \int_{-0,316}^{-0,951} e^{-\frac{t^2}{2}} dt = 0,2048 \tag{6.14}$$

que se compara muy bien con el valor exacto 0,2051.

La pregunta natural es, a partir de que valor de n es posible prescindir de la corrección de Bernstein. La respuesta depende de la tabla que se use.

En efecto, siento (6.10) una exponencial, la interpolación lineal es muy inexacta, y si la tabla de que se dispone no da las diferencias primeras y segundas para una interpolación parabólica, es preferible limitarse a considerar para t valores de una presición igual a la de la tabla (décimo o centésimo), calculándolo con una cifra más para redondear.

Por lo tanto, la regla será: calcular t sin la corrección, y observar de la relación entre 1/2 y $\sqrt{npq}$ si el uso de la corrección alteraría el redondeo.

Para no tener que emplear tablas excesivamente extensas, y poder realizar cálculos de gran precisión en caso necesario, la mejor disposición es la de la colección (27) en la que t se da al centésimo, y también las diferencias primera y segunda, para llegar al milésimo en caso necesario. Pero generalmente en la práctica basta una tabla, con precisión al décimo de t.

VI.9.- Como ejemplo de las ventajas prácticas del uso de la aproximación, se revisará la determinación del número n de observaciones para aproximar el valor de p con precisión mayor de un centésimo y probabilidad $1-\delta = 0,99$, que en el capítulo IV (párrafo IV, 15) resultó, mediante la aplicación de la desigualdad de Tchebicheff, nada menos que 250.000.

La desigualdad básica se transforma en otra correspondiente al desvío reducido, multiplicando todos los términos por $\sqrt{\frac{n}{pq}}$

$$(6.15)\quad P\left(\left|\frac{r}{n}-p\right|<\varepsilon\right) = P\left(-\varepsilon<\frac{r-np}{n}<\varepsilon\right) = P\left(-\varepsilon\sqrt{\frac{n}{pq}}<\frac{r-npq}{npq}<\varepsilon\sqrt{\frac{n}{pq}}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\varepsilon\sqrt{\frac{n}{pq}}}^{\varepsilon\sqrt{\frac{n}{pq}}} e^{-\frac{t^2}{2}}\,dt$$

El producto $\varepsilon\sqrt{\frac{n}{pq}}$, que figura en los límites, debe ser mayor que el desvío reducido necesario para que la integral valga 0,99. Ya sea utilizando una tabla ordinaria, u otra simplicada para resolver problemas de este tipo, en la que se entra por la probabilidad y se halla el desvío reducido mínimo para que la probabilidad de superarlo sea menor que el argumento (como la tabla II de CRAMER(3)), que son muy cómodas porque evitan interpolaciones, que como se señaló anteriormente deberían hacerse utilizando diferencias segundas, se tiene:

$$(6.16)\qquad t = 2{,}5758 \leqslant \varepsilon\sqrt{\frac{n}{pq}} \quad ; \quad n \geqslant \frac{(2{,}5758)^2}{\varepsilon^2}\,pq$$

Como el producto pq no puede superar un cuarto, el valor de n mínimo no será nunca superior a

$$(6.17)\qquad n = \frac{(2{,}5758)^2}{0{,}0001}\times 0{,}25 \cong 12.750$$

casi veinte veces inferior al calculado anteriormente.

El número es con todo elevado, lo que se debe al pequeño valor de δ. Como la ley normal converge a cero bastante rápidamente, se puede esperar que aumentando algo δ se puede disminuir el desvío reducido de tal manera que aunque la precisión se incremente apreciablemente, no se supere el n determinado mediante la desigualdad de Tchebicheff. Feller (4) tiene un ejemplo con $\varepsilon=0{,}001$ y $\delta=0{,}05$, resultando $n=40.000$.

Cuando se conoce el orden de magnitud de p , por ejemplo que no supera 0,10 ó 0,05 (caso de tasas de mortalidad, o de elementos deficientes en una producción industrial), el producto pq, y con él n disminuyen apreciablemente.

Naturalmente el significado de la fijación de δ necesita una discusión especial, que tiene su lugar propio en la PARTE IV (Inferencia Estadística), pero los ejemplos dados son suficientes para apreciar la ventaja del uso de la normal en lugar de la desigualdad de Tchebicheff.

VI.10.- PROBLEMAS TIPICOS DE USO DE LA APROXIMACION NORMAL. Estos problemas son dos :

(1) Dada la probabilidad p , y el numero n de observaciones, hallar la probabilidad de que un cierto desvío no supere cierto valor (o que lo supere, o que esté comprendido dentro de un cierto intervalo.

La solución de este problema consiste en calcular simplemente el desvío reducido, correspondiente, (puesto que se conocen x y $\sqrt{npq}$), y hallar en la tabla la probabilidad buscada. Sea, por ejemplo, un fenómeno aleatorio con probabilidad $p=q=1/2$ y

$n = 100$, y el problema es hallar la probabilidad de que el desvío supere $+10$ (en el caso de cien tiros con una moneda equilibrada, hallar la probabilidad de obtener más de 60 caras en cien tiros).

Utilizando la tabla I de CRAMER (3), que tiene como argumentos los desvíos con precisión de un décimo, hay que utilizar la corrección de Bernstein (porque 1/2 dividido por $\sqrt{npq} = 5$ es 0,1, y seguramente la corrección altera la entrada en la tabla) en forma aditiva (porque siendo el desvío positivo, se encuentra a la derecha de np):

(6.18) $$t_1 = \frac{10 + 0,5}{5} = 2,1$$

Como la tabla referida está dada en la forma

(6.19) $$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_1} e^{-\frac{t^2}{2}} dt$$

hay que tomar el complemento a la unidad para hallar la probabilidad buscada:

(6.16) $$\frac{1}{\sqrt{2\pi}} \int_{t_1}^{+\infty} e^{-\frac{t^2}{2}} dt = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_1} e^{-\frac{t^2}{2}} dt = 1 - 0,98214 = 0,01786$$

(2) Dado p una cierta probabilidad, y un desvío cual es el número máximo n de observaciones para que el desvío no supere esa probabilidad.

Como la probabilidad está dada, basta entrar por ella en la tabla para hallar el desvío reducido, del cual se despeja n. En un ejemplo similar al anterior, dados $p = \frac{1}{2}$, $x = 20$ y $P = 0,10$ (en el caso de una moneda equilibrada, cual es el número máximo de tiros que pueden realizarse para que la diferencia entre el número de caras observado y el valor probable no supere a 20).

Utilizando la tabla II de CRAMER (3) $t = 1,2816$ (porque esta tabla da la probabilidad de superar el valor absoluto del desvío, y la ley normal es simétrica, por lo que hay que entrar por $P = 0,20$) y se tiene:

(6.17) $$t = \frac{x + 0,5}{npq} = 1,2816 \quad \therefore \quad n \cong \frac{(10,5)^2}{(0,218)^2} = \frac{110,25}{0,0475} = 237$$

La discusión del capítulo anterior sobre el juego a color en la ruleta puede completarse, mediante la aplicación de estos problemas tipo: (a) Fijada la apuesta, el número de jugadas y una cierta pérdida (que es función del desvío), hallar su probabilidad. (b) Fijada la apuesta, la pérdida y una cierta probabilidad, hallar el número máximo de jugadas, para que la pérdida no supere el límite fijado, con la probabilidad dada.

VI.10.- En el capítulo anterior se vió como, al considerar la variable aleatoria binomial como suma de variables aleatorias, se pudo extender el teorema de Bernouilli, lográndose demostrar otros teoremas del mismo tipo y aún mas generales, válidos para su -

ma de variables aleatorias cualesquiera, con restricciones que no se refieren a la forma analítica de la ley de probabilidades, sino solamente a sus momentos de segundo orden.

Parece natural preguntarse si la convergencia de la ley binomial, en el sentido ordinario del análisis, a la ley normal, podría también extenderse a la ley de la suma de variables aleatorias, con restricciones únicamente sobre los momentos y no sobre la forma analítica de las leyes de probabilidades de las variables sumandos.

Pero tambien es evidente que el problema así planteado debe presentar serias dificultades. Sin embargo, y con la introducción de algunos métodos especiales para el estudio de las sumas de variables aleatorias, es posible resolverlo en forma relativamente sencilla, al menos para el caso bastante general, como se hará en el Capítulo VIII.

Es indudable que un teorema límite de este tipo, que permitiría tratar una clase muy amplia de variables aleatorias con los métodos tradicionales empleados para la variable binomial, debe ser muy importante, por lo que POLYA le atribuyó la designación de TEOREMA CENTRAL DEL LIMITE (a), expresión feliz que ha perdurado.

(a) "Über eine Aufgabe der Wahscheinlichkeitrechnung betrefrend die Irfahrt in Strassennetz" Mathematische Annalen, vol 84 (1921) pag. 149.

C A P I T U L O VII.

L E Y D E P O I S S O N - V A R I A B L E S A L E A T O R I A S N U M E R A B L E S

VII.1.- Antes de proseguir con el estudio del problema de la extensión del teorema de De Moivre-Laplace a las sumas de variables aleatorias, es conveniente detenerse en la consideración de un cierto tipo de fenómenos aleatorios, aparentemente muy distintos de los que se han venido tratando hasta ahora en el curso, pero que, aparte de presentar interés de por sí, proveen el método para salvar la dificultad señalada en el párrafo VI.8 de la lentitud de convergencia a la ley normal por parte de la binomial, cuando los valores p son cercanos a cero o a la unidad.

En la física y en la técnica se presenta a menudo el caso de un fenómeno aleatorio, en el que la probabilidad de un cierto suceso es una función monótona creciente del tiempo, o más generalmente, de una variable continua.

Por ejemplo, en la desintegración radioactiva, el suceso consiste en el registro de un cierto intervalo de tiempo, y se divide el período de observación en intervalos de esa longitud, la experiencia muestra: (1)La frecuencia relativa del número de intervalos en los que se registró una emisión, frente al total de intervalos observados , presenta una estabilidad tanto mayor cuanto más largo es el período de observación; (2)Si se aumenta la longitud del intervalo, la frecuencia relativa aumenta. Este comportamiento del fenómeno hace razonable pensar que la probabilidad de la observación de una emisión es función monótona creciente del tiempo.

El fenómeno es en realidad algo más complicado. En un cierto intervalo puede observarse ninguna emisión, una emisión, o bien más de una. Si se toman en cuenta nada más que los intervalos en que se registró una sola emisión, o exactamente dos, etc. aparece la estabilidad de las frecuencias relativas y señalada anteriormente, y resulta conveniente introducir una variable aleatoria que, al igual que la binomial, toma sólo valores enteros positivos, incluyendo el cero. Pero a diferencia de esta última variable, su campo de variación no está acotado, pues dado un n suficientemente grande, no existe ninguna razón para descartar como imposible el suceso correspondiente a la observación de exactamente n emisiones.

Podría pensarse que si la probabilidad de registrar n emisiones en un cierto intervalo es muy pequeña, en la práctica se la puede descartar. Esto se puede aceptar como una regla de trabajo en algunas aplicaciones, pero desde el punto de vista de la formulación de una teoría rigurosa, no es posible, porque desaparecería la regla de que la suma de las probabilidades de un sistema completo de sucesos excluyentes es igual

a la unidad.

Por otra parte, la axiomática formulada en el capítulo II, y los teoremas de adición que son su consecuencia, tienen su validez limitada a un número finito de sucesos. Pero esta última dificultad se obvia aceptando que la operación de suma se puede extender a la serie (para un tratamiento riguroso de esta cuestión, ver KOLMOGOROFF(14). Cap. II).

VII.2.- La deducción de la forma analítica de la ley de probabilidad asociada con esta variable aleatoria numerable, se realiza sin mayores dificultades, partiendo del hecho fundamental de que la probabilidad de un suceso es función monótona creciente del tiempo, y además, de que la probabilidad del registro de un cierto número de emisiones en un cierto intervalo, es independiente de si en el intervalo anterior se registró o no alguna emisión, o varias, propiedad que tambien puede verificarse experimentalmente.

A estas propiedades experimentales se agregan dos hipótesis, que pueden parecer muy restrictivas, pero que en realidad no lo son, si se atiende a que los resultados que se obtienen resultan confirmados en la práctica. Ellas son: (1)La función monótona creciente del tiempo f(t), que representa la probabilidad de la emisión de una partícula en el tiempo t , es nula para $t=0$, es derivable a la derecha para $t=0$, y esta derida es distinta de cero; (2)Si $P(t) = f(t)$ es la referida probabilidad, y Q(t) es la probabilidad de no observar ninguna emisión en el tiempo t, la diferencia $1-(P+Q)$ es un infinitésimo con el tiempo, lo que equivales a aceptar que al disminuir suficientemente t, se puede aceptar que $P+Q=1$.

En un intervalo fijo T , como no se conoce la forma de P(t) y Q(t), no es posible afirmar si el error cometido al aceptar $P+Q=1$ es pequeño o no, pero se puede recurrir al artificio de dividirlo en un número n de subintervalos T. En el límite para n indefinidamente creciente será válida la hipótesis, lo que significa que para n suficientemente grande podrá aplicarse la ley binomial y la probabilidad de que en el tiempo T se hayan observado y emisiones será:

$$P_T(y) = \frac{n!}{y!(n-y)!} (\lambda\Delta T)^y (1-\lambda\Delta T)^{n-y} \tag{7.1}$$

en la que f(t) se ha reemplazaso por su desarrollo segun la fórmula de Mac Laurin, y se ha llamado λ a su derivada en el origen.

Teniendo en cuenta que $\Delta T = \frac{T}{n}$ (7.1) se transforma en:

$$P_T(y) \quad \frac{n!}{y!(n-y)!} \left(\frac{\lambda T}{n}\right)^y (1-\frac{\lambda T}{n})^{n-y} \tag{7.2}$$

Agrupando convenientemente los términos, de manera que primero aparezcan los que son independientes de n , luego el resultado de la simplificación de los factoriales, y luego la potencia en que figura n , se tiene:

(7.3) $$P_T(y) = \frac{(\lambda T)^y}{y!} \cdot \frac{n(n-1)\ \ldots\ (n-y+1)}{n^y} \left(1 - \frac{\lambda T}{n}\right)^{n-y}$$

Esta expresión es solo aproximada, pero haciendo crecer n la aproximación mejora, y al pasar al límite se tendrá una expresión exacta.

La operación del pasaje al límite es sencilla. El primer factor es constante. El segundo tiende a la unidad, y el tercero es cociente de potencias de un binomio, en el que el numerador tiende a $e^{-\lambda T}$ y el denominador a la unidad, con lo que se tiene finalmente:

(7.4) $$\lim_{n\to\infty} P_T(y) = \frac{(\lambda T)^y}{y!}\ e^{-\lambda T} = \frac{\alpha^y}{y!}\, e^{-\alpha}$$

conocida con el nombre de l e y d e P o i s s o n. La suma de la serie de las probabilidades es igual a la unidad :

(7.4) $$\sum_{y=0}^{\infty} \frac{\alpha^y}{y!}\, e^{-\alpha} = e^{-\alpha} \sum \frac{\alpha^y}{y!} = e^{-\alpha} \cdot e^{\alpha} = 1$$

como corresponde por tratarse de la suma de las probabilidades de un sistema completo de sucesos excluyentes.

VII.3.- Los momentos de primer y segundo orden al origen se definen formalmente como en IV.10:

(7.5) $$\mu_1 = \sum yP(y) \quad ; \quad \mu_2 = \sum y^2 P(y)$$

pero en una variable aleatoria numerable ya no es posible afirmar "a priori" que los momentos existen, pues su existencia depende de que la serie correspondiente sea convergente. Afortunadamente, en la ley de Poisson no se presentan dificultades de este tipo.

Pasando al cálculo efectivo, para el primer momento se tiene:

(7.6) $$\mu_1 = \sum y \frac{\alpha^y}{y!}\, e^{-\alpha} = e^{-\alpha} \sum_1^{\infty} y \frac{\alpha^y}{y!} = e^{-\alpha} \cdot \alpha \sum \frac{\alpha^{y-1}}{(y-1)!} = e^{-\alpha} \cdot \alpha \cdot e^{\alpha} = \alpha$$

o sea que el igual al parámetro de la ley: n_1

El momento de segundo orden al origen se calcula de la misma manera, aunque el procedimiento resulta algo mas largo:

(7.7) $$\mu_1 = \sum_0^{\infty} y^2 \frac{\alpha^y}{y!} \cdot e^{-\alpha} = e^{-\alpha} \cdot \alpha \sum_1^{\infty} y \frac{\alpha^{y-1}}{(y-1)!} = e^{-\alpha} \cdot \alpha \left(\sum_1^{\infty} \frac{(y-1)\ \alpha^{y-1}}{(y-1)!} + \sum_1^{\infty} \frac{\alpha^{y-1}}{(y-1)!} \right) = \alpha^2 + \alpha$$

Para calcular la varianza, se emplea la relación:

(7.8) $$\sigma^2 = \mu_2 - m^2$$

demostrada en IV.11 y que es de aplicación general porque en esa demostración no intervino la forma particular de la ley de probabilidad, con lo que se tiene

(7.9) $$\sigma^2 = (\alpha^2 + \alpha) - \alpha^2 = \alpha$$

Con el momento de primer orden, o valor probable, y la varianza, se podría apli -

car la desigualdad de Tchebicheff para una descripción sumaria de la ley, mediante una medida de posición y otra de dispersión (ver IV.12), pero en general no será necesario porque la ley de Poisson ha sido tabulada. La tabla original está agotada, pero se encuentra parcialmente reproducida en algunos libros y en colecciones de tablas usuales como la (27), repetidamente citada.

VII.4.- La primera aplicación de la ley de Poisson para estudiar fenómenos aleatorios distribuidos en el tiempo, la realizó VON BORTKIEWICZ (a) a fines del siglo pasado.

Uno de los ejemplos clásicos contenidos en la publicación de este autor, es el estudio de los accidentes mortales debidos a patadas de caballo en 14 regimientos de caballería alemanes, durante los veinte años comprendidos entre 1875 y 1894. En este caso, el intervalo es el regimiento-año, y descartando los datos de cuatro regimientos, que mostraban una frecuencia notablemente alta con respecto a los otros, se tiene en total 200 observaciones, con 113 accidentes mortales, con la siguiente distribución:

Número de accidentes mortales	Años en que se observó ese número	Total de accidentes mortales
0	109	0
1	65	65
2	22	44
3	3	9
4	1	4
		122

Para encontrar el parámetro de la presunta ley de Poisson, basta recordar que es el primer momento de dicha ley, por lo que aplicando el teorema de Bernouilli generalizado (V.9), la suma promediada de los valores observados de la variable aleatoria será una buena aproximación. En el caso, la suma es 122, que dividida por $n = 200$ da 0,61

Entrando en una tabla de la ley de Poisson por el valor del parámetro, se obtienen las siguientes probabilidades, que multiplicadas por 200 dan el número teórico de años:

Número de accidentes mortales	Probabilidad segun ley de Poisson	Número teórico de años en que debía haberse observado
0	0,5437	108,7
1	0,3317	66,3
2	0,1012	20,2
3	0,0202	4,0
4	0,0042	0,8

Para el cálculo se utilizó la tabla incluida en (27), que da los valores del pa-

(a) "Das Gesetz der kleinen Zahlen", Leipzig, 1898.

rametro con precision de un décimo, por lo que hubo que interpolar. En esta última operación se aprovecho la circunstancia de que las probabilidades estan dadas hata el sexto decimal, para calcular diferencias primeras y segundas y hacer una interpolacion parabólica de las probabilidades hasta la cuarta cifra decimal, pues tampoco en este caso la interpolación lineal es aconsejable.

Intuitivamente, la diferencia entre el número teórico de observaciones y el efectivamente registrado no parece de importancia, lo que resulta justificable si se considera que el mecanismo del fenómeno es precisamente del tipo descripto en el comienzo del capítulo, pues se puede aceptar que la probabilidad de recibir una patada mortal es una función monótona del tiempo de exposición, y por lo tanto debería seguir una ley de Poisson.

Pero sin embargo, no está claro con que criterio se debe considerar que las diferencias son significativas o no. Este asunto se tratará con detalle en la Parte Cuarta (Inferencia Estadistica) al estudiar la teoría de la verificación de hipótesis.

VII.5.- Mas generalmente, la ley de Poisson suele describir muy bien la frecuencia de los accidentes, y algunos otros tipos de fenómenos distribuídos en el tiempo. El mismo Von Bortkiewicz mostró en su pubicación que con su ayuda se lograba una buena representación del fenómeno de los suicidios femeninos, utilizando los datos correspondientes a los casos registrados en ocho estados alemanes.

Una aplicación curiosa es la de W. A. Wallis (a) a las vacantes producidas en la Corte Suprema de los Estados Unidos en un período de 96 años, como muestra el cuadro siguiente:

Número de vacantes en el año	Años observados con ese número de vacantes	Probabilidades segun Poisson	Número teórico de años
0	59	0,6065	58,2
1	27	0,3033	29,1
2	9	0,0758	7,3
3	1	0,0144	1,4

En este caso el parámetro α, determinado de la misma manera que en el ejemplo anterior, es igual a 0,5, por lo que no hay dificultades de interpolación, y para el último número se le ha asignado el total de las probabilidades correspondientes a 4 y más (que en la tabla con precisión hasta el millonésimo, llega al 7) pues para los restantes valores la probabilidad es inferior).

(a) "The Poisson Distribution and the Supreme Court", Journal of the American Statistical Association, vol. 31 (1936), pag. 376.-

VII.6.- La ley de Poisson es de interés no sólo para los actuarios que se ocupan del análisis de las estadísticas de accidentes para calcular las primas de los seguros respectivos. Es de fundamental interés tambien en la física y en la técnica.

Al comienzo se hizo referencia al problema de la desintegración radioactiva. Cramer (2) discute detalladamente un ejemplo de este tipo, y en el cuadro siguiente se puede ver una aplicación a las observaciones realizadas por RUTHERFORD Y GEIGER en el caso del polonio:

Observaciones	Períodos con ese número de observaciones	Número teórico de períodos
0	15	14,3
1	56	57,4
2	106	115,2
3	152	154,1
4	170	154,7
5	122	124,2
6	88	83,1
7	50	47,6
8	17	23,9
9-13	16	17,3

El parámetro correspondiente resulta ser $\alpha=4.0$(determinado con precisión hasta la primera cifra decimal).

En 1908, el ingeniero sueco ERLANG utilizó la ley de Poisson para estudiar un problema técnico de la mayor importancia: el dimensionamiento de los canales de interconexión entre centrales telefónicas. Si se quiere asegurarse que una comunicación entre abonados de dos centrales puede establecerse siempre que se pida, los canales deben tener un número elevado de vías, y resultan muy costosos, al mismo tiempo que su capacidad completa se utiliza rara vez. Parece lógico limitar el número de canales, estableciendo una probabilidad suficientemente pequeña para el caso de que una comunicación no pueda atenderse por falta de vía para establecerla, pero ¿Cual es la ley de probabilidades? Estudiando la estadística de llamadas, Erlang halló que seguía muy bien una ley de Poisson.

Este caso fué el primero de una larga serie de aplicaciones técnicas y económicas que han dado lugar a la llamada teoría de las colas, o tiempos de espera, que forma parte de la investigación operativa.

VII.7.- La ley de Poisson se adapta tambien al caso de que la variable no es el tiem-

po, sino otra cualquiera, pero tambien continua y con respecto a la cual se puede considerar que la probabilidad es una función monótona creciente.

La primera aplicación de este tipo la realizó "Studen" (a) al reticulado de un microscopio, y está transcripta detalladamente en el libro de FELLER (4) que tambien contiene una aplicación a la distribución de impactos durante el bombardeo de Londres. Tanto en este caso, como en el "Student", la variable continua de la que puede interpretarse que la probabilidad es una función monótona, es el área.

VII.8.- Históricamente, la ley de Poisson fué descubierta en relación con problemas muy distintos de los hasta ahora considerados. Se trataba simplemente de estudiar la aproximación de la ley binomial por un procedimiento distinto al de De Moivre, y utilizable cuando p es cercano a cero o a la unidad.

La idea de Poisson, en 1832, fué la de estudiar dicho límite del término genérico cuando $np = cte$. Esta hipótesis varía fundamentalmente el problema de Bernouilli y consiguientemente toda la discusión de los capítulos IV y VI, en los que se supone p constante, y por lo tanto np crece indefinidamente con n.

Pero si la convergencia en este procedimiento es más rápida que con el de De Moivre, para un p pequeño, es posible que haciendo $np = \alpha$ se tenga una expresión aproximada útil, lo que efectivamente ocurre.

Tomando $p = 0{,}01$, $n = 50$, y por lo tanto $\alpha = 0{,}5$, se tiene la siguiente comparación de resultados:

r	Ley binomial	Ley de Poisson
0	0,6050	0,6065
1	0,3050	0,3033
2	0,0756	0,0758
3	0,0122	0,0126
4	0,0015	0,0016
5	0,0001	0,0002
	0,9994	1,0000,

Los valores de r superiores a 5 tendrían segun la ley binomial, una probabilidad inferior a cinco cienmilésimos, por lo que no se los ha tomado en cuenta.

Es posible ver que la ley de Poisson como aproximación de la binomial es notablemente precisa, a pesar de que n no es muy grande.

En cambio, el teorema de De Moivre, y la fórmula de Laplace, darían errores muy grandes, debido a la fuerte asimetría que tiene en este caso la ley binomial. En efecto, esta última presenta su máximo, de acuerdo con la discusión del párrafo IV.3 en el

(a) Pseudónimo del estadistico inglés W. Gosset.-

entero siguiente a $np-q = -0,4$, que en este caso es el cero, y como toma solamente valores positivos, resulta una sucesión decreciente de probabilidades. Pero la ley normal es simétrica en torno a $np = 0,5$ y el error correpondiente a la evaluación de la probabilidad acumulada en cero la normal la atribuiría a valores negativos. El error es tanto más perjudicial cuanto que afecta principalmente al valor de mayor probabilidad. Utilizando una tabla con entrada al décimo de desvío, se tendría:

(7.10) $$P(0) = \frac{0,3123}{0,703} = 0,4442$$

con un error relativo por defecto del 27%.

Utilizando la tabla de áreas para calcular la suma de las probabilidades se tendrá :

(7.11) $$P(0 \leqslant r \leqslant 5) = \frac{1}{\sqrt{2\pi}} \int_{-0,4}^{+\infty} e^{-\frac{t^2}{2}} dt = 0,758$$

y nuevamente el error es considerable, mientras que sumando los términos correspondientes de la ley de Poisson, resulta inferior a un milésimo.

VII.9.- A medida que aumenta el parámetro α la ley de Poisson se hace más simétrica, y sus ordenadas difieren menos de las de la ley normal.

Cuando dada la probabilidad $p \leqslant 0,02$, el número n es tan grande como para que el producto $np = 10$, es posible utilizar la aproximación normal con resultados aceptables para valores en la vecindad de np ($|t| \leqslant 2$)

Debidos a su utilidad para la aproximación de la ley binomial cuando una de las dos probabilidades es muy pequeña, la ley de Poisson fué llamada por VON BORTKIEWICZ "ley de los pequeños números", denominación poco feliz que todavía se encuentra en algunos libros de estadística.

C A P I T U L O V I I I

T E O R E M A C E N T R A L D E L L I M I T E.

VIII.1.- FUNCION GENERATRIZ. La introducción del método adecuado para tratar el problema de la adición de variables aleatorias, se debe a LAPLACE. En el Capítulo V no fué necesario utilizarlo, porque allí se trató nada más que de la relación entre los momentos de la variable aleatoria suma y los de las variables sumandos, pero para estudiar el límite de la ley de probabilidad de la variable aleatoria suma, es necesario determinar esa ley en forma más completa de la que permiten hacerlo los momentos de primer orden.

Segun LAPLACE, dada una sucesión finita o infinita de números (que en particular pueden ser probabilidades):

(8.1) $\quad p_0, p_1, p_2, \ldots\ldots, p_n, \ldots$

se llama FUNCION GENERATRIZ de dicha sucesión a una función de un parámetro arbitrario s real o complejo, tal que su desarrollo en serie de potencias del parámetro tenga por coeficientes los términos de la sucesión (8.1).

Si la sucesión es finita, el desarrollo en serie de potencias del parámetro s será un polinomio, y si la sucesión es ilimitada será una verdadera serie.

En el caso de que la sucesión (8.1) sea la de las probabilidades de la ley binomial, se tendrá un polinomio, cuyo término genérico será:

(8.2) $$\frac{n!}{r!\,(n-r)!}\, p^r q^{n-r} s^r$$

y la función generatriz se obtendrá sumando todos los términos, desde $r = 0$ hasta $r = n$ (recordando que convencionalmente $0! = 1$):

(8.3) $$g(s) = \sum \frac{n!}{r!\,(n-r)!}\, q^{n-r}(ps)^r = (ps + q)^n$$

porque al agrupar los términos de igual potencia r (8.2) resulta ser el término genérico del desarrollo de la potencia n - ésima del binomio $(ps + q)$.

Si la sucesión (8.1) es la de las probabilidades de una ley de Poisson, su función generatriz se obtendrá sumando una serie:

(8.4) $$g(s) = \sum \frac{y}{y!}\, e^{-\alpha} \cdot s^y = e^{-\alpha} \sum \frac{(\alpha s)^y}{y!} = e^{-\alpha} e^{\alpha s} = e^{-\alpha(1-s)}$$

El procedimiento para hallar la función generatriz no presenta, en general, dificultades.

VIII.2.- DETERMINACION DE LOS MOMENTOS MEDIANTE LA FUNCION GENERATRIZ. En los ejemplos anteriores se ha visto como al sumar con respecto a r, ésta desaparece como varia-

ble, y la función generatriz depende solamente del parámetro s.

Cuando $g(s)$ es un polinomio, está definida para todo valor de s, real o complejo, pero cuando $g(s)$ es una serie, podría parecer que la fijación de un recinto de convergencia fuera causa de algunos problemas en su uso. Sin embargo, no ocurre así, porque $g(s)$ es convergente para el valor unitario del parámetro (porque entonces se reduce a la suma de todas de las probabilidades, que siempre vale la unidad), que es precisamente donde interesa.

En efecto, existe una relación sencilla entre los momentos de la variable aleatoria y las derivadas de su función generatriz para el valor $s = 1$.

El momento de primer orden de la variable aleatoria es igual a la derivada de la funnción generatriz en el punto $s = 1$.

Formalmente, la derivada de la función generatriz es:

(8.5) $g'(s) = \sum p_r \, s^{r-1} . r$

y cuando $s = 1$, (8.5) se transforma en:

(8.6) $g'(1) = \sum p_r . r$

que es precisamente la definición del momento de primer orden, como se estableció en el Capítulo IV.

Si la función generatriz es un polinomio, la existencia de la derivada no plantea ningún problema. Si es una serie, tampoco presentará, en general, ningun problema, porque las series de potencias son indefinidamente derivables en el interior de su recinto de convergencia, y de la expresión de la suma de la serie podrá establecerse si si el punto $s = 1$ está en la frontera.

Para la variable aleatoria binomial, la función generatriz de su ley de probabilidad dada en (8.3) es inmediatamente derivable, y dando a s el valor unitario en la expresión de esa derivada, se vuelve a tener el momento ya conocido:

(8.7) $g'(s) = n.(ps + q)^{n-1}$; $g'(1) = np$

En el caso de la ley de Poisson, la función generatriz (8.4) está definida para todo valor de s, y la derivada se calcula también fácilmente:

(8.8) $g'(s) = \alpha \, e^{-\alpha(1-s)}$; $g'(1)$

resultado ya obtenido en el Capítulo VII.

Para obtener los momentos de orden superior, los cálculos se complican un tanto. Considerando solamente el de segundo orden, se tiene que formalmente la derivada segunda de la función generatriz es :

(8.9) $g''(s) = \sum p_r s^{r-2} .(n-1).n$

y haciendo $s = 1$:

(8.10) $g''(1) = \sum p_r .(r-1)r = \sum p_r \, r^2 - \sum p_r \, r$

por lo que para obtener el momento de segundo orden al origen, habrá que sumar a la derivada segunda en el punto $s = 1$, el momento de primer orden, o sea la derivada primera en el punto $s = 1$:

(8.11) $\mu_2 = g''(1) + g'(1)$

Para la ley binomial se tiene:

(8.12) $g''(s) = np.(n - 1)p.(ps + q)^{n - 2}$

y aplicando (8.11):

(8.13) $\mu_2 = g''(1) + g'(1) = n^2p^2 - np^2 + np = n^2p^2 + np(1 - p) = n^2p^2 + npq$

que es la expresión conocida del segundo momento al origen.

En el caso de la ley de Poisson:

(8.14) $g''(s) = \alpha^2 e^{-\alpha(1 - s)}$

y aplicando (8.11):

(8.15) $\mu_2 = g''(1) + g'(1) = \alpha^2 + \alpha$

como se sabía.

Para momentos de orden superior al segundo las correcciones son más complicadas, Por ejemplo, para :

(8.16) $g'''(s) = \sum p_r s^{r - 3}.r.(r - 1).(r - 2) = \sum p_r r^3 + \sum p_r r^2 - 2\sum p_r r$

y la fórmula a aplicar sería:

(8.17) $\mu_2 = g'''(1) + g''(1) - g'(1)$

A pesar de la necesidad de efectuar correcciones para los momentos de orden superior al primero, es evidente que las funciones generatrices proveen un método eficiente para la determinación de los momentos, que tiene la ventaja de ser aplicable con generalidad, en lugar de tener que aplicar artificios para el cálculo de los momentos, como ocurrió en los capítulos IV y VII, con la ley binomial y la de Poisson, respectivamente.

VIII.3.- FUNCION GENERATRIZ DE LA SUMA DE DOS VARIABLES ALEATORIAS. Sin embargo, la aplicación más importante de las funciones generatrices no es el cálculo de los momentos sino el dar un procedimiento para la determinación completa de la ley de probabilidad de la suma de dos variables aleatorias, cuando se conocen las leyes de las variables sumandos.

Sea, por ejemplo, la suma de dos variables aleatorias cualesquiera, en la que la ley de probabilidad de la primera está dada por la sucesión $(p_0, p_1, \ldots, p_n, \ldots)$, y la de la segunda por la sucesión $(q_0, q_1, \ldots, q_n, \ldots)$. Sus funciones generatrices serán:

(8.18) $g_1(s) = \sum p_r s^r$; $g_2(s) = \sum q_r s^r$

Si se forma el producto de dichas funciones generatrices, el coeficiente de su término genérico de potencia r será:

(8.19) $(p_0 q_r + p_1 q_{r-1} + \ldots + p_{r-1} q_1 + p_r q_0)$

o sea igual a la probabilidad de que la variable suma tome el valor r, si los sumandos son independientes entre sí, con lo que se ha demostrado el importante TEOREMA :

La función generatriz de la variable aleatoria suma de dos variables independientes, es igual al producto de las funciones generatrices de las variables sumandos.

VIII.4.- Aplicando el teorema que se acaba de demostrar a la suma de dos variables aleatorias binomiales independientes entre sí, con la misma probabilidad fundamental p y órdenes n_1 y n_2, se tiene:

$$g(s) = g_1(s).g_2(s) = (ps + q)^{n_1} .(ps + q)^{n_2} = (ps + q)^{n_1 + n_2} \tag{8.20}$$

La función generatriz de la suma tiene la misma expresión que la función generatriz de una variable aleatoria binomial, con la misma probabilidad fundamental p y orden $n_1 + n_2$, por lo que si se pudiera probar que la transformación de una ley de probabilidad que le hace corresponder una función generatriz es biunívoca, quedaría demostrado que la adición de dos variables aleatorias binomiales con la misma probabilidad fundamental p tiene por resultado una nueva variable aleatoria binomial con la misma probabilidad fundamental p y un orden igual a la suma de los órdenes.

En este caso particular, la demostración del caracter biunívoco de la transformación no ofrece dificultades, debido a que las variables aleatorias son finitas. Dar una función generatriz es lo mismo que dar la sucesión de los momentos, y se vió en el Capítulo IV que si se conocen los momentos desde el de orden cero hasta el de orden n, la ley de probabilidad queda determinada sin ambigüedad, como el conjunto que valores que es solución única de un sistema de ecuaciones lineales compatibles. Como la función generatriz de una ley binomial es un polinomio de grado igual al orden de la variable, existen y son distintas de cero las derivadas hasta el orden en cuestión, y queda por lo tanto demostrado que la variable suma es también una variable binomial cuyo orden se deduce del de los sumandos.

Este resultado parece muy elemental. Sin embargo, si se intenta calcular efectivamente (8.19) para probarlo directamente, se verá que no es nada fácil, y que el uso de las funciones generatrices es practicamente irreemplazable para poder hallar la ley de probabilidad de la variable suma.

Como comprobación, se deducirá la forma de la función generatriz de la variable aleatoria binomial de orden n, partiendo de las funciones generatrices de las variables elementales que toman los valores 0 y 1 con las probabilidades constantes p y q.

La función generatriz de la variable sumando es:

(8.21) $g(s) = q.s^0 + p.s^1 = (ps + q)$

Como el teorema demostrado en el párrafo anterior se puede genralizar a cualquier número (finito) de sumandos, y todos los sumandos son iguales entre sí, la función generatriz de la suma de n variables elementales iguales entre sí, e independientes, será la potencia n-ésima de la función generatriz de la variable elemental, con lo que se vuelve a encontrar (8.3).

VIII.5.- Si se aplica el teorema de que la función generatriz de la suma de dos variables aleatorias independientes es el producto de las funciones generatrices de los sumandos al caso de dos variables aleatorias que siguen una ley de Poisson, la cuestión se complica porque la variable aleatoria suma no es finita, sino numerable.

La función generatriz de la suma será:

(8.22) $g(s) = g_1(s).g_2(s) = e^{-\alpha_1(1-s)}.e^{-\alpha_2(1-s)} = e^{-(\alpha_1+\alpha_2)(1-s)}$

o sea que tiene la forma de una función generatriz idéntica a la de la variable aleatoria que sigue la ley de Poisson, con un parámetro α igual a la suma de los parámetros de las variables sumandos.

Pero probar el caracter biunívoco de la transformación no es, en este caso, elemental, porque siendo la variable suma, como las sumandos, numerable, se tendría que para determinarla por medio de los momentos habría que resolver un sistema de infinitas ecuaciones lineales con infinitas incógnitas, y probar que la solución es única. El problema así planteado excede los límites de este curso, pero en cambio se lo podrá resolver más adelante en una forma que no requiere considerar el sistema de infinitas ecuaciones con infinitas incógnitas, y con la cual se probará que efectivamente se puede afirmar que la suma de dos variables aleatorias que siguen la ley de Poisson, es otra variable aleatoria que sigue también la ley de Poisson, con un parámetro igual a la suma de los parámetros de las variables sumandos.

VIII.6.- FUNCION CARACTERISTICA. Tanto para resolver en forma general el problema de la biunivocidad de la transformación de la ley de probabilidad a la función generatriz, como para calcular en forma más sencilla el cálculo de los momentos de orden superior al primero, y poder aplicar los mismos métodos a variables aleatorias que toman valores no enteros (por ejemplo, la variable frecuencia relativa introducida en el Capítulo IV, o simplemente una variable binomial centrada, o reducida) conviene especializar el campo de variación de los valores del parámetro s ,sobre el que hasta ahora no se había hecho ninguna hipótesis, fuera de considerarlo limitado, en caso necesario al recinto de convergencia de la serie.

Si se limita el campo de variación de s a los valores complejos con módulo uni-

tario s e^{iu}, la función generatriz se transforma en:

(8.23) $\varphi(u) = \sum e^{iru} p_r$

que es llamada f u n c i ó n c a r a c t e r í s t i c a de la variable aleatoria.

LAPLACE utilizó en algunos problemas esta especialización del campo de variación de s , pero sin darle un nombre especial a (8.23), y PAUL LEVY reivindica para Cauchy, posiblemente con razón, el título de descubridor de la función característica en el cálculo de probabilidades ( ).

VIII.7.- CALCULO DE LOS MOMENTOS POR MEDIO DE LA FUNCION CARACTERISTICA. Si se calcula la derivada n-ésima de (8.23):

(8.24) $\varphi^{(n)}(u) = \sum e^{iru} p_r . (ir)^n$

se deduce inmediatamente que:

(8.25) $\mu_n = \dfrac{\varphi^{(n)}(0)}{i^n}$

o sea que e l m o m e n t o d e o r d e n n - é s i m o d e l a v a r i a b l e a l e a t o r i a e s i g u a l a l a d e r i v a d a n - é s i m a d e s u f u n c i ó n c a r a c t e r í s t i c a , e n e l p u n t o c e r o y d i v i d i d a p o r l a p o t e n c i a n - é s i m a d e l a u n i d a d i m a g i n a r i a , sin correcciones aditivas o substractivas, como ocurría en el caso de la función generatriz sin especializar el valor del parámetro.

Para la variable aleatoria binomial, la función característica es, como se deduce inmediatamente:

(8.26) $\varphi(u) = (pe^{iu} + q)^n$

y los momentos de primer y segundo orden al origen resultan:

(8.27) $\varphi'(u) = n(pe^{iu} + q)^{n-1} . ipe^{iu}$; $m = \dfrac{\varphi'(0)}{i} = np$

(8.28) $\varphi''(u) = n(n-1)(pe^{iu} + q)^{n-2} . i^2p^2e^{2iu} + n(pe^{iu} + q)^{n-1} . i^2pe^{iu}$

$\dfrac{\varphi''(0)}{i^2} = n^2p^2 - np^2 + np = n^2p^2 + np(1-p) = n^2p^2 + npq$

En la ley de Poisson, la función característica es:

(8.29) $\varphi(u) = e^{-\alpha(1 - e^{iu})}$

y los momentos de primer y segundo orden al origen se calculan de igual manera:

(8.30) $\varphi'(u) = i\alpha e^{iu} . e^{-\alpha(1 - e^{iu})}$; $m = \dfrac{\varphi'(0)}{i} = \alpha$

(8.31) $\varphi''(u) = i^2\alpha e^{iu} . e^{-\alpha(1 - e^{iu})} + i^2\alpha^2 e^{2iu} . e^{-(1 - e^{iu})}$

$\mu_2 = \dfrac{\varphi''(0)}{i^2} = \alpha^2 + \alpha$

VIII.8.- CAMBIOS DE VARIABLES EN LA FUNCION CARACTERISTICA CORRESPONDIENTES A TRANSFORMACIONES DE LA VARIABLE ALEATORIA. Los ejemplos dados en el párrafo anterior pueden ser poco convincentes acerca de las ventajas que presenta el uso de la función característica con respecto a la función generatriz, pues si bien se han evitado correcciones aditivas o substractivas en el cálculo de los momentos de orden superior al primero, ello ha sido a costa de alargar la operación de derivación con la introducción de exponenciales que las complican.

Pero aparte de otras ventajas que se veran en el párrafo siguiente, también en el cálculo de momentos las funciones características son más cómodas de usar en general que las generatrices para el cálculo de momentos, pues las transformaciones de centrado y reducido de las variables aleatorias, de uso constante en el cálculo de probabilidades, tienen su correspondencia en cambios de variables, con los cuales se determinan las nuevas funciones características mediante las que se pueden calcular los momentos de las variables transformadas. En cambio, al centrar y reducir una variable aleatoria sus valores dejan de ser enteros, y la función generatriz no es directamente aplicable.

En el caso del centrado de la variable aleatoria, la nueva función característica será por definición:

(8.32) $\varphi^*(w) = \sum p_r e^{i(r - m)w}$

y resulta inmediatamente que la nueva función característica se deduce de la de la variable no centrada, simplemente multiplicandola por un factor igual a e^{-imu}:

(8.33) $\varphi^*(u) = e^{-ium}\,\varphi(u) = \sum p_r e^{i(r - m)u}$

Para la variable aleatoria binomial, se tiene:

(8.34) $\varphi^*(u) = e^{-inpu}(pe^{iu} + q)^n = (pe^{iqu} + qe^{-pu})^n$

y por derivación se obtiene que el primer momento centrado es nulo, y la varianza igual a $\sigma^2 = npq$.

A su vez, la función característica de la variable aleatoria reducida se deduce de la de la variable centrada mediante el sencillo cambio de variables $\sigma u = z$. En efecto, se tiene que $u = \frac{z}{\sigma}$, de donde:

(8.35) $\varphi^*(u) = \sum p_r e^{i(r-m)u} = \sum p_r e^{i(r - m)\frac{z}{\sigma}} = \varphi^{**}(z)$

Este cambio de variables tiene importancia fundamental en la demostración del teorema central del límite.

Cabe hacer notar que la práctica de indicar con un asterisco la función característica de la variable centrada, y con dos y un cambio en la denominación de la variable para la función característica de la variable reducida, no es usual en el cálculo de probabilidades. Se la ha introducido y se la utilizará en este capítulo para evitar confusiones por falta de familiaridad, pero más adelante se prescindirá de ella.

VIII. 9.- TEOREMA DE INVERSION DE LAS FUNCIONES CARACTERISTICAS. Otra ventaja considerable que presenta el uso de las funciones características es que existe una fórmula de inversión de las mismas, por medio del cual es posible pasar de la función característica a la ley de probabilidad.

En efecto, si se multiplica la función característica por un factor exponencial de la forma e^{iuv}, en el cual u es el parámetro o variable de la función característica y v un nuevo parámetro, y al producto se lo integra entre $-\pi, +\pi$, dividiendo la integral por $\frac{1}{2\pi}$:

$$(8.36) \qquad f(v) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \varphi(u) . e^{-iuv} du$$

la nueva función de v que se obtiene tiene la importante propiedad de que para valores de v iguales a los que adopta la variable aleatoria, f(v) es igual a la probabilidad, y para valores de v que no coinciden con ninguno de la variable aleatoria, f(v) es nula, que es el teorema de inversión.

La demostración es completamente elemental, pues se funda únicamente en las propiedades conocidas de las funciones trigonométricas. En efecto, reemplazando en (8.36) la función característica por su expresión, y aplicando la propiedad distributiva de la integración respecto de la suma (en el caso de que la función característica esté definida por medio de una serie, hay que probar previamente su convergencia uniforme), se tiene:

$$(8.37) \qquad f(v) = \frac{1}{2} \sum p_r \int_{-\pi}^{+\pi} e^{i(r - v)u} du$$

y la integral es igual a:

$$(8.38) \qquad \int_{-\pi}^{+\pi} e^{i(r - v)u} du = \int_{-\pi}^{+\pi} \left[\cos (r - v)u + i \operatorname{sen} (r - v)u\right] du$$

Si v es igual a un cierto r de la variable aleatoria, en el sumando correspondiente de (8.37) se verificará que el seno es nulo, y el coseno será igual a uno, con lo que la integral será igual a 2π. En los otros términos, los argumentos del seno y del coseno no seran nulos, salvo en un punto del intervalo de integración, pero sus integrales serán nulas, porque las integrales del seno y del coseno son nulas en el intervalo $-\pi, +\pi$. Sólo será distinto de cero el término en que $v = r$, pero como está multiplicado por p_r y dividido por $\frac{1}{2\pi}$, en definitiva se verifica que en f(v) en ese punto es igual a p_r.

Si v no es igual a ningun valor de la variable aleatoria, todos los términos de la suma de (8.37) seran nulos, con lo que queda probado el teorema de inversión.

Este teorema permite resolver en general el problema de la biunivocidad de la transformación de la ley de probabilidad a la función generatriz, que había quedado pendiente al discutir anteriormente la suma de dos variables que seguían la ley de Poisson. Cons-

truyendo las funciones características de los sumandos, la función característica de la suma tendrá también la forma correspondiente a una ley de Poisson, y como un procedimiento inequívoco para pasar de la función característica a la ley de probabilidad, una función característica corresponde a una ley de probabilidad y sólo a una, con lo queda probado que la suma de dos variables aleatorias que siguen la ley de Poisson obedece también a otra ley de Poisson cuyo parámetro es la suma de los parámetros.

Debe hacerse notar que en realidad no hubiera sido necesario pasar por la función característica para demostrar esta propiedad aditiva de la ley de Poisson, porque al probar el caracter biunívoco de la correspondencia entre leyes de probabilidad y funciones características, quedó también probada la biunivocidad entre leyes de probabilidad y funciones generatrices. Dada una función generatriz, queda determinada una función característica, simplemente restringiendo el parámetro a tomar unicamente valores de módulo unitario, y a una función característica le corresponde solamente una función generatriz, porque si esto no fuera cierto, se tendría que habría dos funciones generatrices que coincidirían en todos los puntos de un arco de curva (el círculo de radio unitario del plano complejo, que corresponde al campo de definición de la función característica) pero dos funciones analíticas que coinciden en un arco de curva, coinciden en todo su campo de existencia, y las funciones generatrices son analíticas por ser series de potencias.

VIII.10.- PLANTEO DEL PROBLEMA CENTRAL DEL LIMITE. Conforme se señaló al final del capítulo VI, este problema consiste en estudiar bajo qué condiciones la ley de probabilidades de una suma de variables aleatorias tiende, en el límite para un número de sumandos indefinidamente creciente, a la ley normal.

El estudio directo de la cuestión es practicamente imposible. Pero mediante el uso de las funciones generatrices (o características), cuyas propiedades más importantes se acaban de estudiar, es posible hallar una solución relativamente simple.

En efecto, en lugar de permanecer en el campo de las leyes de probabilidad, en el que la suma no es una operación definida en tal forma que sea posible hallar procedimientos generales para deducir la forma analítica de la ley de la variable suma partiendo de las de las variables sumandos, y mucho menos para efectuar un pasaje al límite, se recurre al artificio de pasar al campo de las funciones características o al de las generatrices, en el que en cambio a la suma de variables aleatorias corresponde una operación sencilla que es el producto. Una vez establecidas las condiciones para que el producto de las transformadas tienda a un límite que no dependa de una especificación detallada de las leyes de probabilidad, se vuelve al campo de estas últimas.

Este procedimiento es un caso particular del empleo de métodos similares, muy común

en el análisis matemático, especialmente para la solución de ecuaciones diferenciales, y conocidos comunmente bajo la denominación de cálculo operacional. El método empleado para demostrar el teorema central del límite, es un ejemplo del cálculo operacional de Fourier (a).

Por lo tanto, el procedimiento de demostración del teorema central del límite consiste en las siguientes tres etapas:

(1) Estudio del límite del producto infinito de funciones características (debido a la necesidad de la inversión, conviene emplear funciones características y no generatrices) el que se realizará demostrando un teorema de Laplace;

(2) Demostración que la correspondencia entre funciones características y leyes de probabilidad se mantiene en el límite (teorema de continuidad). Solamente se enunciaran las condiciones, dejando la demostración para un apéndice;

(3) Inversión efectiva del límite de las funciones características, probando que corresponde a la ley normal.

En los puntos (1) y (3) se seguirá a VON MISES (21), mientras que para el teorema de continuidad (en el apéndice), se utilizará el método de FELLER (4).

VIII.- 11. TEOREMA DE LAPLACE PARA EL LIMITE DEL PRODUCTO DE UNA FAMILIA INFINITA DE FUNCIONES. Este teorema es válido para funciones cualesquiera, que cumplan unicamente la condición de tener un máximo en un punto $x=a$, comun para todas, además de las derivadas primeras y segundas en ese mismo punto, y asegura, si se cumplen ciertas hipótesis que se enunciaran en el curso de la demostración, que el límite del producto infinito de las funciones es la exponencial $e^{-\frac{t^2}{2}}$.

Mediante un cambio de variables, es posible hacer que el máximo común para todas en el punto $x=a$ sea la unidad, con lo que se tiene definida la familia de la siguiente manera:

$$(8.38) \qquad f_i(a) = 1 \quad ; \quad f'(a) = 0 \quad ; \quad f''(a) = -k_i^2$$

Si se considera la familia de los logaritmos $F_i(x) = \log f(x)$, evidentemente esta familia cumple también con la condición de tener un máximo común en el punto $x = a$ (máximo que será igual a cero), por lo que las derivadas primeras serán también nulas, y las derivadas segundas, debido a que $f_i(a)=1$, seran también negativas, y además iguales a $-k_i^2$:

$$(8.39) \qquad F_i'(a) = \frac{f'(a)}{f(a)} = 0 \quad ; \quad F_i''(a) = \frac{f''(a)f^2(a) - f'(a)f'(a)}{f^2(a)} = f''(a) = -k_i^2$$

Desarrollando $F_i(x)$ por la fórmula de Mac Laurin:

(a) Para una información sobre los métodos de cálculo operacional, en general, ver el artículo de TRICOMI, en Annales de l'Institut Henry Poincaré, vol.8 (1938), pag.111.
El cálculo operacional de Fourier es tratado en una gran cantidad de textos especializados, como los de BOCHNER, TITCHMARSH, CARSLAW y WIENER, pero sin prestar especial atención a su uso en el cálculo de probabilidades.

(8.40) $$F_i(u) = \frac{(u-a)^2}{2} k_i^2 + R_i$$

donde R_i es el resto (que para u suficientemente próximo de a es proporcional al producto de la tercera potencia del incremento por la derivada tercera, en un punto intermedio).

El logaritmo del producto de n funciones de la familia $f_i(u)$ será la suma de los logaritmos correspondientes, cada uno de loscuales es igual a (8.40), obteniéndose:

(8.41) $$\log \prod_{i=1}^{n} f_i(u) = \sum_{i=1}^{n} F_i(u) = -\frac{(u-a)}{2} s_n^2 + \sum_{i=1}^{n} R_i$$

expresión en la que se ha reemplazado a la suma $\sum_{i=1}^{n} k_i^2$ por s_n^2.

Si se introducen las hipótesis auxiliares de que las derivadas segundas en el punto $u = a$ tienen una cota inferior K distinta de cero, y que en un cierto intervalo en torno de a las derivadas terceras estan superiormente acotadas, en forma uniforme por una constante P, mediante el cambio de variables $w = (x - a)s_n$ es posible establecer la siguiente acotación, válida en un intervalo que contiene a a:

(8.42) $$\left| \log \prod_{i=1}^{n} f_i(w) + \frac{w^2}{2} \right| \leqslant \left| \frac{(w-a)^3 P}{3!\sqrt{n}\, K} \right|$$

Pasando al límite para n indefinidamente creciente, se tiene que si w se mantiene dentro de un intervalo de longitud fija:

(8.43) $$\lim_{n\to\infty} \log \prod_{i=1}^{n} f_i(w) = -\frac{w^2}{2} \therefore \lim_{n\to\infty} \prod_{i=1}^{n} f_i(w) = e^{-\frac{w^2}{2}}$$

o sea que el límite del producto infinito de una familia de funciones que cumple con las condiciones (8.38) y las hipótesis adicionales sobre la acotación inferior de las derivadas terceras en un entorno de a, es la exponencial $e^{-\frac{w^2}{2}}$, como se quería demostrar.

VIII.12.- APLICACION DEL TEOREMA DE LAPLACE A LAS FUNCIONES GENERATRICES Y CARACTERISTICAS. Dada una familia de funciones características, es evidente que cumplen siempre con la primera y con la tercera de las condiciones de (8.38), en el punto a 0, porque su valor en cero es la suma de las probabilidades, o sea la unidad, y su derivada segunda es la varianza multiplicada por el cuadrado de la unidad imaginaria, verificandose que la varianza (suma de productos de cuadrados por probabilidades) es siempre positiva. La segunda condición (nulidad de la primera derivada), se cumplirá, en cambio, solamente para las funciones características de una clase especial de variables aleatorias, que tienen nulo su primer momento, o sea en las v a r i a b l e s a l e a t o r i a s. c e n t r a d a s.

El cambio de variables introducido en el curso de la demostración, equivale, conforme a lo visto en el párrafo respectivo (VIII.8), a la normación de las variables aleato-

rias por la dispersión de la suma.

Por lo tanto, el teorema de LAPLACE, expresado en términos de funciones caracte - rísticas, sería el siguiente: la función característica de la suma de n variables aleatorias centradas, y normadas por la dispersión de la suma, tiende al límite $e^{-\frac{w^2}{2}}$ cuando el número de sumandos crece indefinidamente, si se cumplen ciertas condiciones de acotación de sus momentos de segundo y de tercer orden.

Estas condiciones de acotación de los momentos equivalen a asegurar que las variables aleatorias son del mismo orden de crecimiento, en probabilidad, es decir, que la probabilidad de que el cociente de dos valores observados sea cercano a cero tiende a cero, ya que la acotación de las probabilidades de una de ellas no puede ser infinitamente pequeña con respecto a la otra de otra, por la acotación inferior de las varianzas. Tanto en el teorema de DE MOIVRE, ya discutido en el capítulo VI, como en el que se acaba de demostrar, la normación de las variables aleatorias por la dispersión de la suma es esencial para la existencia del límite, porque sin ella no sería posible probar que el segundo miembro de (8.42) tiende a cero para n indefinidamente creciente. En efecto, la hipótesis de la acotación inferior de las varianzas implica la divergencia de la serie cuyas sumas parciales son las s_n^2, y este hecho, aplicado directamente al primer término del segundo miembro de (8.41), implica que su valor disminuye más allá de todo límite al aumentar n. Al efectuar el cambio de variables, la sucesión de las parábolas con máximo en $u=0$ que representan a dicho término, es modificada, convirtiendose en una constante, y el orden de divergencia de la serie de las varianzas resulta ser suficiente para anular la contribución del resto, como se muestra en (8.42). Solo la normación hace desaparecer la divergencia del primer término del segundo miembro de (8.41), mientras que asegura la anulación, en el límite del resto.

VIII.13. INVERSION DEL LIMITE DEL PRODUCTO DE LAS FUNCIONES CARACTERISTICAS. Si se admite que cuanda la función límite hallada es continua en el origen, es tambien una función característica de una cierta variable aleatoria, problema este último de solucion dificil pero del que se da en el apéndice al presente capítulo una demostración relativamente elemental (teorema de continuidad), con la aplicación de la fórmula de inversión (8.37) será posible establecer la ley de probabilidad que sigue la variable aleatoria límite de la suma cuando el número de sumandos crece indefinidamente.

La fórmula de inversión es:

$$f(v) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \varphi(u)\, e^{-iuv}\, du \tag{8.44}$$

en donde la función caracteristica es la de la variable no centrada.

Si se trata de aplicarla a una función característica de una variable aleatoria centrada, será necesario multiplicar dicha función característica por un factor exponencial, como se explicó detalladamente en el párrafo VIII.8, y a fin de mantener la igualdad en la fórmula (8.44), será necesario dividir por el mismo factor exponencial.

Tratandose de la función característica de una suma de variables aleatorias independientes (que es el producto de las funciones características de los sumandos), el factor exponencial a introducir será un producto de factores, que se reduce a una suma de exponentes. Teniendo en cuenta la forma de dicho factor exponencial que resulta de (8.32), será en este caso $e^{-i\sum m_i u} = e^{-iM_n u}$, y para dividir habrá simplemente que cambiar el signo del exponente, con lo que efectuando la introducción del factor y su inverso en (8.44), agrupando los términos exponenciales e introduciendo la notación de función característica de variables centradas empleada en el párrafo VIII.1, se tiene finalmente:

(8.45) $$f(v) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} e^{-i(v-M_n)u} \prod \varphi_i^*(u)\, du$$

Efectuando ahora el cambio de variables típico del teorema de LAPLACE:

(8.46) $$us_n = w \quad ; \quad s_n du = dw$$

se tendrá la nueva expresión:

(8.47) $$s_n f(v) = \frac{1}{2\pi}\int_{-\pi s_n}^{+\pi s_n} e^{-itw} \prod \varphi_i^{**}\left(\frac{w}{s_n}\right) dw$$

donde se ha agrupado en $t = \frac{(v-M_n)}{s_n}$ las variables que no intervienen en la integración. Pasando al límite para $n \to \infty$:

(8.48) $$\lim_{n\to\infty} s_n f(v) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itw}\, e^{-\frac{w^2}{2}}\, dw$$

La validez de este pasaje al límite bajo el signo integral necesita en realidad una demostración especial, porque además de las condiciones habituales para realizarlo, es necesario tener en cuenta que el intervalo de integración se convierte en un intervalo infinito, y precisamente el teorema de LAPLACE en la forma demostrada anteriormente requiere una acotación del intervalo de la variable para anular el resto. Afortunadamente esta demostración no presenta grandes dificultades (puede consultarse en el tratado de VON MISES (21)), y obliga a introducir solamente una condición adicional, que es la acotación superior de las derivadas segundas, con lo que en realidad se completa adecuadamente la interpretación de que el teorema se refiere a variables aleatorias de igual orden de magnitud en probabilidad.

Dando por sentado entonces la validez de (8.48), solo resta calcular efectivamente la integral, para lo que hay que separar la variable de integración de t, agrupando los exponentes y completando los cuadrados:

(8.49) $$\lim_{n\to\infty} s_n f(v) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{w^2+2itw}{2}}\, dw = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{w^2+t^2-t^2+2itw}{2}}\, dw$$

obteniendose el cuadrado de un binomio, con lo que finalmente, resulta:

(8.50) $$\lim_{n\to\infty} s_n f(v) = \frac{1}{2\pi} e^{-\frac{t^2}{2}}\int_{-\infty}^{+\infty} e^{-\frac{(w+it)^2}{2}}\, d(w+it) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

porque la integral es igual a la raíz cuadrada de 2π.

Cuando v es igual a uno de los valores de la variable aleatoria suma, f(v) es la probabilidad correspondiente, como se vió al demostrar la fórmula de inversión (8.37), de manera que cuando n es suficientemente grande se tiene la siguiente expresión aproximada:

(8.51) $$f(r) = \frac{1}{s_n\sqrt{2\pi}}\, e^{-\frac{(r-Mn)^2}{2s_n}}$$

que es la forma local, o puntual, del teorema central del límite, que puede tambien expresarse en forma integral:

(8.52) $$P\left(t_1 \leqslant \frac{r - M_n}{s_n} \leqslant t_2\right) = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{t^2}{2}}\, dt$$

VIII.14.- APLICACION A LA SUMA DE VARIABLES ALEATORIAS IGUALMENTE DISTRIBUIDAS. Cuando las variables aleatorias son iguales, es decir tienen igual campo de variación e igual ley de probabilidad, es evidente que valen las hipótesis del teorema de LAPLACE, por lo que su consecuencia, que es el teorema central del límite, tambien será válido (siempre que existan los momentos que intervienen en su formulación y demostración). Como caso particular, se tiene entonces el teorema de DE MOIVRE, demostrado independientemente, pero existen otros casos no tan sencillos como este último, en el cual es necesaria la demostración general que se acaba de dar, como ocurre por ejemplo en el paseo al azar, discutido en el Cap. V. Si $p=q=1/2$, se tiene que $m=0$ y $\sigma^2=1$ y aplicando (8.52) se tiene:

(8.53) $$P(x_1 \leqslant x \leqslant x_2) = \frac{1}{\sqrt{2\pi n}} \int_{\frac{x_1}{\sqrt{n}}}^{\frac{x_2}{\sqrt{n}}} e^{-\frac{x^2}{2n}}\, dx$$

con lo que se completa con mucho mas detalle la descripción lograda en el capítulo citado mediante una ley de los grandes números, ya que es posible calcular la probabilidad (en el caso, la proporción de la población de partículas) correspondiente a un intervalo dado. Esta descripción, que por simplicidad se ha referido al caso simétrico en que $p=q$ es ejemplo de un tipo de descripción de sistemas de gran utilidad e importancia en física.

VIII.15.- APLICACION A LA SUMA DE VARIABLES ALEATORIAS DESIGUALMENTE DISTRIBUIDAS. En este caso no es tan sencillo como en el anterior verificar que se cumplen las hipótesis de validez, y en realidad se halló indispensable continuar el estudio del problema hasta encontrar la condición necesaria y suficiente, ya que las condiciones dadas son solamente suficientes, y en el caso general podría ocurrir que una familia de variables aleatorias no las cumpliera. Esta condición fué hallada por el matemático finlandés LINDEBERG (a) como una nueva hipótesis suficiente, y años despúes FELLER logró probar que también era necesaria (ver apéndice). La condición de LINDEBERG, se refiere solamente a los momentos de segundo orden centrados.

(a)"Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung". Mathematische Zeitschrift, vol. 15 (1922) pag. 211-225.

A P E N D I C E A L C A P I T U L O V I I I.

VIII.A.1.- TEOREMA DE CONTINUIDAD. La transformación de leyes de probabilidad a funciones características define un isomorfismo entre ambos conjuntos, en el que a la operación de suma en uno le corresponde la de multiplicación en el otro.

Pero para poder demostrar el teorema central del límite, hay que probar que la transformación es continua, es decir, que al límite de la suma le corresponde el límite del producto, y lo que es más delicado, que el límite del producto es tambien una función característica. En otros términos, considerando a las funciones características como puntos en el espacio, podría ocurrir que el conjunto de puntos correspondientes al conjunto de funciones características no fuera cerrado, y que el límite de una sucesión de puntos no pertenezca al conjunto. Esta cuestión se aclarará más adelante con un ejemplo.

La demostración del teorema de continuidad consta, pues, de dos partes. La primera establece que la convergencia de las funciones características es condición necesaria y suficiente para la convergencia de las leyes de probabilidad, y la segunda cuales son las condiciones que debe cumplir ese límite para ser una función característica.

VIII . A.2.- Al demostrar la unicidad de la transformación, en (VIII.9), mediante la fórmula de inversión, se vió que a cada función generatriz corresponde una función característica y recíprocamente, de manera que el trabajar con un tipo u otro depende de con cual de las dos funciones resulta mas simple. Para la inversión, era más conveniente utilizar las funciones características, pero para el teorema de continuidad, y tratándose de leyes de probabilidad de variables aleatorias discretas, es preferible utilizar las funciones generatrices.

Dada una sucesión de variables aleatorias definidas para valores enteros de la variable (en particular, esa sucesión puede ser la de las sumas parciales de una familia infinita de variables), de la que se supone que sus leyes convergen, o sea que para el valor j la sucesión de probabilidades p_{ij} converge a un cierto valor p_j, quedará demostrada la parte necesaria del teorema de continuidad si se encuentra que la sucesión de funciones generatrices converge.

Si $g_i(s)$ es el término genérico de la sucesión de funciones generatrices :

(8.A.1) $$g_i(s) = p_{io} + p_{i1}s + p_{i2}s^2 + \cdots + p_{ij}s^j + \cdots$$

y $g(s)$ es la función generatriz de la sucesión de los límites de las columnas de coeficientes p_{ij}:

(8.A.2) $$g(s) = p_o + p_1 s + p_2 s^2 + \dots + p_j s^j + \dots$$

De conformidad con la definición de límite, para que la sucesión de las $g_i(s)$ converja a $g(s)$ en un punto s, debe verificarse que dado un ε positivo y arbitrario y a partir de un cierto i, el valor absoluto de la diferencia sea menor que ε:

(8.A.3) $$|g_i(s) - g(s)| < \varepsilon$$

Como por hipótesis, cada una de las sucesiones p_{ij} converge a p_j, para cada una de ellas se verificará que a partir de un cierto i:

(8.A.4) $$|p_{ij} - p_j| < \frac{\varepsilon}{2(n+1)}$$

El i en (8.A.4) dependerá en general del j, pero no es necesario establecer ninguna condición adicional de convergencia uniforme si la demostración se restringe al recinto $|s| < 1$

Dentro de dicho recinto se verificará la desigualdad:

(8.A.5) $$|g_i(s) - g(s)| \leqslant \sum_{j=0}^{n} |p_{ij} - p_j| s^j + 2\frac{s^{n+1}}{s-1}$$

en la que a partir de un cierto índice n se ha reemplazado el valor absoluto de la diferencia de las probabilidades por su cota superior 2 (porque cada una no puede superar la unidad).

Dado un cierto ε, habrá un cierto índice $j = n$ con el cual el segundo término del segundo miembro será menor que $\varepsilon/2$, y bastará un i suficientemente grande para que cada uno de los sumandos del primer término del segundo miembro sea menor que $\varepsilon/2(n+1)$, con lo que se tiene:

(8.A.6) $$|g_i(s) - g(s)| \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2}$$

y queda demostrada la parte necesaria.

La restricción a los valores de s que cumplen $|s| < 1$ fué necesaria para poder reemplazar términos del segundo miembro de (8.A.5) por una serie geométrica convergente, quedando un número finito de sumandos, lo que hace innecesario la convergencia uniforme de las sucesiones p_{ij}, y no resta generalidad a la demostración.

Si la sucesión de las funciones generatrices converge, para demostrar la parte suficiente del teorema hay que utilizar la construcción conocida por el nombre de procedimiento diagonal de Weierstrass, a fin de probar que cada uno de las sucesiones p_{ij}, converge a un cierto valor p_j.

Tomando la primer sucesión p_{io}, es evidente que sus términos estan acotados entre cero y la unidad, porque se trata de probabilidades, y conforme al teorema de Bolzano, existirá una subsucesión convergente, cuyo límite se designará por p_o^*. Seleccionando en la segunda sucesión p_{i1} los términos que corresponden a la subsucesión convergente anterior, se tiene una nueva sucesión acotada, que admitirá una subsucesión convergente

con límite p_1^*. En la primera sucesión, la subsucesión correspondiente a la que es convergente a p_1^* en la segunda, está comprendida dentro de otra que es convergente a p_0^*, y tendrá el mismo límite, por lo que es facil ver que de esta menera se construye una subsucesión tal que para todo j las p_{ij} tienen un límite p_j^*.

Construyendo ahora la función generatriz:

(8.A.7) $$\overset{*}{g}(s) = p_0^* + p_1^* s + p_2^* + \dots + p_j^* s^j$$

esta resulta como límite de la sucesión de funciones generatrices correspondiente a la sucesión de leyes en la que las probabilidades tenían el límite p_j^*, pero como esta sucesión de funciones generatrices está comprendida en otra que por hipótesis es convergente se verificará que:

(8.A.8) $$\overset{*}{g}(s) = g(s)$$

Pero esto implica demostrar que la sucesión de las leyes es convergente, porque si no fuera así, podrían seleccionarse subsucesiones correspondientes a límites distintos, y (8.A.8) demuestra que todos los límites tienen que ser iguales, con lo que queda probada la parte suficiente.

VIII.A.3.- Si la sucesión de leyes de probabilidad es tal que para $i=n$ los n primeros p_{ij} son nulos, es evidente que todos los límites $p_j = 0$, y el límite no es una ley de probabilidades.

Para hallar la propiedad de la función generatriz límite que asegure que no se verifica este caso excepcional, hay que recordar que en una función generatriz el valor para $s=1$ es la unidad, porque es la suma de las probabilidades. En la función límite de una sucesión de funciones generatrices se cumplirá también que $g(1)=1$, porque para $s=1$ es límite de una sucesión constantemente igual a 1, pero si la función es discontinua ya no se verifica que la suma de las probabilidades (que es la suma de los coeficientes), valga 1. Precisamente en el ejemplo excepcional dado, la función límite es nula para todo valor de s, salvo para $s=1$, en donde tiene una discontinuidad, y toma el valor 1.

Si se fija la condición de que la función límite sea continua en el punto $s=1$, se tendrá:

(8.A.9), $$g(1) = \lim_{s \to 1} \sum_{j=0}^{\infty} p_j s^j = \sum_{j=0}^{\infty} p_j$$

y se puede asegurar entonces que la función límite es la función generatriz de una ley de probabilidad, eliminando así el casoexcepcional.

VIII.A.4.- Pasando de las funciones generatrices a las características, el teorema de continuidad se enunciará entonces: la condición necesaria y

suficiente para que una sucesión de leyes de probabilidad converja a un límite que sea una ley de probabilidad, es que la sucesión de funciones características tenga como límite una función continua en el punto $u=0$.

Como la función que aparece como límite en el teorema de Laplace es precisamente continua en el punto cero, queda completada entonces la demostración del teorema central del límite.

VIII.A.5.- El análisis detallado del fenómeno aleatorio elemental, aquel que consiste en la realización o no realización de un suceso, que se llevó a cabo en los capítulos IV y VI, con la introducción de la variable aleatoria binomial y la demostración de los teoremas de Bernouilli y de De Moivre-Laplace, se puede resumir en las dos siguientes proposiciones:

(1) Al crecer n el crecimiento del desvío es menos rápido, en probabilidad, que el de n. En otros términos, dada una cierta constante ε, positiva y arbitraria, la probabilidad de que el cociente entre el desvío y n supere en valor absoluto a , tiende al límite cero cuando n crece indefinidamente:

(8.A.10) $$\lim_{n\to\infty} P\left(\left|\frac{r-np}{n}\right| > \varepsilon\right) \longrightarrow 0$$

Esta afirmación, que es otra manera de enunciar el teorema de Bernouilli, se precisa estableciendo exactamente cual es el orden de crecimiento del desvío:

(2) Al crecer n el crecimiento del desvío es del mismo orden, en probabilidad, que el de $\sqrt{n}$. O sea, que dada una constante t, la probabilidad de que el valor absoluto del cociente entre el desvío y la raíz cuadrada de n sea menor que un número proporcional a t, cuando n crece indefinidamente tiende a una constante, que está dada por la fórmula integral de Laplace:

(8.A.11) $$\lim_{n\to\infty} P\left(\left|\frac{r-np}{\sqrt{n}}\right| < t\sqrt{pq}\right) = \frac{1}{2\pi}\int_{-t_1}^{+t_1} e^{-\frac{t^2}{2}}\,dt$$

En los capítulos V y VIII se generalizaron estas dos proposiciones, mediante las leyes de los grandes números y el teorema central del límite, a las sumas de variables aleatorias, sin otra hipótesis para el caso de variables aleatorias iguales que la existencia de los momentos de primer y segundo orden, que intervienen en el enunciado o en la demostración, y con ciertas limitaciones sobre el comportamiento de los momentos de segundo orden y tambien del de tercero, cuando se trataba de variables aleatorias no iguales.

De esta manera, y mediante los teoremas límites, el método de tratamiento para el fenómeno aleatorio elemental se extendió a fenómenos aleatorios más complejos, obtenien-

do la posibilidad de encontrar nuevas constantes asintóticas cuya existencia no era evidente si se considera solamente el comportamiento descripto como característico de los fenómenos aleatorios, y pudiendo formular enunciados con tanta más precisión cuanto mayor sea el número de observaciones, con lo que queda removida la incertidumbre propia de todo fenómeno aleatorio y reducido su tratamiento, en casos muy generales, a un método standard.

Es evidente, por lo tanto, que los teoremas límites con cuyo auxilio se han obtenido estos resultados, constituyen el núcleo básico del cálculo de probabilidades.

VIII.A.6.- Las demostraciones dadas son solamente suficientes, pero desde fecha relativamente reciente se conocen también las condiciones necesarias y suficientes.

La condición necesaria y suficiente para la ley de los grandes números fué hallada por Kolmogoroff (a) y la del teorema central del límite por Feller (b).

Comparando las fechas de publicación de estas memorias con las del teorema de Bernouilli (1713) y el de De Moivre (1732), puede parecer sorprendente que haya transcurrido tanto tiempo entre el descubierto de los primeros resultados y su generalización más amplia.

Este hecho se debe a que prácticamente el cálculo de probabilidades se mantuvo ajeno al gran desenvolvimiento matemático del siglo XIX, circunstancia que hoy no aparece como fácilmente explicable. Sim embargo, posiblemente tuvo gran influencia el que el método propio del cálculo de probabilidades, la transformación a funciones generatrices y características, cayó en desuso, y con él la posibilidad de estudiar sus teoremas más importantes.

La crítica al método de las funciones generatrices, durante el siglo XIX, se debió a un análisis superficial del uso que de las mismas hizo Laplace. En efecto, en su gran " Theorie Analytique des Probabilités ", las funciones generatrices son empleadas con frecuencia para integrar ecuaciones en diferencias finitas que aparecen en numerosos problemas, como se verá en el Cap.IX. Boole y otros matemáticos utilizaron otros procedimientos más eficaces para tratar las diferencias finitas, y se pasó por alto la importancia que tenían las funciones generatrices para los problemas básicos del cálculo de probabilidades, a lo contribuyó el mismo Laplace, que emplea una notación extremadamente confusa que hace dificilmente inteligibles muchas partes de su obra.

El estado actual del cálculo de probabilidades, especialmente en lo que se refiere a los teoremas límites, se haya expuesto en las obras de GNEDENKO-KOLMOGOROFF (10) y

(a)"Uber die Summen durch den Zufall bestimmter unabhängigen Grössen" Mathematische Annalen, vol 99, pag. 309-319.
(b)"Uber der zentrallen Grenzwertsatz der Wahrscheinlichkeitrechunng", Mathematische Zeitschriff, vol. 40 (1935), pag. 521-559.

LOEVE (18), que usan recursos matemáticos superiores. Una exposición más elemental, pero muy clara y completa, es la de GNEDENKO (7), que puede considerarse sin duda como el mejor libro de texto que hasta ahora se haya publicado en la materia.

Tambien tiene considerable interés la obra de Paul Levy, a quien conjuntamente con Von Mises se debe la revitalización del cálculo de probabilidades, y que en lo que tiene contacto con este curso se halla expuesta en (16) y (17).

Una exposición de divulgación, de lectura recomendable, se encuentra en una conferencia de FELLER (a).

(a) "The Fundamental Limit Theorems in Probability", Bulletin of the American Mathematical Society, vol 51 (1945), pag. 800-832.

CAPITULO IX.

CADENAS DE PROBABILIDADES. SUCESOS DEPENDIENTES.

IX.1.- En muchos problemas, la aplicación directa de los teoremas fundamentales no permite calcular las probabilidades buscadas, sino únicamente establecer ciertas relaciones entre ellas, a partir de las cuales se las despeja utilizando métodos especiales.

Sea, por ejemplo, calcular la probabilidad de la aparición de una racha de longitud r en una sucesión de Bernoulli (ver Capítulo IV), entendiendose que se ha producido una racha de longitud r cuando una de las dos alternativas se ha presentado por lo menos r veces en sucesión.

El problema no está completamente formulado si no se fija también la longitud de la sucesión de Bernouilli, pues evidentemente la probabilidad de la aparición de la racha varía con n. Si $n=r$, dicha probabilidad es igual a p^r, y si $n<r$ es igual a cero porque entonces la racha es imposible. Pero si $n>r$, el problema se complica.

DE MOIVRE halló la solución recurriendo a un artificio importante, que como se dijo más arriba, consiste en establecer una relación entre las probabilidades de observación de la racha, así definida, para varios valores de la longitud de la sucesión de Bernouilli.

Llamando y_n a la probabilidad de observar la racha en una sucesión de longitud n, es evidente que y_{n+1} corresponde a un suceso que se puede descomponer en la suma de otros dos excluyentes: (1) La racha se observó ya en el rango n; (2) La racha se produce de tal manera que puede observarse solamente en el rango $n+1$.

Las probabilidades de estos dos sucesos excluyentes se pueden calcular sin dificultades, admitiendo que se conocen las y_i. En efecto, la del primero es, por definición, y_n, y la del segundo es, a su vez, el producto de las probabilidades de otros tres sucesos independientes entre sí: (1) La racha no apareció en el rango $n-r$ (probabilidad igual a $1-y_{n-r}$); (2) En la observación de rango $n-r+1$ no se dió el suceso que forma la racha, a cuya aparición individual se asigna la probabilidad p (por lo tanto, la probabilidad de este suceso será $1-p=q$); (3) En las r últimas observaciones se repitió el suceso que forma la racha (probabilidad igual a p^r). Finalmente, se tiene:

$$y_{n+1} = y_n + (1 - y_{n-r})qp^r \tag{9.1}$$

Esta relación entre las y_i correspondientes a tres rangos distintos determina completamente la sucesión de las y_i, porque dado un n determinado es posible calcular y_n efectivamente en una forma única, debido a que se tiene:

$$y_1 = y_2 = \cdots = y_{r-1} = 0 \quad ; \quad y_r = p^r \tag{9.2}$$

y con estas c o n d i c i o n e s i n i c i a l e s es posible determinar por recurrencia cualquier y_i.

IX.2.- El mismo DE MOIVRE aplicó un procedimiento similar al problema de la ruina de los jugadores, al que se aludió al final del Capítulo V, y que es uno de los clásicos del cálculo de probabilidades.

Sea dos jugadores que juegan entre sí un cierto juego de azar. Para uno de los jugadores, la probabilidad de ganar en una jugada es p, y la perder q (para el otro seran, respectivamente, q y p, pues cuando uno gana el otro pierde), y no existe una tercera alternativa (empate).

La apuesta de cada uno de los jugadores es unitaria, y la ganancia neta (ver Capítulo V, párrafo V.13) será, por consecuencia, también unitaria. Las reglas de juego establecen que hay que abonar la apuesta antes de cada jugada, debido a lo cual un jugador se arruinará (no podrá continuar jugando), si su pérdida acumulada llega a igualar su capital, y el problema consiste en calcular las probabilidades que tiene cada uno de arruinarse, las que corresponden al triunfo del otro.

Si el capital de uno de los jugadores es igual a x, y procediendo de igual manera que en problema anterior de las rachas, se supone conocida la probabilidad de arruinarse con ese capital, llamada q_x, es posible tambien establecer una relación entre las probabilidades correspondientes a distintos valores de x:

(9.3) $$q_x = pq_{x+1} + qq_{x-1}$$

porque cuando el jugador de capital x realiza una jugada, si gana aumenta su capital en una unidad, y si pierde lo disminuye también en una unidad, y el suceso "arruinarse cuando se dispone de un capital x" se descompone en la suma de dos sucesos excluyentes: (1) arruinarse después de haber ganado la primer jugada; (2) arruinarse después de haber perdido la primera jugada.

Si la suma de los capitales de ambos jugadores es igual a C, es evidente que $q_C=0$ porque cuando uno de los jugadores llega a acumular el capital de los dos, el otro se ha arruinado, y como el juego se interrumpe, la probabilidad de arruinarse corresponde al suceso imposible.

Recíprocamente, $q_0=1$, porque cuando un jugador pierde su capital, el juego se interrumpe, y la probabilidad de arruinarse corresponde al suceso cierto.

En forma simétrica, se puede plantear la relación correspondiente a las p_x (que son las probabilidades de ganar y corresponden a las de ruina del otro jugador):

(9.4) $$p_x = pp_{x+1} + qp_{x-1}$$

De manera mas elaborada que en el caso anterior, es posible demostrar que tanto la sucesión de las q_x como la de las p_x, que a diferencia de las y_n de las rachas son sucesiones finitas, estan determinadas unívocamente por las c o n d i c i o n e s e n l o s l í m i t e s, que se acaban de dar.

IX.3.- FELLER (4) señala que este problema de la ruina de los jugadores es completamente equivalente al del paseo al azar lineal de una partícula, realizado en las mismas condiciones descriptas en el Capítulo V (párrafo V.8), sólo que con m u r o s a b s o r b e n t e s en las abcisas 0 y C, o sea puntos en los que la partícula queda retenida si llega a ellos, problema que presenta considerable interés para la física.

IX.4.- Las c a d e n a s d e p r o b a b i l i d a d e s, como se llama a las relaciones del tipo de (9.1), (9.3) y (9.4), son en realidad ejemplos de e c u a c i o n e s e n d i f e r e n c i a s f i n i t a s, a las que los matemáticos del siglo XVIII y primera mitad del XIX otorgaron gran importancia, aunque hoy están un tanto olvidadas, salvo para su uso en problemas de interpolación.

Sin embargo, la integración de las ecuaciones en diferencias finitas, como se llama al procedimiento de encontrar una expresión de las y_i o de las q_x, en función de i o de x (función definida sólo para valores enteros de la variable), presenta un gran interés, en el caso de las cadenas de probabilidades, porque aparte de que las soluciones por recurrencia pueden conducir a un número de operaciones elevado, como en el caso de (9.1), si n es grande, o ser practicamente imposibles, como en (9.3) y en (9.4), muchas veces no interesa tanto calcular una probabilidad determinada como el conocimiento de las propiedades asintóticas de la sucesión de probabilidades.

Un ejemplo del uso de la técnica clásica de la integración de las ecuaciones en diferencias finitas puede verse en el tratamiento que hace FELLER (4) del problema de la ruina de los jugadores, pero también, siguiendo a LAPLACE es posible aplicar funciones generatrices, (como hace USPENSKY (22), Cap. V, para el problema de las rachas).

Tanto con un método como con el otro, a menudo resultan también operaciones demasiado complicadas o numerosas, y ha sido necesario hallar métodos numéricos aproximados, cuyo iniciador fué también DE MOIVRE.

Del conjunto de métodos utilizados para tratar las cadenas de probabilidades se seleccionará en este curso el de las funciones generatrices, en la forma desarrollada por FELLER (4) para los llamados s u c e s o s r e c u r r e n t e s.

IX.5.- Segun este último autor, un suceso recurrente está caracterizado por las siguientes dos propiedades: (1) Su definición permite determinar, conocidos los resultados de n observaciones independientes si se ha producido o no y en qué rango de observación; (2) Cada vez que se produce, las observaciones que constituyen el suceso recurrente no pueden ser utilizadas para definir una nueva ocurrencia.

El ejemplo de las rachas servirá para aclarar estas propiedades. Sea una racha

de longitud $r=4$, y en la observación n-ésima termina una sucesión de seis observaciones de la alternativa que constituye la racha. Conforme a las propiedades enunciadas, la racha se habrá producido en la observación de rango n-2 (propiedad 1), y no en las de rango n-1 y n-2 (por la propiedad 2, una vez producida la racha en la observación n-2, hay que empezar a contar de nuevo para decidir que se está ante una nueva ocurrencia).

Con esta definición, el problema de calcular la probabilidad de una racha es distinto del planteado por DE MOIVRE, conforme se vió anteriormente. En efecto, en el párrafo IX.1, la probabilidad y_n no correspondía a que la racha hubiera terminado en la observación n-ésima, sino simplemente a que se hubiera presentado con anterioridad una sucesión de por lo menos r resultados iguales, y quedaba totalmente indeterminado el saber cuantas rachas se habían producido si la sucesión de resultados iguales presentaba una longitud mayor que r.

IX.6.- Una vez definido un suceso recurrente, y dada una sucesión ilimitada de observaciones, se pueden definir a su vez dos sucesiones ilimitadas de probabilidades: (1) La sucesión u_j correspondiente a la aparición incondicional del suceso en la observación de rango j ;(2) La sucesión f_j de probabilidades de aparición p o r p r i m e r a v e z en la observación de rango j.

Las probabilidades f_j corresponden a sucesos excluyentes, y por lo tanto la suma de n cualesquiera de ellas no puede superar la unidad. La serie de todas las f_j es de términos positivos, y como segun la observación anterior, sus sumas parciales están acotadas superiormente, será siempre convergente.

Si la suma de la serie de las f_j es la unidad, se dice que el suceso recurrente es c i e r t o, porque tiene una probabilidad 1 de aparecer al menos una vez. Si la suma de la serie de las f_j es igual a $f<1$, se dice que el suceso recurrente es i n c i e r t o, porque tiene una probabilidad $1-f$ de no aparecer nunca.

Las probabilidades u_j no corresponden a sucesos excluyentes, por lo que no existe acotación superior para sus sumas parciales, y la serie puede ser convergente o divergente (pero nunca oscilante, porque es una serie de términos positivos).

IX.7.- TEOREMA FUNDAMENTAL DE LOS SUCESOS RECURRENTES. Conforme se vió en el capítulo anterior, dadas las sucesiones u_j y f_j, es posible definir las funciones generatrices de ambas, como funciones de un cierto parámetro real o complejo s, tales que su desarrollo en serie de potencias del parámetro tenga por coeficientes los términos de las sucesiones:

(9.5) $$U(s) = \sum_j u_j s^j \quad ; \quad F(s) = \sum_j f_j s^j$$

pero en general no se cumplirá que para $s=1$ las funciones generatrices valen la unidad. En efecto, las sucesiones u_j y f_j no corresponden a leyes de probabilidad de una variable aleatoria. En la primera, los sucesos no son excluyentes, y de la segunda no forma parte la probabilidad de que el suceso no aparezca nunca, la que podrá ser cero o tener un valor no nulo, segun que el suceso sea cierto o incierto.

Pero entre las sucesiones u_j y f_j existe una relación fundamental, y por otra parte evidente:

(9.6) $$u_n = f_1 u_{n-1} + f_2 u_{n-2} + \cdots + f_{n-1} u_1 + f_n$$

dado que el suceso de aparecer incondicionalmente en la observación de rango n-ésimo se puede descomponer en n sucesos excluyentes, segun que haya aparecido por primera vez en la primera observación y vuelto a aparecer después de n-1 observaciones, y así sucesivamente. Por la independencia de las observaciones, supuesta en la definición de sucesos recurrentes, la sucesión de las u_j permanece invariante si se desplaza el origen.

En la expresión (9.5), todos los términos menos el último son productos en los que los índices de las probabilidades suman n. Para hacer (9.5) totalmente simétrica es conveniente completar la sucesión de las u_j con un término $u_0=1$, y tambien la de las f_j con $f_0=0$, obteniendose entonces:

$$u_n = f_0 u_n + f_1 u_{n-1} + \cdots + f_n u_0$$

Si se multiplica ambos miembros de (9.6) por s^n, resulta que el coeficiente de la potencia n-ésima de s en el desarrollo de la función generatriz $U(s)$ (completada con el término de rango 0) tiene la forma correspondiente al producto de las funciones generatrices $U(s)$ y $F(s)$. Como (9.6) es válida solamente desde 1 en adelante, sumando todas las expresiones similares, se tendrá:

(9.7) $$U(s) - 1 = U(s)F(s)$$

de donde resulta inmediatamente:

(9.8) $$U(s) = \frac{1}{1 - F(s)}$$

que es el t e o r e m a f u n d a m e n t a l d e l o s s u c e s o s r e c u r r e n t e s.

IX.8.- Si en (9.8) se da a s el valor 1, $F(1)$ será la suma de la serie de las f_j y si el suceso es cierto se tendrá:

(9.9) $$F(1) = 1$$

y si el suceso es incierto:

(9.10) $$F(1) < 1$$

A su vez, U(1) será la suma de la serie de las u_j, de lo que resulta inmediatamente: La condición necesaria y suficiente para que un suceso sea cierto es que la serie de las u_j sea divergente. Y tambien: La condición necesaria y suficiente para que un suceso sea incierto es que la serie de las u_j sea convergente.

La adición de $u_0=1$ no origina inconvenientes, porque como no se trata de una ley de probabilidades, el valor numérico de la función generatriz para $s=1$ no interesa para el cálculo de momentos. Lo que importa es si la serie es convergente o divergente, comportamiento que no es alterado por la adición de una constante.

La sucesión de las u_j, es, a veces, más fácil de determinar que la de las f_j, y por lo tanto el teorema demostrado permitirá en ciertos casos establecer en forma indirecta si un suceso es cierto o incierto, en lugar de sumar la serie de las f_j, así como deducir importantes consecuencias, como se verá al tratar las cadenas de Markoff.

IX.9.- APLICACION A LA TEORIA DE LAS RACHAS. Definida una racha de longitud r en una sucesión ilimitada de Bernouilli, como un suceso recurrente (ver párrafo IX) y no en la forma en que lo hizo DE MOIVRE, el primer problema es determinar si es un suceso cierto o incierto.

Dada la independencia de las observaciones que caracteriza a la sucesión de Bernouilli, la probabilidad de que en el rango n termine una sucesión de r observaciones iguales, de probabilidad p, es igual a p^r. Teniendo en cuenta la definicion de racha como suceso recurrente, la aparición de una sucesión de r observaciones iguales puede descomponerse en la suma de r sucesos excluyentes, cada uno de los cuales consiste en la observación de una racha en el rango $n-j$, y más j sucesos iguales de probabilidad p en los rangos siguientes.

La probabilidad de cada uno de estos sucesos elementales será:

(9.11) $$u_{n-j}p^j$$

con el índice j variando de 0 a $r-1$.

La probabilidad del suceso suma será igual a la suma de las probabilidades y se tendrá:

(9.12) $$p^r = u_n + u_{n-1}p + \cdots + u_{n-r+1}p^{r-1}$$

Multiplicando ambos miembros de (9.12) por s^n, y sumando a partir de $n=r$ se tendrá en el primer miembro una serie geométrica de razón (ps) y primer término igual a $(ps)^r$. Limitando el campo de variación de s con la condición de que su valor absoluto sea menor que la unidad, dicha serie será convergente y de suma igual a :

(9.13) $$(ps)^r(1 + s + \dots) = \frac{p^r s^r}{1 - s}$$

En el segundo miembro, la suma de cada uno de los términos multiplicados por s^n será la función generatriz de las u_j (menos $u_0 = 1$) multiplicada por $(ps)^{n-j}$, o sea

(9.14) $$[U(s) - 1]\,[1 + ps + \dots + (ps)^{r-1}] = [U(s) - 1]\,\frac{1 - p^r s^r}{1 - ps}.$$

Igualando (9.14) y (9.15):

(9.15) $$[U(s) - 1] = \frac{(1 - ps)p^r s^r}{(1 - s)\,(1 - p^r s^r)}$$

de donde se despeja U(s)

(9.16) $$U(s) = \frac{1 - s + qp^r s^{r+1}}{(1 - s)(1 - p^r s^r)}$$

Si la serie U(s) fuera convergente para $s = 1$, su suma sería igual al límite para $s = 1$ del segundo miembro de (9.17). Pero como este límite es infinito, la serie de las u_j es divergente, y por lo tanto la racha es un suceso cierto.

IX.10.- TIEMPOS DE RECURRENCIA. Tratándose de un suceso cierto, es posible considerar entonces que la sucesión de las f_j define la ley de probabilidad de una cierta variable aleatoria numerable (si el suceso fuera incierto, habría una probabilidad para un valor infinito, que sería una forma convencional de decir que existe una cierta probabilidad de que el suceso no aparezca y no se trataría de una variable aleatoria propia), llamada tiempo de recurrencia, porque dada la independencia que caracteriza a la sucesión de Bernouilli, se puede tomar como origen cualquier observación, y no sólo la sucesión de las u_j, como se vió al demostrar el teorema fundamental en el párrafo IX.7 , sino también la de las f_j es siempre la misma, de manera que f_j puede interpretarse como la probabilidad de que el suceso, una vez observado, vuelva a aparecer al transcurrir exactamente un tiempo j (la expresión tiempo en realidad se refiere al número de observaciones).

La introducción de esta variable aleatoria permite lograr una descripción del fenómeno de las rachas en una forma relativamente simple y muy completa.

Si se llama N_x a una nueva variable aleatoria, definida como el número de ocurrencias en x observaciones y T a la variable aleatoria tiempo transcurrido hasta la realización del suceso y veces, es evidente que:

(9.17) $$Pr(N_x \geqslant y) = P(T_y \leqslant x)$$

porque la probabilidad de que el número de veces que el suceso se observa en un total fijo de x observaciones sea igual o supere a y es evidentemente igual a la probabilidad de que el tiempo transcurrido para y observaciones sea igual o inferior que x .

Pero la variable T_y es simplemente la suma de y variables aleatorias independientes entre sí, e iguales a la variable tiempo de recurrencia que se acaba de definir, por lo que le serán aplicables los teoremas límites, si dicha variable aleatoria numerable posee los momentos cuya existencia requieren esos teoremas. En particular el teorema central del límite permitirá resolver el problema de hallar la probabilidad de observar un número dado de rachas en n observaciones, sumamente difícil de hallar con otros procedimientos.

Para ello habría que determinar $F(s)$, lo que es posible porque conociendo $U(s)$ basta aplicar el teorema fundamental, y calcular las derivadas correspondientes en el punto $s=1$.

Aunque las operaciones son un tanto laboriosas, la simplicidad de descripción que se logra es notable. Las expresiones del primer momento y de la varianza, así como algunos ejemplos numéricos, correspondientes al primer momento o t i e m p o m e d i o d e r e c u r r e n c i a, se pueden ver en el libro de FELLER (4).

IX.11.- SUCESOS DEPENDIENTES. CADENAS FINITAS DE MARKOFF DE PRIMER ORDEN. Dado un conjunto finito de sucesos o e s t a d o s E_0, E_1, ... E_n, se dice que constituyen una cadena finita de Markoff de primer orden cuando dado un estado está definida una distribución de probabilidades correspondiente al pasaje a otro estado, o a la permanencia en el mismo estado:

(9.18) $$P(E_j/E_i) = p_{ij}$$

En otros términos, la probabilidad de alcanzar un estado E_j depende únicamente del estado anterior E_i, pero no de los estados anteriores a E_i, lo que constituye un tipo especial y muy importante de dependencia, como se verá en el párrafo IX.14.

El conjunto de p r o b a b i l i d a d e s d e t r a n s i c i ó n p_{ij} puede ser convenientemente representado mediante una matriz cuadrada de orden igual al número de estados, y en la que los términos de la fila i son las probabilidades de transición al estado j (individualizado por la columna):

(9.19) $$|P| = \begin{vmatrix} p_{oo} & p_{o1} & \cdots & p_{on} \\ \cdots & \cdots & \cdots & \cdots \\ p_{no} & p_{n1} & \cdots & p_{nn} \end{vmatrix}$$

Como cada fila corresponde a una distribución de probabilidades sobre un sistema completo de sucesos excluyentes, su suma tiene que ser necesariamente igual a la unidad, llamandose a toda matriz que cumple la condición de tener sus términos no negativos y la suma de sus filas igual a la unidad, m a t r i z e s t o c á s t i c a.

Para considerar completamente descripto el comportamiento de un sistema representado por una cadena de Markoff de primer orden, sería necesario dar una d i s t r i b u c i ó n i n i c i a l d e p r o b a b i l i d a d e s P_i, y la probabilidad de una evolución del sistema que se inicia en el estado E_i, pasa al E_j, luego al E_k hasta llegar en un número finito de etapas al estado E_h, sería por aplicación del teorema del producto en el caso general (ver párrafo III.8):

(9.20) $$P(E_i - E_j - E_k - \ldots - E_h) = P_i p_{ij} p_{jk} \ldots p_{gh}$$

donde $(E_i - E_j - \ldots - E_h)$ representa la sucesión de estados, llamada c a d e n a o t r a y e c t o r i a, que constituye la evolución del sistema. Naturalmente, en una trayectoria los estados pueden hallarse en cualquier orden y no solamente en el dado por los índices atribuidos por razones de identificación, a condición que ninguna de las probabilidades de transición involucradas sea nula, lo que determinaría la imposibilidad de la trayectoria.

La distribución inicial, salvo en el importante caso de las c a d e n a s e s t a c i o n a r i a s que se definiran y estudiaran más adelante, depende de las circunstancias particulares en que se considere cada sistema.

IX.12.- EJEMPLOS DE CADENAS FINITAS DE MARKOFF DE PRIMER ORDEN. (1) E s t r u c t u r a e s t a d í s t i c a d e u n l e n g u a j e e s c r i t o. En el párrafo III.14 se hizo referencia a la circunstancia de que en un lenguaje escrito los distintos signos no son equiprobables, sino que dependen del anterior. La estructura estadística del lenguaje podría describirse, segun esta observación, mediante una matriz cuadrada de orden igual al número de signos del alfabeto más uno (para incluir el espacio en blanco) y las probabilidades de transición se determinarian aproximadamente computando la frecuencia relativa de los digramas (o grupos de letras) que aparecen en un texto escrito suficientemente largo, una vez fijada la letra inicial (a).

Una trayectoria de esta cadena de Markoff representaría un texto escrito en el idioma en cuestión. Sin embargo, si se determina la matriz de probabilidades de transición, y se realiza la extracción de una trayectoria conforme a los métodos que se describirán en la parte cuarta, el texto obtenido será en general incoherente, aunque fonéticamente recordará al idioma. Esto se debe a que un lenguaje es en realidad una cadena de Markoff de orden superior (la probabilidad de la aparición de un signo o del espacio en blanco depende no solamente del signo anterior, sino del conjunto de los tres o cuatro signos anteriores), y no de primer orden.

(a) En los libros sobre criptografía como el citado más adelante en el párrafo IX.21, se dan las frecuencias relativas experimentales de los digramas mas corrientes.

(2) P a s e o a l a z a r. En el paseo al azar lineal (ver párrafo V.7), el estado E_i es simplemente la abcisa que ocupa la partícula en un instante determinado. El caso general, en el que no hay limitación de la abcisa que la partícula puede ocupar, requeriría para su tratamiento una cadena infinita de Markoff, porque el número de estados posibles es infinito, y cuya teoría ha sido satisfactoriamente desarrollada, pero que al igual que las de orden superior citadas más arriba, no seran consideradas en este curso.

Pero si el conjunto de posiciones que puede ocupar la partícula es finito, entonces sí es posible representar el paseo al azar lineal mediante una cadena de Markoff finita. En el caso de los muros absorbentes , que es identico al problema de la ruina de los jugadores (ver párrafo IX.3), la matriz de probabilidades de transición sería la siguiente:

$$(9.21) \qquad |P| = \begin{vmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ q & 0 & p & 0 & \dots & 0 \\ 0 & q & 0 & p & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{vmatrix}$$

Cuando la partícula llega al estado E_o , o al E_n , queda retenida, lo que se indica con la probabilidad 1. En cualquiera de los estados intermedios, tiene probabilidad q de volver al estado anterior, y p de pasar al estado siguiente. Si se estudia al sistema a partir del estado E_j (abcisa inicial, o distribución inicial del capital de los jugadores), la distribución inicial de probabilidades P_i correspondiente se reduce a $P_j = 1$, y todas las demás $P_i = 0$.

Con escasas modificaciones a la matriz (9.21), se puede describir otros esquemas de interés para la física. Sea por ejemplo, el caso de los m u r o s r e f l e c t o r e s, en el que la partícula, al llegar a uno de los extremos de su recorrido, es rebotada a la posición inmediata anterior. La matriz correspondiente, que no requiere mayor explicación, es la siguiente:

$$(9.22) \qquad |P| = \begin{vmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & q & 0 & p & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 \end{vmatrix}$$

También puede considerarse el caso en que los muros no son perfectamente elásticos, y hay una absorción parcial, al que corresponde la matriz siguiente:

$$(9.23) \qquad |P| = \begin{vmatrix} q & p & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & q & 0 & p & \cdots & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & q & 0 & p & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & q & p \end{vmatrix}$$

(3) M o d e l o d e d i f u s i ó n d e E h r e n f e s t: En 1907, los esposos Ehrenfest propusieron un modelo de difusión extremadamente simple (a), a fin de responder a ciertas objeciones presentadas contra las teorías de Boltzmann. Sea un sistema de n partículas, las que estan distribuídas en dos recipientes, A y B. El estado del sistema está completamente caracterizado por el conocimiento del número de partículas que se encuentra en uno de los recipientes, por ejemplo el A. Si en un cierto instante se encuentra en el estado E_i (i moléculas en el recipiente A), el sistema puede pasar al estado E_{i-1} o E_{i+1} segun que una molécula de A pase a B (transición E_i a E_{i-1}) o de B a A (transición E_i a E_{i+1}). Si se atribuye a estas transiciones una probabilidad proporcional al número de partículas que se encuentre en el recipiente de origen de la partícula que pasa al otro, o sea $P(E_{i-1}/E_i) = i/n$ y $P(E_{i+1}/E_i) = \frac{n-i}{n}$, es posible obtener importantísimas consecuencias que se discutirán más adelante (párrafo IX.23).

En forma de cadena de Markoff, la matriz de probabilidades de transición es la siguiente:

$$(9.24) \qquad |P| = \begin{vmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1/n & 0 & 1-1/n & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 2/n & 0 & 1-2/n & \cdots & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1-2/n & 0 & 2/n & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1-1/n & 0 & 1/n \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 \end{vmatrix}$$

(a) P. und T. Ehrenfest: "Uber zwei bekannte Einwände gegen Boltzmannsche H-Theorem" Physical. Zeit. vol. 8 (1908), pag. 311-314.

IX.13.- PROBABILIDADES DE TRANSICION DE ORDEN SUPERIOR. LAS CADENAS DE MARKOFF COMO CONJUNTO DE SUCESOS RECURRENTES. La transición de un estado E_i a otro E_j en dos etapas puede realizarse pasando como estado intermedio por cualquier estado del conjunto que constituye la cadena de Markoff (incluyendo como estado intermedio a los mismos E_i y E_j), salvo que alguna de las probabilidades de transición involucradas sea nula. Como cada una de las trayectorias es excluyente de las demás, la probabilidad de transición de E_i a E_j puede expresarse como la suma de los productos $p_{ik} \cdot p_{kj}$ para todo el campo de variación de k, siendo nulos los productos correspondientes a las trayectorias imposibles:

$$p_{ij}^{(2)} = \sum_{k=0}^{n} p_{ik} p_{kj} \tag{9.25}$$

indicando el índice superior (con paréntesis, para diferenciarlo de un exponente), el número de etapas, y pudiendo verificarse sin dificultad que

$$\sum_j p_{ij}^{(2)} = 1 \tag{9.26}$$

Como (9.25) es la expresión del término genérico del producto de la matriz de probabilidades de transición $|P|$ por si misma, se tiene que el conjunto de las probabilidades de transición en dos etapas se puede representar convenientemente mediante el cuadrado de la matriz P, la que conforme con (9.26), es tambien una matriz estocástica. Procediendo por inducción, se tiene

$$p_{ij}^{(t)} = \sum_{k=0}^{n} p_{ik}^{(t-1)} p_{kj} \tag{9.27}$$

y el conjunto de probabilidades de transición en t etapas estará dado por la potencia de exponente t de la matriz $|P|$, habiéndose adoptado el índice t para indicar el número de etapas porque en la mayoría de las aplicaciones de las cadenas de Markoff se supone que cada etapa transcurre en la unidad de tiempo, siendo más cómodo referirse al tiempo transcurrido que al número de etapas.

Dado un estado genérico E_i, queda por lo tanto univocamente determinada la sucesión de probabilidades de llegar al estado E_j en un tiempo t, y tambien la de las probabilidades de llegar p o r p r i m e r a v e z al estado E_j en un tiempo t ya que las probabilidades correpondientes $f_{ij}^{(t)}$ estan definidas por las $p_{ij}^{(t)}$:

$$f_{ij} = p_{ij} \quad ; \quad f_{ij}^{(2)} = p_{ij}^{(2)} - p_{ij} \quad ; \ldots ; \quad f_{ij}^{(t)} = p_{ij}^{(t)} - \sum_{m=1}^{t-1} f_{ij}^{(m)} p_{jj}^{(t-m)} \tag{9.28}$$

Es evidente que la serie $\sum_{t=1}^{\infty} f_{ij}^{(t)}$ es siempre convergente, porque el llegar por primera vez a un estado procediendo de otro es excluyente con todas las alternativas similares, y las sumas parciales estaran acotadas por la unidad (ver párrafo IX.6)

Como las propiedades de los sucesos recurrentes definidos en el párrafo IX.5, dependen en realidad de la existencia de las dos sucesiones de probabilidades de aparición y de aparición por primera vez y podría haberse prescindido de la condición de independencia en su definición, las cadenas de Markoff pueden tratarse como un conjunto de sucesos recurrentes, haciendo las convenciones $f_{ij}^{(0)} = 0$ y $p_{ij}^{(0)} = 1$.

A la relación

$$p_{ij}^{(t)} = \sum_{m=1}^{t} f_{ij}^{(m)} p_{jj}^{(t-m)} \tag{9.29}$$

que con las convenciones anteriores es otra forma de escribir la última expresión de (9.28), y es a homóloga de la (9.6) de los sucesos recurrentes, LOEVE (18) la ha llamado acertadamente r e l a c i ó n c e n t r a l de las cadenas de Markoff, porque efectivamente ocupa una posición central en el estudio de sus propiedades.

IX.14.- LAS CADENAS FINITAS DE MARKOFF COMO MODELO MATEMATICO DE SISTEMAS FISICOS. Esta propiedad de las cadenas de Markoff, que una vez observado un estado toda la evolución futura de la cadena queda determinada, en términos de probabilidades, con exclusión de lo que ocurrió anteriormente al estado observado, presenta un gran interés. En efecto, como señala LOEVE (18), gran parte de nuestro conocimiento de las leyes de la naturaleza se refiere a los fenómenos n o h e r e d i t a r i o s, es decir que dado un estado se puede describir la evolución futura del sistema físico con prescindencia de la historia anterior, y cuyo ejemplo típico es la mecánica del punto material.

Las cadenas finitas de Markoff presentan no sólo un ejemplo sencillo de modelo probabilístico no hereditario, que es el homólogo de los modelos matemáticos deterministas que usa la física clásica sino que como se verá mas adelante (párrafo IX.22), permite tambien justificar en una forma relativamente simple como los modelos deterministas necesitan ser complementados por modelos probabilísticos.

IX.15.- CLASIFICACION DE LOS ESTADOS DE LAS CADENAS DE MARKOFF. Conforme a lo visto en el párrafo IX.13, dado un estado de una cadena de Markoff caracterizada por una matriz de probabilidades de transición, queda inmediatamente determinado un conjunto de $n+1$ sucesiones de probabilidades de transición de orden superior ($n+1$ es el número de estados que constituye la cadena) y otro conjunto de $n+1$ sucesiones de probabilidades de alcanzar por primera vez un estado determinado.

De estos conjuntos, interesa seleccionar ciertas sucesiones particulares, que corresponden a la vuelta al estado considerado, y a la vuelta por primera vez. Estas sucesiones se identifican con los nombres de p r o b a b i l i d a d e s d e r e -

c u r r e n c i a (que corresponden a las $p_{jj}^{(t)}$) y de p r o b a b i l i d a - d e s d e r e c u r r e n c i a p o r p r i m e r a v e z (o sea las $f_{jj}^{(t)}$). Para simplificar la notación, se convendrá con identificar estas probabilidades con un sólo índice inferior, ya que no existe posibilidad de confusión.

Llamando f_j a la suma de la serie de las probabilidades de recurrencia por primera vez:

(9.30) $$f_j = \sum_{t=1}^{\infty} f_j^{(t)}$$

se tendrá una primera clasificación de los estados de una cadena de Markoff segun que f_j sea igual o menor que la unidad. En el primer caso, el estado se llamará p e r s i s t e n t e, (porque tiene probabilidad 1 de volver a aparecer, aunque como se verá más adelante, la interpretación intuitiva puede ser errónea), y en el segundo t r a n s i t o r i o (porque tiene una probabilidad a $1-f_j$, distinta de cero, de no volver a aparecer).

Recordando las convenciones $p_j^{(0)} = 1$ y $f_j^{(0)} = 0$, y construyendo las funciones generatrices $P(s) = P_j^{(t)} . s^t$ y $F(s) = f_j^{(t)} . s^t$, por aplicación inmediata del teorema fundamental de los sucesos recurrentes (ver párrafo IX.7) se obtiene el siguiente teorema: l a c o n d i c i ó n n e c e s a r i a y s u f i c i e n t e p a r a q u e u n e s t a d o s e a p e r s i s t e n t e (t r a n s i t o r i o) e s l a d i v e r g e n c i a (c o n v e r g e n c i a) d e l a s e r i e d e l a s $p_j^{(t)}$.

Es notable que con sólo este teorema y la relación central (9.29) se pueden obtener una gran parte de las propiedades importantes de las cadenas de Markoff.

IX.16.- El uso combinado del teorema demostrado en el párrafo IX.15, y de la relación central (9.29) permite demostrar que una cadena finita de Markoff no puede estar constituída únicamente por estados transitorios.

Como la serie de las probabilidades de $p_j^{(t)}$ es convergente, se verificará que da un ε positivo y arbitrario, habrá un t_1 tal que para todo $t > t_1$ se tendrá:

(9.31) $$p_j^{(t)} < \varepsilon$$

y mediante la relación central es posible probar que habrá también un t_2 tal que para $t > t_2$ se verifica:

(9.32) $$p_{ij}^{(t)} < \varepsilon$$

cualquiera sea el índice i.

En efecto, la sumatoria que figura en la relación central (9.29) se puede descomponer en otras dos utilizando un índice superior m_1:

(9.33) $$p_{ij}^{(t)} = \sum_{m=1}^{t} f_{ij}^{(m)} . p_j^{(t-m)} = \sum_{m=1}^{m_1} f_{ij}^{(m)} . p_j^{(t-m)} + \sum_{m=m_1+1}^{t} f_{ij}^{(m)} . p_j^{(t-m)}$$

y prescindiendo en la primer sumatoria de las $f_{ij}^{(m)}$, y de las $p_j^{(t-m)}$ en la segunda, que tratándose de probabilidades serán números positivos que no pueden superar la unidad, se tiene:

(9.34) $$p_{ij}^{(t)} \leqslant \sum_{m=1}^{m_1} p_j^{(t-m)} + \sum_{m=m_1+1}^{t} f_{ij}^{(m)}$$

La serie de las $f_j^{(m)}$ es siempre convergente, y la de las $p_j^{(t)}$ lo es también porque se trata de un estado transitorio. Por lo tanto, es posible elegir un número m_1 tal que el resto de la serie de las $f_j^{(m)}$ representado por la segunda sumatoria sea menor que $\varepsilon/2$, y fijado ese m_1, se podrá elegir un t_2 tal que la suma de los m_1 términos que integran la primera sumatoria sea inferior tambien a $\varepsilon/2$, con lo que se tendrá:

(9.35) $$p_{ij}^{(t)} < \varepsilon$$

para todo $t > t_2$. El número t_2 dependerá de i, pero como dicho índice tiene un campo finito de variación, basta tomar el t_2 mayor para asegurar que (9.35) se verifica con independencia de i, como se quería demostrar.

Si una cadena de Markoff estuviera formada solamente por estados transitorios, basta elegir un t_2 para que haciendo variar j todas las $p_{ij}^{(t)}$ que forman la fila i sean menores que $\varepsilon/n+1$ (siendo $n+1$ el número de éstados de la cadena) en cualquier matriz de probabilidades de transición de orden mayor que t_2, en la que t_2 se ha elegido como el mayor de los que para cada j verifican (9.35) debido a lo cual la suma de cualquiera de las filas será menor que ε, en contradicción con la propiedad de que tiene que ser necesariamente igual a la unidad, con lo que se ha probado la imposibilidad de la existencia de una cadena finita de Markoff con todos sus estados transitorios (la demostración no es válida para el caso infinito).

IX.17.- PROPIEDADES DE CLAUSURA DE LOS ESTADOS PERSISTENTES. Si desde un estado persistente E_j es posible alcanzar un estado E_k, habrá un índice t para el que ciertamente podrá escribirse:

(9.36) $$p_j^{(t)} = p_{jk}^{(x)} \cdot p_{kk}^{(y)} \cdot p_{kj}^{(z)}$$

siendo $x+y+z=t$ y las tres probabilidades del segundo miembro distintas de cero. En efecto, $p_{jk}^{(x)}$ no puede ser nula por hipótesis (sino el estado E_k no podría ser alcanzdo desde el E_j), y tampoco $p_{kj}^{(z)}$ (porque sino el estado E_j no sería persistente, ya que sino existiera un z con esa condición, una vez alcanzado el E_k el sistema no podría volver al E_j). Lo que necesita una condición adicional es demostrar que y no puede ser nulo (en cuyo caso el resultado sería trivial), para lo que hay que fijar que tanto x como z son la duración mínima para las transiciones que indican los subíndices y que t es mayor que la suma de esos mínimos, en cuyo

caso entonces tampoco $p_k^{(y)}$ puede ser nula para todo t (porque si $p_j^{(t)}$ fuera nula para todo t mayor que x+y mínimo, E_j no sería persistente, ya que la divergencia de la serie de las $p_j^{(t)}$ implica que posee infinitos términos no nulos).

Una vez fijados lo tiempos mínimos x, z, haciendo variar y varía tambien t y es evidente que el comportamiento de $p_k^{(y)}$ con respecto a la divergencia de la serie tiene que ser el mismo que el de $p_j^{(t)}$, ya que los términos de la primera sucesión difieren de los de la segunda solamente en un factor constante, llegándose a la importante conclusión que d e s d e u n e s t a d o p e r s i s t e n t e s o l a m e n t e s e p u e d e a l c a n z a r o t r o e s t a d o p e r s i s t e n t e.

El conjunto de estados que pueden alcanzarse desde un estado dado se denomina c l a u s u r a de ese estado, y con esta nomenclatura el resultado anterior se puede reformular diciendo que l a c l a u s u r a d e u n e s t a d o p e r s i s t e n t e e s t á f o r m a d a u n i c a m e n t e p o r e s t a d o s p e r s i s t e n t e s. Es evidente que los estados persistentes que integran la clausura de otro estado persistente t i e n e n t o d o s l a m i s m a c l a u s u r a (porque si alguno tuviera una clausura menor, los estados no incluidos en ella no serían persistentes, y si tuviera una clausura mayor, los estados de esa nueva clausura serían alcanzables desde los estados desde a los que su vez se alcanza al estado supuestamente excepcional).

En cambio, desde un estado transitorio se puede alcanzar un estado persistente, porque una demostración similar a la anterior para probar que el estado alcanzado sería también transitorio no es válida, ya que $p_{kj}^{(z)}$ podría ser nula sin incurrirse en ninguna contradicción y en general no podría escribirse una relación del tipo de la (9.31).

En el caso del paseo al azar con muros absorbentes (ver párrafo IX.12, por ejemplo (2)), cada uno de los estados extremos es persistente y su clausura está constituída por el mismo estado (un estado persistente con esta propiedad particular se llama a b s o r b e n t e). Los demás estados son transitorios, porque tienen una probabilidad distinta de cero de no volver a aparecer (que es por lo menos la de una trayectoria que incluya un estado persistente).

IX.18.- CLASIFICACION DE LOS ESTADOS PERSISTENTES. Dado un estado persistente, es posible definir la variable aleatoria tiempo de recurrencia (ver párrafo IX.10) y calcular su primer momento:

(9.37) $$\tau_j = \sum_{m=1}^{\infty} m . f_j^{(m)}$$

Si la serie que define dicho momento es divergente, el estado se llama n u l o y se conviene en decir que el tiempo medio de recurrencia es infinito(en realidad, el tiempo de recurrencia no puede ser infinito, y en un estado nulo lo único que ocurre es que el promedio de los tiempos de recurrencia supera cualquier número, es decir, que tiempos de recurrencia muy grandes no tienen una probabilidad pequeña, a diferencia de lo que ocurre con los estados no nulos, en los que el promedio de los tiempos de recurrencia converge en probabilidad al tiempo medio (9.38), lo que equivale a acotar la probabilidad de los tiempos de recurrencia muy grandes).

Si (9.37) es finita, puede ocurrir que el estado sea p e r i ó d i c o, o sea que las $p_j^{(t)}$ sean nulas salvo para un conjunto de valores de t que es múltiplo de un número (el período). En los ejemplos del párrafo IX.12, tanto el paseo al azar con muros reflectores como el modelo de Ehrenfest son ejemplos de cadenas de Markoff, con estados periódicos, porque siendo nula la diagonal principal de la matriz de probabilidades de transición, es evidente que no es posible volver a un estado dado sino en un número par de etapas, y el período sería igual a dos.

Si el estado pérsistente no es nulo ni periódico, se llama e r g ó d i c o. El paseo al azar con muros elásticos parcialmente absorbentes es un ejemplo de cadena de Markoff con estados ergódicos.

Tanto para los estados persistentes ergódicos como para los nulos vale para el comportamiento asíntótico de sus probabilidades de recurrencia, la siguiente relación:

(9.38) $$\lim_{t\to\infty} p_j^{(t)} = p_j = \frac{1}{\tau_j}$$

que para los periódicos se transforma en otra muy similar:

(9.39) $$\lim_{t\to\infty} p_j^{(\pi t)} = p_j = \frac{\pi}{\tau_j}$$

en la que π es el período.

Conforme a la relación (9.39), para los estados nulos se verifica que

(9.40) $$\lim_{t\to\infty} p_j^{(t)} = 0$$

o sea que los estados nulos corresponden al caso de divergencia de la serie de las $p_j^{(t)}$ con términos indefinidamente decrecientes.

Aplicando la demostración utilizada en el párrafo IX.16 para los estados transitorios, se tiene que también para los nulos vale:

(9.41) $$\lim_{t\to\infty} p_{ij}^{(t)} = 0$$

cualquiera sea el subíndice i .

Utilizando el mismo método empleado para probar las propiedades de clausura de los estados persistentes, resulta inmediatamente que desde un estado persistente es posible alcanzar únicamente un estado persistente, de su misma clase, ya que las $p_j^{(t)}$ y $p_k^{(y)}$ difieren solamente en una constante, y tendrán por lo tanto el mismo compor-

tamiento asintótico (simultáneamente un límite nulo, distinto de cero o un comportamiento periódico). En otros términos, la clausura de un estado persistente está formada por estados persistentes de su misma clase.

Una consecuencia inmediata es que en una cadena finita de Markoff no pueden existir estados persistentes nulos. En efecto, en la matriz de probabilidades de transición (de cualquier orden t), en la fila que corresponde a un estado nulo seran iguales a cero todas las probabilidades ubicadas en columnas correspondientes a estados no nulos, y seran distintas de cero únicamente las correspondientes al pasaje a estados nulos de su clausura. Pero como para cada estado nulo vale la relación (9.41), es evidente que para un indice t suficientemente grande la suma de las probabilidades de la fila de un estado nulo es menor que ε (igual demostración que para el caso de una cadena que estuviera formada únicamente por estados transitorios, en el párrafo IX.16), en contradicción con la necesidad de que la suma sea igual a la unidad.

IX.19.- CADENAS IRREDUCIBLES. PROBABILIDADES ESTACIONARIAS. Una cadena de Markoff se llama i r r e d u c i b l e cuando la clausura de cualquiera de sus estados coincide con el conjunto de estados que constituyen la cadena (también suele llamarse t r a n s i t i v a). En el párrafo IX.12 todos los ejemplos, salvo el del paseo al azar con muros absorbentes, corresponden a cadenas irreducibles, porque desde un estado de la cadena se puede alcanzar cualquier otro en un número finito de etapas.

Evidentemente, una cadena irreducible sólo puede estar formada por estados persistentes (si hubiera estados transitorios, como la totalidad de los estados no puede pertenecer a esa categoría, habría tambien estados persistentes, cuya clausura no incluiría a los transitorios, y la cadena no respondería a la definición de irreducibilidad), y como no es posible que existan estados nulos, y todos deben ser de igual clase, se concluye que los estados pueden ser unicamente o todos ergódicos, o todos periódicos, lo que constituye el primer teorema de las cadenas irreducibles.

Las cadenas irreducibles poseen la importantísima propiedad de que, asintóticamente, la probabilidad de alcanzar un estado resulta independiente del estado inicial o sea que se verifica:

(9.41) $$\lim p_{ij}^{(t)} = p_j$$

generalización de la relación (9.33) probada para los estados transitorios, y extendida sin ninguna variante a los estados nulos.

La demostración depende también de una partición adecuada de la relación central (9.29):

(9.43) $$p_{ij}^{(t)} = \sum_{m=1}^{t} f_{ij}^{(m)} \cdot p_j^{(t-m)} = \sum_{m=1}^{m_1} f_{ij}^{(m_1)} \cdot p_j^{(t-m)} + \sum_{m=m_1+1}^{t} f_{ij}^{(m)} \cdot p_j^{(t-m)}$$

La serie $f_{ij} = \sum_{m=1}^{\infty} f_{ij}^{(m)}$ es siempre convergente (ver párrafo IX.13), y por la propiedad enunciada sin demostración en el párrafo anterior, $\lim_{t\to\infty} p_j^{(t)} = p_j$. Por lo tanto, prescindiendo de los $p_j^{(t-m)}$ en la segunda sumatoria, es posible fijar un m_1 tal que el resto de la serie de las $f_{ij}^{(m)}$ sea menor que $\varepsilon/2$ y un t_1 tal que para todo $t > t_1$ se verifique $p_{ij}^{(t-m)} < p_j + \varepsilon/2f_{ij}$ si m es menor que m_1, en la primer sumatoria, con lo que se obtiene en definitiva:

$$p_{ij}^{(t)} \leqslant (p_j + \frac{\varepsilon}{2f_{ij}}) \sum_{m=1}^{m_1} f_{ij}^{(m)} + \frac{\varepsilon}{2} \tag{9.44}$$

Como para cada m_1 se puede fijar el t_1 necesario para la validez de (9.43), en el límite para $m_1 \to \infty$ se tendrá:

$$\lim_{t\to\infty} p_{ij}^{(t)} \leqslant (p_j + \frac{\varepsilon}{2f_{ij}}) f_{ij} + \frac{\varepsilon}{2} = p_j \cdot f_{ij} + \varepsilon \tag{9.45}$$

Volviendo a (9.42), la segunda sumatoria es siempre positiva (suma de productos de probabilidades), de manera que si se prescinde de ella, y en la primera se elige un t suficientemente grande como para que $p_{ij}^{(t-m)} > p_{ij} - \frac{\varepsilon}{f_{ij}}$, se tendrá:

$$p_{ij}^{(t)} \geqslant (p_j - \frac{\varepsilon}{f_{ij}}) \sum_{m=1}^{m_1} f_{ij}^{(m)} \tag{9.46}$$

pasando al límite:

$$\lim_{t\to\infty} p_{ij}^{(t)} \geqslant (p_j - \frac{\varepsilon}{f_{ij}}) f_{ij} = p_j f_{ij} - \varepsilon \tag{9.47}$$

Como la cadena es irreducible, $f_{ij} = 1$, y finalmente, combinando (9.45) y (9.47) y reemplazando f_{ij} por su valor:

$$\lim_{t\to\infty} p_{ij}^{(t)} = p_j \tag{9.48}$$

como se quería demostrar.

Por consiguiente, se obtiene el segundo teorema: en un sistema representado por una cadena irreducible de Markoff, después de transcurrir un tiempo suficientemente largo, es posible prescindir de la distribución estacionaria, ya que siendo la probabilidad de hallar al sistema en un estado E_j igual a $\sum P_i p_{ij}^{(n)}$ (segun (9.20)), al pasar al límite para n infinito se obtiene p_j.

Si la cadena es periódica, es necesario precisar el significado de la distribución estacionaria. En toda cadena periodica, de período π es posible dividir el conjunto C de estados en t subconjuntos disjuntos C_i, tales que desde un estado comprendido en el subconjunto C_i es posible pasar en un etapa solamente a uno o varios del subconjunto C_{i+1}, en orden cíclico (en el modelo de Ehrenfest, de período $\pi = 2$ existen dos subconjuntos formados, respectivamente, por los estados de índice par e impar, y desde un estado par se puede alcanzar en una etapa solo un estado impar, y re

cíprocamente).

Las probabilidades de transición $p_{ij}^{(t)}$ son distintas de cero, para t múltiplo del período, si E_i y E_j pertenecen al mismo subconjunto. Dentro de cada subconjunto, la suma de las probabilidades de orden $t = n\pi$ es igual a la unidad (las otras son nulas) y el límite de cada una de ellas es π/μ_j

Como hay π subconjuntos, la suma de las inversas de los tiempos medios de recurrencia tambien es unitaria y definen por lo tanto una distribución estacionaria.

IX.20.- TEOREMA ERGODICO. Dado que a partir de un estado inicial es posible definir unívocamente la sucesión de probabilidades $p_{ij}^{(t)}$, si se adopta una variable aleatoria que toma el valor uno cuando aprece el estado j, y cero cuando aparece otro estado, la frecuencia relativa de aparición verifica, por aplicación del teorema de Poisson (ver V.11):

(9.49) $$\lim_{t\to\infty} \left(\left| \frac{x_{ij}}{t} - \frac{\sum p_{ij}^{(t)}}{t} \right| > \varepsilon \right) = 0$$

en donde x_{ij} es el número de apariciones del estado E_j en una trayectoria de longitud t que comienza con el estado E_i. Pero como por aplicación de un conocido teorema de análisis, la suma promediada de una sucesión que tiende a un límite p_j, tiene ese mismo límite, se verificará:

(9.50) $$\lim_{t\to\infty} \left(\left| \frac{x_{ij}}{t} - p_j \right| > \varepsilon \right) = 0$$

y como p_j es independiente del estado inicial E_i, se tiene el teorema ergódico de las cadenas estacionarias de Markoff: l a f r e c u e n c i a r e l a t i v a d e l a s a p a r i c i o n e s d e u n e s t a d o e n u n a c a d e n a e s t a c i o n a r i a (o i r r e d u c i b l e), c o n v e r g e e n p r o b a b i l i d a d a u n a c o n s t a n t e i n d e p e n d i e n t e d e l e s t a d o i n i c i a l.

IX.21.- APLICACION A LA ESTRUCTURA ESTADISTICA DE UN LENGUAJE. Un lenguaje puede ser representado, al menos en forma aproximada, por una cadena irreducible de Markoff de primer orden, ya que es evidente que dada una letra, en un texto suficiente largo se alcanzará siempre otra cualquiera del alfabeto, o sea que la clausura de una letra es toda la cadena. Esta cadena evidentemente no es periódica, de manera que por aplicacion del teorema ergódico, la frecuencia relativa de aparición tenderá a un límite igual a la inversa del tiempo medio de recurrencia, con independencia de cual sea la letra inicial del texto.

Esta constancia de la frecuencia relativa de aparición de las letras (y del espacio en blanco) en un idioma, cualquiera sea el comienzo del texto, es un hecho descubier

to experimentalmente por los criptógrafos hace varios siglos, y empleado para la construcción de claves y su descifre (b).

A título ilustrativo, se dan las frecuencias reltivas de aparición (expresadas en %) de los signos comunes de algunos idiomas corrientes(c):

(1)Castellano: Espacio en blanco:17%.- E ó A:11%.- I u O:7%.- N ó S:6%.- L ó R:5%.

(2)Inglés: Intervalo:17%.- E:10%.- T:8%.- A:7%.- I ó S:6%.- O, N y R:5%.

(3)Francés: Intervalo:17%.- E:16%.- S:8%.- A, R ó N:6%.- I, T, U, L, O:5%.

(4)Alemán: Intervalo:14%.- E:14%.- N:9%.- A, S, I, R:6%.- D ó T:5%.- V:4%.

Dadas las diferencias, se comprende como es un dato fundamental para los criptógrafos el conocer el idioma en que está escrito un texto en clave antes de intentar su descifre (b) , y recíprocamente, por qué en algunos casos se han resucitado lenguas muertas para usarlas en la transmisión de mensajes especialmente importantes que requerían gran seguridad de cifrado (caso del Navajo, en la segunda guerra mundial).

Dado que la frecuencia de aparición corresponde a la inversa del tiempo medio de recurrencia, una vez conocida la frecuencia del intervalo o espacio en blanco, es posible deducir la longitud media de las palabras. Así en el castellano será de cuatro letras y fracción, aproximándose a cinco, mientras que en el alemán es de más de seis letras (hay que recordar que al computar el tiempo de recurrencia se cuenta también la aparición del signo, de manera que a la inversa de la frecuencia relativa del intervalo hay que restar la unidad para obtener la longitud media de las palabras.

La representación del idioma (más generalmente,de una fuente de información) mediante una cadena de Markoff, permitió a Shannon desarrollar su famosa teoría matemática de la información, que resuelve los problemas mencionados en el párrafo III.14, y que cada día encuentra aplicaciones más extendidas (a).

IX.22.- APLICACIONES A LA FISICA. Tambien mediante las cadenas finitas de Markoff es posible explicar el papel que desempeña el cálculo de probabilidades en la física.

A principios de siglo se produjo una polémica, que luego se hizo célebre, entre Zermelo y Boltzmann. El primero invocó un teorema rigurosamente demostrado por Poincaré, segun el cual todo sistema dinámico conservativo tiene la propiedad de que sus estados vuelven a presentarse, después de un cierto tiempo de retorno, propio de cada esta

(a) Una exposición rigurosa de la teoría de la información,primero en forma simplificada (para fuentes discretas), y después más general puede verse en A. I. KHINCHIN: "Mathematical Foundations of Information Theory", New York, Dover, 1957.

(b) Una introducción sumaria, pero bastante ilustrativa, de los métodos de construccion y descifre de claves, puede verse en R. CEILLIER: "La Crytographie", Presses Universitares de France (Collección Que Sais-je? Nº 116), Paris, 1945.

(c) Como toda determinación experimental de probabilidades, las frecuencias relativas presentan oscilaciónes, y los datos suelen presentar pequeñas diferencias segun los autores. En los ejemplos, se ha seguido a G. Guilbaud (obra citada en III.14)

do (este es un enunciado muy simplificado del teorema de Poincaré), lo que significaba una contradicción con el comportamiento irreversible de los sistemas termodinámicos que estudiaba el segundo.

La respuesta de Boltzmann a Zermelo no niega la validez del teorema del ciclo de Poincaré, sino que afirma que todo estado tiene un probabilidad de ser observado, y que a probabilidades muy pequeñas corresponden tiempos de retorno muy largos, de manera que aunque la irreversibilidad no es teóricamente imposible, en la práctica es inobservable. Pero Boltzmann no pudo demostrar rigurosamente sus afirmaciones.

En el caso especial de que el sistema físico puede ser representado por una cadena de Markoff, finita e irreducible, la respuesta de Boltzmann queda demostrada por la existencia de una distribución estacionaria (párrafo IX.19), con la única salvedad de que las probabilidades de observación que se atribuyen a los estados deben interpretarse únicamente como límites, después de transcurrido un tiempo suficientemente largo de evolución del sistema. Los tiempos medios de recurrencia tienen una relación inversa con dichas probabilidades límites, y el teorema ergódico aclara la significación de la impracticabilidad del comportamiento reversible, ya que los estados con pequeña probabilidad y tiempos de reversibilidad muy grande presentan una frecuencia relativa de aparición muy pequeña.

Naturalmente, para considerar levantada completamente la objeción de Zermelo, es necesario, entre otros puntos, demostrar rigurosamente el teorema ergódico, para sistemas bastante más generales que las cadenas finitas de Markoff, lo que recien se logró treinta años más tarde de abierta la polémica entre Zermelo y Boltzmann (a).

IX.23.- La discusión anterior puede ejemplificarse mediante el estudio del modelo de Ehrenfest, descripto en el párrafo IX.12 (ejemplo (3)).

Si en este modelo el estado inicial consiste en la concentración de las n moléculas de una cierta masa de un gas en uno de los dos recipientes, el segundo principio de la termodinámica establece un sentido de difusión hacia el recipiente inicialmente vacío, hasta que se alcanza el equilibrio térmico, con n/2 moléculas en cada uno. La vuelta al estado inicial sería imposible.

Sin embargo, si se examina cuidadosamente el modelo, se advierte que no existe en él nada que impida la reversibilidad que establece el teorema del ciclo de Poincaré.

(a) Las primeras demostraciones (que corresponden a casos distintos)m son las siguientes: G. D. BIRKHOFF: "Proof of the Ergodic Theorem" Proc. Nat. Acad. of USA, vol 17 (1931)pag. 650-660; J. V. NEUMANN: "Proof of the quasi ergodic hypothesis", Id. vol 18 (1932), pag 70-82; T.CARLEMAN: "Applications de la théorie des equations integrales linéaires aux equations differentiels non linéaires", Acta Mat. , vol 59 (1932) pag. 63-87. Dada la importancia que tiene la teoría ergódica, ya sea para la mecánica estadística o como rama de las matemáticas, existe una extensa literatura especializada. Un resumen del estado actual de la cuestión puede verse en LOEVE (18), Cap. IX.

Pero la solución de la cuestión se halla en la determinación de la distribución estacionaria, que existe cualquiera sea n porque la cadena es irreducible (el hecho de que, además, sea periódica no introduce, en este caso, ninguna dificultad particular).

Después de transcurrido un tiempo suficientemente largo, la probabilidad de un estado se puede admitir que coincide con su límite, pero para el pasaje de un estado a otro vale siempre la matriz de probabilidades de transición (9.24), lo que permite establecer la siguiente relación de recurrencia entre las probabilidades estacionarias:

(9.51) $$p_j = (1 - \frac{j-1}{n})p_{j-1} + \frac{j+1}{n} p_{j+1}$$

que no es más que el resultado que se obtiene al pasar al límite para $t \to \infty$ en la expresión (9.27), dada la forma especial que en esta cadena tienen las probabilidades p_{kj}.

Para los estados extremos (E_o y E_n), la relación (9.50) se reduce a:

(9.51) $$p_o = \frac{1}{n} p_1 \quad ; \quad p_n = \frac{1}{n} p_{n-1}$$

y como el conjunto de los estados forma un sistema completo de sucesos excluyentes, debe verificarse además:

(9.53) $$\sum_{j=0}^{n} p_j = 1$$

condicion que implica eliminar el período, y que la distribución limite será de las inversas de los tiempos medios (ver párrafo IX. 19, final).

Utilizando (9.51) y (9.52), es posible expresar todas las p_j en función de p_o, la que se despejará a partir de (9.53).

Para p_1 se obtiene directamente:

(9.54) $$p_1 = np_o$$

Para p_2, se reemplaza el valor de p_1 en la relación de recurrencia correspondiente a p_1:

(9.55) $$np_o = p_o + \frac{2}{n} p_2 \quad ; \quad p_2 = \frac{n(n-1)}{2} p_o$$

multiplicando numerador y denominador por $(n-2)!$ resulta finalmente:

(9.56) $$p_2 = \binom{n}{2} p_o$$

Las expresiones (9.54) y (9.56) indican que la relación de dependencia entre p_j y p_o es posiblemente de la forma $p_j = \binom{n}{j} p_o$. Como ya se lo ha probado para dos valores de j, es necesario demostrar que si se verifica para j se verifica para $j+1$, con lo que quedaría demostrado para todo j, lo que se realiza utilizando nuevamente (9.51):

(9.57) $$\binom{n}{j} p_o = (1 - \frac{j-1}{r}) \binom{n}{j-1} p_o + \frac{j+1}{n} p_{j+1}$$

$$\therefore \quad p_{j+1} = \left(\frac{n}{j+1} \binom{n}{j} - \binom{n}{j-1} + \frac{j-1}{n} \binom{n}{j-1} \right) p_o$$

Reemplazando los coeficientes binomiales por su expresión explícita en función de factoriales, y después de algunas simplificaciones, el paréntesis de (9.57) resulta igual a $\frac{(n-1)!}{(n-j-1)!}$, con lo que se ha demostrado que para cualquier índice j:

(9.58) $$p_j = \binom{n}{j} p_o$$

Para determinar p_o se parte de la condición (9.53):

(9.59) $$\sum p_j = p_o \sum \binom{n}{j} = 1$$

y la suma de los coeficientes binomiales es igual a 2^n, lo que resulta de que en:

(9.60) $$(p+q)^n = \sum \binom{n}{j} p^j q^{n-j}$$

si $p+q=1$, el único sistema de valores para el que $p^j q^{n-j}$ es constante e independiente de j es $p=q=\frac{1}{2}$, con lo que se obtiene la expresión final buscada de las probabilidades estacionarias:

(9.61) $$p_j = \binom{n}{j} \frac{1}{2^n}$$

o sea que la distribución estacionaria de las inversas de los tiempos medios de recurrencia es la misma que resultaría de asignar probabilidad $\frac{1}{2}$ a cada molecúla para estar en uno de los dos recipientes.

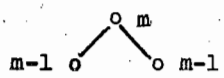
Una vez determinada esta distribución estacionaria (por un método que es en general aplicable para cualquier cadena finita de Markoff), es posible entonces discutir el equilibrio termodinámico y la irreversibilidad en la forma en que lo hace FELLER (4).

La distribución (9.61) es simétrica, por lo que la probabilidad máxima corresponderá a $n/2$. Este máximo, sin embargo, será muy pequeño si n es grande, y lo que importa probar es que en un intervalo que contiene a $n/2$ y de muy pequeña magnitud respecto a n se concentra casi toda la probabilidad, propiedad asintótatica característica de la ley binomial (ver párrafo IV.13), precisada por la fórmula de Laplace (ver párrafo VI.5).

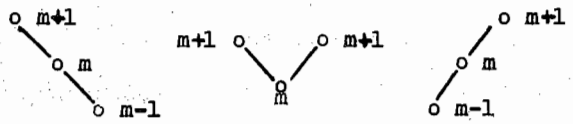
Dado que para calcular la probabilidad de un desvío con respecto a $n/2$, hay que dividir por $\sqrt{npq}$, si $n=10.000$, $p=q=1/2$ $\sqrt{npq}=50$, y como la probabilidad de que el desvío reducido supere el valor absoluto 4,42 es inferior a 0,000 001, resulta que después de transcurrido un cierto tiempo, el sistema se encontrará con probabilidad superior a 0,999 99 en un estado tal que el número de moléculas de gas en cada recipiente estará comprendido entre 4.779 y 5.221. Como la concentración relativa de la probabilidad aumenta con n, y el número de moléculas de una masa gaseosa se cuenta por millones (el número de Avogadro es del orden de 10^{23}), la probabilidad de estados tan alejados del equilibrio que sean detectables por observación de un desequilibrio térmico es pequeñisima, y su frecuencia relativa de observación practicamente nula.

Mediante el modelo probabilístico de las cadenas finitas de Markoff, ha sido posible completar el modelo determinista de Poincaré, precisando su significado y aclarando lo que debe entenderse por irreversibilidad.

IX.24.- La discusión original del modelo de Ehrenfest se orienta de una manera distinta de la aquí indicada. Suponiendo un estado E_m que corresponde a una distribución fuertemente asimétrica (m cercano a n), se calcula la probabilidad de los segmentos de trayectoria -t-1, t, t 1, en los que en el tiempo t el sistema se encuentra en el estado E_m y se prueba que la que corresponde a la sucesión de estados E_{m-1}- E_m - -E_{m-1} , que simbólicamente se identifica



es abrumadoramente mayor que la de las otras posibles



El significado epistemológico de los teoremas límites a que se aludió en el párrafo V.13 se ha transformado, debe adimitirse que un tanto sorpresivamente, en la explicación del concepto de irreversibilidad. Sin el cálculo de probabilidades, hubiera sido imposible contestar la objeción de Zermelo, y todo el campo de los fenómenos irreversibles escaparía a la tentativa de una descripción matemática.

Naturalmente, esta discusión es sólo esquemática, y deja de lado importante problemas de fundamentación(a).

IX.25.- CADENAS REDUCIBLES. Cuando en una cadena de Markoff existen estados transitorios, deben también existir estados persistentes, puesto que se ha demostrado que es imposible que todos los estados sean transitorios.

Si E_j es un estado persistente, y E_i un estados transitorio, por el mismo

(a) Un desarrollo elemental, sistemático y prolijo, de las aplicaciones físicas del cálculo de probabilidades, y posiblemente uno de los más claros y completos publicados hasta ahora, a pesar de su anterioridad a la demostración de los teoremas ergódicos, está contenido en la parte cuarta del tratado de V. Mises (20). Para una discusión mas detallada del modelo de Ehrenfest se puede consultar, entre otros M. KAC:"Probability and Related Topics in Physical Sciences", Interscience Publishers, New York, 1959.

Una dilucidación de los conceptos básicos de la mecánica estadística, con especial atencion al cálculo de probabilidades, puede verse en: P. und T. EHRENFEST:" Begriffliche Grundlagen der Statischen Mechanik" (Encyclop.d.mathWissench., Vol. IV. art. 32), Berlín, 1909.

Finalmente, para un tratamiento de la mecánica estadística con los recursos que modernamente ha provisto el cálculo de probabilidades (teorema ergódico y teorema central del límite); A. I. KHINCHIN:"Mathematical Foundations of Statistical Mechanics", New York, Dover, 1949.

método utilizado en el párrafo IX.19 se demuestra:

(9.62) $$\lim_{t\to\infty} p_{ij}^{(t)} = p_j f_{ij}$$

en donde p_j es la probabilidad estacionaria del estado persistente, y f_{ij} la suma de la serie de las probabilidades de alcanzar por primera vez el estado persistente a partir del transitorio, y que a diferencia de lo demostrado en IX.19 no puede ser igual a la unidad si desde el estado transitorio se pueden alcanzar distintos estados persistentes.

En una cadena de este tipo, una trayectoria se puede iniciar en un estado transitorio, pero en cuanto alcanza un estado persistente, permanecerá dentro de la clausura del persistente y no volverá a aparecer ningun estado transitorio, lo que explica la denominación de cadenas reducibles, puesto que el conjunto de estados observables se ha reducido.

El ejemplo del paseo al azar con muros absorbentes (o lo que es equivalente, el problema de la ruina de los jugadores), consiste precisamente en una cadena reducible, y los dos estados persistentes no pertenecen a la misma clausura, por lo que un problema importante es calcular la p r o b a b i l i d a d d e a b s o r c i ó n, o sea, calcular la probabilidad de alcanzar una clausura dada a partir de una estado transitorio tambien dado, la que es distinta, en general, para cada clausura, y para lo que es necesario emplear métodos algebraicos fundados en la estructura particular de las matrices de probabilidades de transición.

Sin entrar en el detalle de los métodos de cálculo de las probabilidades de absorción, es interesante probar que la suma de las probabilidades de absorción a los diversos estados absorbentes o clausuras de estados persistentes, es siempre igual a la unidad, o sea que la probabilidad de una trayectoria compuesta únicamente por estados transitorios es nula (en el caso del problema de la ruina de los jugadores, como los estados absorbentes corresponden a la ruina de un jugador y a la interrupción del juego, el teorema se traduce en la afirmación de que el juego tiene probabilidad cero de continuar indefinidamente, y que con probabilidad uno se arruinará algun jugador)

Para ello conviene efectuar una partición de la matriz de probabilidades de transición, previa una reordenación de sus filas y columnas. Cada clausura de estados persistentes será una submatriz cuadrada, (porque dentro de esa clausura cada estado puede ser alcanzado desde otro) y la prolongación de sus filas seran todos términos nulos porque de un estado de la clausura no se puede pasar a otra clausura ni a un transitorio. En particular, la submatriz correspondiente a un estado absorbente será de primer orden.

Si se supone que existen dos clausuras de estados persistentes, la matriz particionada tendrá, despues del reordenamiento la siguiente estructura:

$$(9.63) \qquad |P| = \begin{vmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ T_1 & T_2 & T \end{vmatrix}$$

P_1 y P_2 seran submatrices cuadradas correspondientes a las clausuras de estados persistentes. T será la submatriz cuadrada correspondiente a las transiciones de estados transitorios entre si, y T_1 y T_2 las submatrices (en general rectangulares) de las probabilidades de transición de los estados transitorios a los estados persistentes.

Al formar potencias sucesivas de esta matriz, tendrán en general la siguiente estructura:

$$(9.64) \qquad |P|^t = \begin{vmatrix} P_1^t & 0 & 0 \\ 0 & P_2^t & 0 \\ T^{(t)} & T^{(t)} & T^t \end{vmatrix}$$

en donde P_1^t, P_2^t y T^t seran efectivamente las potencias t-ésimas de las submatrices respectivas, y $T_1^{(t)}$ y $T_2^{(t)}$ tendran una estructura complicada que no interesa precisar en detalle, a los fines de esta demostración.

Cuando t crece indefinidamente, los términos de $T_1^{(t)}$ y $T_2^{(t)}$ tienden a límites del tipo $p_j f_{ij}$, como se indicó anteriormente. Los términos de T^t, por lo demostrado en IX.17 tienden a cero, y como P^t es una matriz estocástica, es claro que la suma de las probabilidades de absorción tiene por límite la unidad, porque la suma de los terminos de cada fila que pertenecen a T^t tiende a cero.

Un segundo problema interesante, es calcular la ley de probabilidades de la variable t i e m p o d e a b s o r c i ó n, o sea la probabilidad de que la cadena persista en estados transitorios, pasando exactamente en el tiempo fijado a un estado persistente (en el caso de la ruina de los jugadores, corresponde a la probabilidad de la duración del juego), y que tambien se resuelven por métodos algebraicos que dependen de la estructura particular de las matrices de probabilidades de transición.

Desde el punto de vista del comportamiento asintótico de las trayectorias puede prescindirse, por lo que se ha visto, de los estados transitorios. El cálculo de las probabilidades de absorción se ha realizado en muchos casos utilizando modelos distintos del de las cadenas de Markoff, pero tiene interés la reconsideración de los problemas clásicos con este algoritmo, porque permite introducir los recursos del algebra de matrices, a menudo más eficientes que los métodos de integración de ecuaciones en diferencias finitas, o que el empleo de la función generatriz, mediante el cual Laplace resolvió un gran número de ellos.

T E R C E R A P A R T E

T E O R I A D E L A S V A R I A B L E S A L E A T O R I A S C O N T I N U A S

CAPITULO X. VARIABLES ALEATORIAS UNIDIMENSIONALES.

X. 1.- Es frecuente el caso de que un fenómeno aleatorio consista en un conjunto de sucesos de potencia igual a la del continuo. Esto ocurre, por ejemplo, cuando el resultado de una experiencia es una medición, que, al menos teoricamente, puede adoptar cualquier valor dentro de un intervalo finito o infinito.

En la práctica, dado que todo procedimiento de medición posee una precisión más allá de la cual es imposible llegar, podría pensarse que bastaría con la consideración de variables aleatorias numerables, pero la misma teoría de las variables aleatorias discretas no queda completada sin el uso de variables aleatorias continuas, que resultan introducidas naturalmente por el teorema de De Moivre y Laplace.

X.2.- En la caracterización de los fenómenos aleatorios, de que se hizo uso en el cap. I, se dirá que se trata de un fenómeno aleatorio continuo cuando la frecuencia relativa de las observaciones en un intervalo arbitrario de una variable continua, tiende a ser constante al aumentar indefinidamente el número de observaciones.

Para construir un modelo matemático adecuado, sería necesario poder definir una función aditiva de conjunto o medida(que a cada conjunto asigna un número no negativo, y que a la unión de dos conjuntos no rampantes asigna la suma de los números correspondientes), y esta definición tendría que ser aplicable no solamente a intervalos sino a conjuntos cualesquiera, o por los menos a familias muy amplias de conjuntos, y que desde luego contengan a los intervalos.

Aunque el problema es difícil, ha sido resuelto con suficiente generalidad mediante la teoría de la medida, que se trata en la teoría de funciones de variable real (a).

Como en este curso no se presupone el conocimiento de la teoría de la medida, con los refinamientos especiales necesarios para su uso por el cálculo de probabilidades, el tratamiento de las variables aleatorias continuas se limitará a los casos más simples en que los conjuntos son solamente intervalos y la función aditiva tiene propiedades sencillas.

X.3.- Una medida aplicable a intervalos, o segmentos, no tiene porque ser homogénea

(a) En algunos textos de cálculo de probabilidades, como CRAMER (2) GNEDENKO KOLMOGOROFF (10) y LOEVE (18) se encuentran resúmenes de la teoría de la medida necesaria.

es decir, no tiene por qué asignar igual medida (en adelante se dirá probabilidad) a segmentos de igual longitud.

El problema queda resuelto si es posible construir una función monótona creciente $F(x)$, tal que su límite para $+\infty$ es la unidad, y para $-\infty$ es cero. A cualquier intervalo finito le corresponde como probabilidad, la diferencia de los valores de la función monótona creciente en los extremos del intervalo.

Una función monótona creciente puede tener discontinuidades, pero esas discontinuidades seran únicamente de primera especie, es decir, la función tiene en todo punto un límite a la derecha y otro a la izquierda. En un punto de discontinuidad, la diferencia entre los dos límites, o sea el salto de la función, será la probabilidad finita asignada al valor correspondiente de la variable aleatoria.

Con esta definición, la atribución de probabilidades a un segmento se completa conviniendo en tomar como valor de la función el límite a la derecha, con lo que la diferencia de valores en los extremos corresponde al intervalo semicerrado a la derecha. La unión y la intersección (en número finito) de intervalos semicerrados a la derecha tiene por resultado conjuntos del mismo tipo, por lo que la aplicación a esta clase de segmentos de las operaciones entre sucesos, conduce a conjuntos para los que las probabilidades correspondientes continúan siendo definidas. La cuestión se complica cuando el número de las operaciones a realizar con los sucesos es infinita (a).

Las funciones monótonas crecientes acotadas pueden presentar sólo un número finito o infinito numerable de discontinuidades de primera especie, y por lo tanto existiran siempre puntos de continuidad (que formarán un conjunto con la potencia del continuo). En dichos puntos de continuidad, la función monótona creciente no asigna una probabilidad finita a la variable aleatoria.

X.4.- Toda función que cumpla con las propiedades anteriormente indicadas:

(10.1) $\Delta x . \Delta F(x) \geqq 0$; $\lim_{x \to +\infty} F(x) = 1$; $\lim_{x \to -\infty} F(x) = 0$

se llama f u n c i o n d e d i s t r i b u c i ó n, y la convención sobre su valor permite asignarle, como se ha dicho, la siguiente interpretación en términos de probabilidades:

(10.2) $F(a) = P(x \leqslant a)$

$F(b) - F(a) = P(a \leqslant x \leqslant b)$ si $F(x)$ es continua en a ;

$F(b) - F(a) = P(a < x \leqslant b)$ si $F(x)$ es discontinua en a ;

o sea, que la función de distribución caracteriza la ley de probabilidades.

La ley de probabilidad de una variable aleatoria binomial, dada por (4.1), puede interpretarse como definiendo los saltos de una función de distribución, que es discon

(a) Ver nota en el párrafo anterior.

tinua en los $n+1$ puntos de abcisa entera comprendidos entre 0 y n, y constante entre los puntos de discontinuidad. La función de distribución así construída constituye lo que se llama una f u n c i ó n e s c a l e a.

La ley de Poisson posee una función de distribución con infinito numerable puntos de discontinuidades. La fórmula de Laplace define, en cambio, una función de distribución, que carece de discontinuidades:

$$P(x \leqslant a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{t}{2}} dt = F(a) \tag{10.3}$$

X.5.- La derivada de una función de distribución, cuya integral en un intervalo de continuidad es igual al incremento de la función de distribución, y por lo tanto a la probabilidad de que la variable aleatoria tome valores en ese intervalo, se llama f u n c i ó n d e d e n s i d a d d e p r o b a b i l i d a d, y dado el caracter monótono creciente de su primitiva será siempre n o n e g a t i v a.

Cuando las funciones de distribución de variables aleatorias continuas no presentan discontinuidades, es cómodo identificarlas mediante sus funciones de densidad de probabilidad, a condición de no olvidar que, en la eventualidad de un cambio de variables, hay que tener presente también el elemento diferencial, sin lo cual no quedaría invariante la integral sobre un intervalo, o sea la probabilidad.

En lo sucesivo, se consideraran solamente variables aleatorias continuas con su ley de probabilidad definida de esta manera, dejando de lado el problema más general de las funciones de distribución discontinuas.

X.6.- Partiendo de la ley de Poisson (7.4), es posible deducir inmediatamente la ley de probabilidad de la variable aleatoria continua t i e m p o t r a n s c u r r i d o h a s t a l a r e a l i z a c i ó n d e u n a o b s e r v a c i ó n.

La probabilidad de efectuar al menos una observación en el intervalo de tiempo ΔT será (poniendo en evidencia el tiempo ΔT):

$$1 - e^{-\lambda \Delta T} \tag{10.4}$$

que es una función de distribución continua que toma valores no nulos a partir del origen tiempo, punto en el cual es nula (en otros términos, está definida solamente, para valores positivos de la variable tiempo), porque es monótona creciente (el término substractivo disminuye al crecer ΔT), su límite para $\Delta T \to +\infty$ es la unidad y para $\Delta T = -\infty$ es cero (haciendo la convención de que es identicamente nula para todo valor negativo de ΔT).

Derivando respecto de ΔT, se tiene:

(10.5) $$f(\Delta T) = \lambda e^{-\lambda \Delta T}$$

que es la función de densidad de probabilidad.

X.7.- Otra ley de probabilidad que se deduce facilmente es la correspondiente al cuadrado de una ley normal:

(10.6) $$P(x^2 \leqslant y) = F(+\sqrt{y}) - F(-\sqrt{y})$$

derivando la expresión, se tiene:

(10.7) $$f(y) = \frac{1}{2\sqrt{y}}\left(f(+\sqrt{y}) + f(-\sqrt{y})\right)$$

Teniendo en cuenta que el integrando de (10.3) es $\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$, y reemplazando en (10.7):

(10.8) $$f(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$$

que es la función de densidad de probabilidad, conocida como l e y d e J i c u a d r a d o y que tiene gran importancia en inferencia estadística (como se verá en el cap. XIV).

X.8.- Más generalmente, cualquier función n o n e g a t i v a e integrable en todo su campo de definición, puede ser adoptada como función de densidad de probabilidad, a c o n d i c i ó n d e n o r m a l i z a r l a d i v i d i e n d o l a p o r e l v a l o r d e s u i n t e g r a l, con lo que las funciones de distribución y de densidad tendran la forma siguiente:

(10.9) $$F(x) = \frac{1}{I}\int_{-\infty}^{x} \psi(x)\, dx \quad ; \quad f(x) = \frac{1}{I}\psi(x)$$

en donde I es la integral sobre todo el campo de existencia.

Un ejemplo importante lo constituyen las integrales eulerianas de segunda especie

(10.10) $$\Gamma(n) = \int_0^{\infty} x^{n-1} e^{x}\, dx$$

que Euler introdujo para interpolar la función factorial, ya que integrando por partes se tiene:

(10.11) $$\Gamma(n) = \left[-x^{n-1} e^{-x}\right]_0^{\infty} + (n-1)\int_0^{\infty} x^{n-2} e^{-x}\, dx = (n-1)\,\Gamma(n-1)$$

y continuando con el proceso, si n es un número entero se llega a:

(10.12) $$\Gamma(n) = (n-1)(n-2)\ .\ .\ .\ \Gamma(1)$$

y como $\Gamma(1) = 1$, se tiene finalmente

(10.13) $$\Gamma(n) = (n-1)!$$

y como la integral (10.10) está definida para todo valor de n mayor que cero se tiene la fórmula de interpolación buscada.

Conforme a lo visto, dividiendo el integrando de (10.10) por el valor de la integral, a fin de que la integración sobre todo el campo de existencia, tenga como resultado la unidad, se tiene una nueva ley de probabilidad, definida unicamente para valores positivos de la variable aleatoria, y cuyas funciones de distribución y de densidad serán:

(10.14) $$F(X) = \frac{1}{\Gamma(n)} \int_0^X x^{n-1} e^{-x} dx \quad ; \quad f(x) = \frac{1}{\Gamma(n)} x^{n-1} e^{-x}$$

Utilizando las integrales eulerianas de primera especie (introducidas para interpolar los coeficientes binomiales), es posible tambien definir la ley de probabilidad de una variable aleatoria restringida al intervalo finito (0,1):

(10.15) $$F(X) = \frac{1}{B(p,q)} \int_0^X x^{p-1} (1-x)^{q-1} dx; \quad f(x) = \frac{1}{B(p,q)} x^{p-1} (1-x)^{q-1}$$

con la condición de que los parámetros p y q sean positivos.

X.9.- Estas variables aleatorias eulerianas, cuyas leyes de probabilidad se acaban de definir en una forma aparentemente arbitraria, tienen una gran importancia, por aparecer en los problemas de inferencia estadística.

Especialmente, las variables gamma, completadas con la adición de un parámetro positivo en el exponente, desempeñan un papel fundamental. La integral gamma generalizada, de orden n y parámetro α :

(10.16) $$\Gamma(n) = \int_0^\infty x^{n-1} e^{-\alpha x} dx$$

permite definir una familia de variables aleatorias, a las que pertenecen la ley exponencial y la de Ji cuadrado. Para determinar sus funciones de distribución y de densidad de probabilidad, hay que calcular primeramente la integral (10.16) a fin de obtener el factor de normalización, lo que se realiza simplemente mediante el cambio de variables $\alpha x = z$, $\alpha dx = dz$:

(10.17) $$\Gamma(n) = \frac{1}{\alpha^n} \int_0^\infty z^{n-1} e^{-z} dz = \frac{\Gamma(n)}{\alpha^n}$$

con lo que se tiene:

(10.18) $$F(X) = \frac{\alpha^n}{\Gamma(n)} \int_0^X x^{n-1} e^{-\alpha x} dx \quad ; \quad f(x) = \frac{\alpha^n}{\Gamma(n)} x^{n-1} e^{-\alpha x}$$

La ley exponencial (10.5) es evidentemente una gamma generalizada de primer orden y parámetro λ :

(10.19) $$f(\lambda T) = \frac{\lambda}{\Gamma(1)} e^{-\lambda T}$$

Si la ley de Ji cuadrado se pone en la forma:

(10.20) $$f(y) = \frac{(\frac{1}{2})^{\frac{1}{2}}}{\sqrt{\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}$$

sería una gamma generalizada de orden 1/2 y parámetro 1/2 , si se puede probar que $\Gamma(1/2)=\sqrt{\pi}$, lo que no ofrece dificultad, utilizando en (10.10) la transformación de Gauss $x = z^2$, $dx = 2zdz$

(10.21) $$\Gamma(n) = 2\int_0^{\infty} z^{2n-1}\, e^{-z^2}\, dz$$

Haciendo $n = 1/2$

(10.22) $$\Gamma(\tfrac{1}{2}) = 2\int_0^{\infty} e^{-z^2}\, dz = \int_{-\infty}^{+\infty} e^{-z^2}\, dz = \sqrt{\pi}$$

y la ley de probabilidad de Ji cuadrado toma la forma buscada:

(10.23) $$f(y) = \frac{(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})}\, y^{-\frac{1}{2}}\, e^{-\frac{y}{2}}$$

X.9.- MOMENTOS DE LAS VARIABLES ALEATORIAS CONTINUAS. Por analogía con la definición de momento de las variables aleatorias discretas, se define

(10.24) $$\mu_n = \int_{-\infty}^{+\infty} x^n f(x)\, dx$$

como el momento de orden n-ésimo de la variable aleatoria continua, si la integral existe. El intervalo de integración es el de definición de la variable aleatoria, y si ésta está limitada en uno o en ambos sentidos, se conviene en completarla con la función idénticamente nula fuera de su campo de definición, con lo que la integral (10.24) tiene validez general.

X.10.- Aplicando la definición (10.24) a la ley normal, se tiene

(10.25) $$m = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}}\, dt = 0$$

porque el integrando es evidentemente una función impar, y por lo tanto la integral (10.25) es nula en todo intervalo simétrico en torno al origen, de manera que pasando al límite en forma simétrica se obtiene el resultado indicado.

Para calcular la varianza, como debido a (10.25), la variable es centrada, basta aplicar la fórmula general e integrar por partes:

(10.25) $$\sigma^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2}}\, dt = \frac{1}{\sqrt{2\pi}} \left[-t e^{-\frac{t^2}{2}}\right]_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\, dt = 1$$

con lo que resulta que la ley normal en la expresión que se emplea para la fórmula de Laplace, es una ley reducida (ver párrafo VI.4).

En cambio, si se tiene en cuenta la forma utilizada en el teorema de De Moivre

(6.10), el momento de primer orden será

$$(10.26)\qquad \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{+\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx = \frac{1}{2\pi\sigma}\int_{-\infty}^{+\infty}(x-m)e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx + \frac{m}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}\,d\frac{x}{\sigma} = m$$

para cuyo cálculo simplemente se ha sumado y restado m a la variable x, resultando la primera integral nula y la segunda unitaria. Por lo tanto, en el teorema de De Moivre se verifica la importante propiedad que el primer momento de la variable aleatoria utilizada para aproximar la ley binomial, tiene el mismo momento que ésta última, lo que se extiende a la varianza; sin más que tomar como variable de integración $\frac{(x-m)}{\sigma}$ e integrar por partes:

$$(10.27)\qquad \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\frac{(x-m)^2}{\sigma^2}\,e^{-\frac{(x-m)^2}{2\sigma^2}}\,d\frac{(x-m)}{\sigma} = \sigma^2\left[-(x-m)\,e^{-\frac{(x-m)^2}{2\sigma^2}}\right]_{-\infty}^{+\infty} + \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}\,d\frac{(x-m)}{\sigma} = \sigma^2$$

Los momentos superiores de orden impar de la ley normal reducida, y los centrados de orden impar de la ley normal no reducida, son evidentemente todos nulos, por la misma razón que el de primer orden. Los momentos pares de orden superior, en cambio, existen siempre y se calculan mediante una ley de recurrencia inmediata:

$$(10.28)\qquad \mu_4 = \frac{\sigma^4}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\frac{(x-m)^4}{\sigma^4}\,e^{-\frac{(x-m)^2}{2\sigma^2}}\,d\frac{(x-m)}{\sigma} =$$

$$\mu_4 = \left[\frac{\sigma^4}{2\pi}\;\frac{-(x-m)^3}{\sigma^2}\,e^{-\frac{(x-m)^2}{2\sigma^2}}\right]_{-\infty}^{+\infty} + \frac{3\sigma^4}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\frac{(x-m)^2}{2}\,e^{-\frac{(x-m)^2}{2\sigma^2}}\,d\frac{(x-m)}{\sigma} = 3\sigma^4$$

y en general:

$$(10.29)\qquad \mu_{2n} = 1\cdot 3\cdot 5\cdot\ldots\cdot(2n-1)\,\sigma^{2n}$$

X.10.- FUNCION CARACTERISTICA DE LAS VARIABLES ALEATORIAS CONTINUAS. También por analogía con las funciones características de las variables aleatorias discretas, se define como función característica de una variable aleatoria continua a la siguiente expresión:

$$(10.30)\qquad \varphi(u) = \int_{-\infty}^{+\infty} e^{iux} f(x)\,dx$$

en la que se ha hecho la convención habitual de completar la definición de la función de densidad de probabilidad, en caso que su campo de variación esté limitado, con una función idénticamente nula.

Es evidente que (10.30) existe siempre, poque el factor es de módulo unitario, y las condiciones de convergencia o divergencia de una integral no se alteran si se multiplica el integrando por un factor de valor absoluto unitario y la integral de la función de densidad de probabilidad, por la condición impuesta a su primitiva, que es la

función de distribución, existe también siempre.

En cambio, si se quisiera generalizar la noción de función generatriz, se chocaría con la dificultad de que en cada caso habría que investigar la convergencia de la integral.

Admitiendo la derivación bajo el signo integral:

$$\varphi^{(n)}(u) = i^n \int_{-\infty}^{+\infty} e^{iux} x^n f(x)dx \tag{10.31}$$

y dando a t el valor cero:

$$\frac{\varphi^{(n)}(0)}{i^n} = \int_{-\infty}^{+\infty} x^n f(x)dx \tag{10.32}$$

con lo que se tiene un método para el cálculo de momentos, similar al utilizado con ventaja en las variables aleatorias discretas.

También es posible demostrar la validez de la importante propiedad de que la función característica de la suma de n variables aleatorias independientes, es igual al producto de las funciones características de las variables sumandos.

X.11.- La función característica de la variable aleatoria normal reducida:

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{iut} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{-t^2-2iut}{2}} dt \tag{10.33}$$

se calcula sin dificultad, simplemente completando el cuadrado de un binomio en el exponente:

$$\varphi(u) = e^{-\frac{u^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(t-iu)^2}{2}} dt = e^{-\frac{u^2}{2}} \tag{10.34}$$

Para calcular la función característica de la variable aleatoria normal no reducida, de primer momento m y varianza σ^2:

$$\varphi(u) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{+\infty} e^{iux} e^{-\frac{(x-m)^2}{2\sigma^2}} dx \tag{10.35}$$

podría utilizarse el mismo método, pero es más instructivo partir de la función característica de la variable reducida y utilizar los cambios de variables y transformaciones indicados en el párrafo VIII.8, y que se emplearon para la inversión de la función característica en el teorema central del límite (ver párrafo VIII.13).

La función característica de la variable centrada se halla efectuando el cambio de variables $u = \sigma w$ en la función característica de la variable reducida (el inverso del empleado para la demostración del teorema de Laplace en el cap. VIII):

$$\varphi^{**}(\sigma w) = \varphi^{*}(w) = e^{-\frac{\sigma^2 w^2}{2}} \tag{10.36}$$

Para pasar de la función característica de la variable reducida a la de la variable sin reducir, se multiplica por el factor e^{imw}:

$$\varphi(w) = e^{imw}\,\varphi^{x}(w) = e^{imw-\frac{\sigma^2 w}{2}} \tag{10.37}$$

que es la expresión de la función característica de la variable aleatoria normal no reducida (conviene recordar, como se dijo en el párrafo VIII.8 que en los textos de probabilidades no se acostumbra utilizar notaciones especiales para las funciones características, segun que la variable sea centrada o reducida, ni tampoco cambiar la denominación de la variable después de las transformaciones). Este tipo de operaciones es de uso constante en el cálculo de probabilidades y sus aplicaciones.

Mediante (10.37), y teniendo en cuenta la unicidad de la transformación entre variables aleatorias y funciones características, se demuestra inmediatamente que la suma de dos variables aleatorias normales, cuya función característica sería:

$$\varphi(u) = e^{i(m_1+m_2)u-(\sigma_1^2+\sigma_1^2)\frac{w^2}{2}} \tag{10.38}$$

es otra variable aleatoria normal, cuyos parámetros son las sumas de los parámetros correspondientes de las variables sumandos, en correspondencia con la identidad entre parámetros y momentos, y las propiedades aditivas de estos últimos, que se generalizan sin dificultad a las variables aleatorias continuas.

X.12.- La función característica de la variable aleatoria gamma generalizada será:

$$\varphi(u) = \frac{\alpha^n}{\Gamma(n)}\int_0^{\infty} e^{iux}\,x^{n-1}\,e^{-\alpha x}\,dx = \frac{\alpha^n}{\Gamma(n)}\int_0^{\infty} x^{n-1}\,e^{-(\alpha-iu)x}\,dx = \frac{\alpha^n}{(\alpha-iu)^n} \tag{10.39}$$

de donde se deduce, por ejemplo, que la función característica de la variable aleatoria Ji cuadrado es:

$$\varphi(u) = \frac{(\frac{1}{2})^{\frac{1}{2}}}{(\frac{1}{2}-iu)^{\frac{n}{2}}} = \frac{1}{(1-2iu)^{\frac{n}{2}}} \tag{10.40}$$

La suma de los cuadrados de n variables aleatorias normales, reducidas e independientes entre sí, tendrá por función característica la potencia n-ésima de (10.40) y teniendo en cuenta la correspondencia entre (10.40) y la ley (10.23), resulta que su densidad de probabilidad será:

$$f(y) = \frac{(\frac{1}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}\,y^{\frac{n}{2}-1}\,e^{-\frac{y}{2}} \tag{10.41}$$

que se conoce con la denominación de función de densidad de probabilidad de la variable aleatoria Ji cuadrado de orden n, de gran importancia en estadística.

El primer momento y la varianza de esta ley, se obtendran por derivación de (10.

.40)

(10.42) $$\varphi'(u) = \frac{n}{2}\,\frac{2i}{(1-2iu)^{\frac{n+2}{2}}} = \frac{ni}{(1-2iu)^{\frac{n+2}{2}}}$$

(10.43) $$\frac{\varphi'(0)}{i} = n$$

(10.44) $$\varphi''(0) = \frac{n}{2}\,\frac{(n+2)}{2}\cdot\frac{4i^2}{(1-2iu)^{\frac{n+4}{2}}} = \frac{ni}{(1-2iu)^{\frac{n+4}{2}}}$$

(10.45) $$\frac{\varphi''(0)}{i^2} = n^2 + 2n \quad ; \quad \sigma^2 = (n^2+2n) - n^2 = 2n$$

X.13.- LEYES EXCEPCIONALES. Como ejemplo de ley de probabilidad excepcional, que no posee ningun momento, se dará la ley de Cauchy.

Sea un centro de emisión de partículas que se supone puntual y una distancia d de una pantalla ilimitada. Sobre esta pantalla se toma como origen de coordenadas el pié de la perpendicular a la misma, bajada desde el centro de emisión . Si la probabilidad de emisión de una partícula desde el centro y con una dirección comprendida dentro de un angulo dado se acepta que es proporcional a la magnitud del ángulo, el problema consiste en hallar la ley de probabilidad correspondiente al impacto sobre la pantalla ilimitada.

Sobre la pantalla, la probabilidad de que una partícula la intercepte en el segmento (0,x) será proporcional al arco cuya tangente es x/d, lo que define una función de distribución. La función de densidad de probabilidad se hallará derivando:

(10.46) $$f(x) = \frac{d}{dx}\,\frac{1}{\pi}(\text{arc tg}\,\frac{x}{d}) = \frac{1}{\pi}\,\frac{1}{d(1+\frac{x^2}{d^2})} = \frac{1}{\pi}\,\frac{d}{d^2+x^2}$$

El factor $\frac{1}{\pi}$ resulta de la forma en que se ha definido la ley de probabilidad angular, mediante una función de distribución que resulta proporcional al arco y que debe ser normalizada por π a fin de que la diferencia entre los valores extremos de la variable sea unitaria.

Si para simplificar se toma como unidad sobre la pantalla a la distancia d , se tendrá:

(10.47) $$f(x) = \frac{1}{\pi}\,\frac{1}{1+x^2}$$

Como es fácil verificar, la integral que representaría el momento de primer orden es divergente, y por lo tanto lo serán también las que corresponden a los momentos de orden superior.

A pesar de estas singularidades, esta ley de probabilidad, cuyo estudio presenta también interés por el problema del que resulta, está vinculada con leyes continuas

no excepcionales, como la normal, conforme se verá en el capítulo XII, y con la distribución de Student, de gran importancia en estadística y que se estudiará en el capítulo XV.

C A P I T U L O XI

TEOREMAS LIMITES PARA LAS VARIABLES ALEATORIAS CONTINUAS

XI.1.- Los teoremas límites estudiados en los capítulos V (leyes de los grandes números) y VIII (teorema central del límite), que establecen el fundamento de los métodos asintóticos (válidos en el límite , y por lo tanto como aproximación para n suficientemente grande) utilizados en las variables aleatorias discretas para aplicarles el mismo tratamiento que a la ley binomial, si es que se las puede considerar como sumas de un número n de variables aleatorias, se extienden sin dificultad para las variables aleatorias continuas.

El uso sistemático de las funciones características permite dar no sólo demostraciones más sencillas, sino también condiciones de validez más amplias, aceptando la validez de un teorema de continuidad mas general que el utilizado en el Cap. VIII y que es debido a Paul Levy (16).

XI.2.- Las demostraciones del capítulo V, supuesta la existencia de los momentos de primer y de segundo orden, se aplican directamente a las variables aleatorias continuas, obteniendose las mismas leyes de los grandes números.

Pero aplicando las funciones características es posible ampliar la validez del teorema de Bernouilli generalizado.

Dada una sucesión de variables aleatorias idénticamente distribuídas, que poseen primer momento, la función característica podrá desarrollarse por la fórmula de Mac Laurin:

(11.1) $$\varphi(u) = 1 + imt + o(t)$$

Como lo que interesa es hallar la función característica de la suma promediada de las variables aleatorias, y ésta es igual a la suma de las variables, dividida cada una de ellas por n , se tendrá:

(11.2) $$\varphi\left(\frac{u}{n}\right) = 1 + \frac{imt}{n} + o\left(\frac{t}{n}\right)$$

La función característica de la suma será el producto, o sea , en este caso, la potencia n-ésima:

(11.3) $$\left[\varphi\left(\frac{u}{n}\right)\right]^n = \left[1 + \frac{imt}{n} + o\left(\frac{t}{n}\right)\right]^n$$

y pasando al límite para n infinito, se tiene

(11.4) $$\lim_{n\to\infty} \left[\varphi\left(\frac{u}{n}\right)\right]^n = e^{imt}$$

que es la función característica de una variable aleatoria que toma el valor m con probabilidad unitaria, y tiene probabilidad cero para cualquier intervalo que no incluye a m (variable aleatoria degenerada), por lo que aplicando el teorema de continuidad, queda demostrado el siguiente teorema. debido a KHINTCHINE: L a s u m a p r o m e d i a

da de n variables aleatorias identicamente distribuidas y que poseen primer momento m, converge en probabilidad a m.

XI.3.- El teorema central del límite, demostrado en los párrafos VIII.11 a VIII.13 vale también sin ningun cambio para las variables aleatorias continuas. Pero conviene ampliar su validez disminuyendo el orden de los momentos que intervienen en su demostración, como se ha hecho en el párrafo anterior con el teorema de Bernouilli generalizado.

En el caso importante de que las variables aleatorias estan identicamente distribuídas, es posible eliminar el momento de tercer orden sin ninguna dificultad.

Si existen los momentos de primer y segundo orden, y las variables centradas, su función característica tendrá la forma:

(11.5) $$\varphi(u) = 1 - \frac{\sigma^2 u^2}{2} + o(u^2)$$

Normando por la dispersión de la suma $\sqrt{n}\,\sigma$:

(11.6) $$\varphi\left(\frac{u}{\sqrt{n}\,\sigma}\right) = 1 - \frac{u^2}{2n} + o\left(\frac{u^2}{n}\right)$$

Formando la función característica de la suma, y pasando al límite para n infinito:

(11.7) $$\lim_{n\to\infty}\left[\varphi\left(\frac{u}{\sqrt{n}\,\sigma}\right)\right]^n = \lim_{n\to\infty}\left[1 - \frac{u^2}{2n} + o\left(\frac{u^2}{n}\right)\right]^n = e^{-\frac{u^2}{2}}$$

o sea el teorema de LINDEBERG y LEVY: La suma de n variables aleatorias idénticamente distribuídas y con segundo momento, centrada y normada por la dispersión de la suma tiene como límite la ley normal reducida.

Una notación conveniente es la de indicar los parámetros de la ley normal en un paréntesis, colocando primero el valor esperado y después la dispersión. Con esta notación, la ley normal reducida, se denominaría ley normal (0,1).

XI.4.- Como aplicación del teorema de LINDERBERG y LEVY, se lo utilizará para la determinación aproximada de la ley de Ji cuadrado de orden n (ver párrafo X.12).

Dado que esta ley está definida como la suma de n variables aleatorias identicamente distribuídas, y los momentos de primer y segundo orden de la suma son n y 2n , respectivamente, para n suficientemente grande, la variable reducida

(11.8) $$t = \frac{y - n}{\sqrt{2n}}$$

tendrá muy aproximadamente una distribución normal.

El interés de este resultado reside en que la convergencia es bastante rápida, y todavía es posible mejorarla, si en vez de la ley de Ji cuadrado se toma la de $\sqrt{2y}$ que converge aún más rapidamente a la normal (2n-1, 1), segun observó R. A. FISHER.

En el cuadro siguiente se comparan los valores exactos hasta la tercera cifra mediante la aproximación normal y la de Fisher, para lo que se usó las siguientes fórmulas de transformación, en las que $n=30$:

(11.9) $$t_{2p} = \frac{y_p - 30}{\sqrt{30}} \quad \therefore \quad y_p = \sqrt{30}\,t_{2p} + 30$$

(11.10) $$t_{2p} = \sqrt{2y_p} - \sqrt{29} \qquad y_p = \tfrac{1}{2}(t_{2p} + \sqrt{29})^2$$

El valor de y que se supera con una probabilidad p (indicada multiplicando por 100, de acuerdo con la práctica usual de los libros de estadística, a fin de ahorrar cero y comas) se ha denominado y_p, y con t_{2p} al desvío normal reducido que es superado en valor absoluto con probabilidad 2p, utilizandose para esta última la tabla II de CRAMER (3), habiéndose empleado una tabla más completa para $p=1$

p	t_{2p}	Aproximación normal	Aproximación de R.A.FISHER	Valor exacto segun tabla
10	1,2816	39,927	40,165	40,265
5	1,6449	42,741	43,488	43,773
1	2,3263	48,020	50,750	50,892

La aproximación lograda mediante el método de R.A.FISHER es tan buena, que usualmente las tablas de Ji cuadrado llegan solamente a $n=30$.

C A P I T U L O XII.

V A R I A B L E S A L E A T O R I A S M U L T I D I M E N S I O N A L E S

XII.1.- Al estudiar las sumas de variables aleatorias discretas, en el capítulo V, se introdujo la noción de distribución conjunta de variables aleatorias (párrafo V.3), que después no se continuó utilizando porque el empleo de las funciones generatrices, o características, permitió evitarla.

Pero existe una variedad de casos en los cuales la necesidad de estudiar una distribución conjunta se presenta naturalmente, como por ejemplo cuando se observan simultáneamente dos o más magnitudes. En particular, el conjunto de las variables de estado de un sistema físico, si se admite que cada una de ellas sigue una ley de probabilidad, identifica una distribución conjunta.

Tambien aparece naturalmente una distribución conjunta cuando se estudian grupos de observaciones de una misma variable aleatoria, problema de gran importancia en estadística, en la cual estos grupos reciben el nombre de m u e s t r a s d e n e l e m e n t o s, o si se trata de determinar la existencia de una posible relación de dependencia entre dos variables aleatorias que se observan separadamente, como es el caso del estudio de la herencia de los caracteres físicos, en el que por ejemplo, se agrupan las estaturas de los padres y de los hijos.

XII.2.- VARIABLES ALEATORIAS BIDIMENSIONALES. Si un suceso está identificado por un par de observaciones numéricas, se dice que está definida una variable aleatoria bidimensional y la función de distribución conjunta se define de la siguiente manera:

(12.1) $$F(x,y) = P(X \leqslant x, Y \leqslant y)$$

Si esta función de distribución es suficientemente regular, será posible definir una función f(x,y), llamada función de densidad de probabilidad, tal que la probabilidad de que una observación se encuentre en el rectángulo de lados paralelos a los ejes coordenados, identificado por las abcisas (a,b) y las ordenadas (c,d),(ver figura 2) es igual a la integral doble

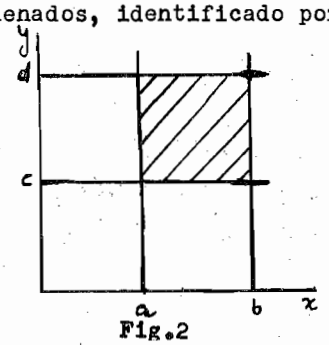


Fig.2

(XII.2) $$\int_c^d \int_a^b f(x,y)\,dx\,dy$$

La probabilidad de que la abcisa sea menor que un cierto valor x , sin establecer limitaciones a la ordenada, o sea de que el punto representativo se encuentre en un semiplano limitado por la recta vertical de abcisa x será:

(12.2) $$F_1(x) = \int_{-\infty}^{x} du \int_{-\infty}^{+\infty} f(u,y)\, dy$$

llamda d i s t r i b u c i ó n m a r g i n a l d e a v a r i a b l e X.

De igual manera, la distribución marginal de la variable y será:

(12.3) $$F_2(y) = \int_{-\infty}^{y} dv \int_{-\infty}^{+\infty} f(x,v)dx$$

y las funciones de densidad de probabilidad correspondientes, siempre en la hipótesis de regularidad de la función de distribución, serán:

(12.4) $$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)dy \quad ; \quad f_2(y) = \int_{-\infty}^{+\infty} f(x,y)dx$$

XII.3.) INDEPENDENCIA. Dadas dos variables aleatorias continuas, cuyas leyes de probabilidad constituyen las leyes marginales de su distribución conjunta, se dice que s son independientes si para dos sucesos identificados por las relaciones $a \leq x \leq b$; $c \leq y \leq d$ las probabilidades correpondientes cumplen la relación establecida en (III.5), cualesquiera sean los pares de valores (a,b) y (c,d):

(12.5) $$P(a \leq x \leq b,\ c \leq y \leq d) = P(a \leq x \leq b)\ P(c \leq y \leq d)$$

Pasando al límite para $a = -\infty$ y $c = -\infty$, se tiene

(12,6) $$F(x,y) = F_1(x)\ F_2(y)$$

y las funciones de densidad de probabilidad cumpliran la siguiente relación:

(12.7) $$f(x,y) = f_1(x)\ f_2(y)$$

Las tres relaciones (12.5) a (12 .7) son equivalentes, en el sentido de que de c ualqiera de ellas se deducen las otras dos.

XII.4.- MOMENTOS. Los momentos de una variable aleatoria bidimensional se definen

(12.8) $$\mu_{i,j} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^i y^j f(x,y)dxdy$$

o usandó la notación de esperanza matemática, que significaba multiplicar por las funciones de densidad de probabilidad e integrar sobre todo el recinto de existencia:

(12.9) $$\mu_{i,j} = E(X^i Y^j)$$

La suma de los exponentes da el orden del momento, y segun esta definición, existen dos momentos de primer orden, y tres momentos de segundo orden:

(12.10) $$m_1 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy \quad ; \quad m_2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)dxdy$$

$$\mu_{2,0} = \iint x^2 f(x,y)dxdy \ ; \ \mu_{1,1} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy \ ; \ \mu_{0,2} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2 f(x,y)dxdy$$

Como en las variables aleatorias estudiadas anteriormente, tienen especial importancia los momentos centrados de segundo orden:

$$(12.11)\qquad \sigma_1^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-m_1)^2 f(x,y)\,dxdy \quad ; \quad \mu'_{1,1} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-m_1)(x-m_2)f(x,y)\,dxdy$$

$$\sigma_2^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(y-m_2)^2 f(x,y)\,dxdy$$

XII.4.- VARIABLES ALEATORIAS DEGENERADAS. Es posible que la variable aleatoria sea sólo aparentemente bidimensional, y que toda la probabilidad esté concentrada sobre una curva, o en un punto.

El estudio del caso en que la curva se reduce a una recta es sencillo, y tienen gran importancia. Si se introduce la forma linear auxiliar $u(X-m_1)+(Y-m_2)v$ en la que u y v son parámetros arbitrarios, se tiene que:

$$(12.12)\qquad E\left[u(X-m_1) + v(Y-m_2)\right]^2 = \sigma_1^2u^2 + 2\mu'_{11}vu + \sigma_2^2v^2$$

Evidentemente, la forma cuadrática no puede ser negativa, porque su construcción consiste en la integración de productos de probabilidades por cuadrados.

Si existe un par de valores de u_o y v_o, tal que la forma cuadrática es nula, sin que lo sean simultáneamente v_o y u_o toda la distribución está concentada sobre la recta de ecuación:

$$(12/13)\qquad u_o(X-m_1) + v_o(Y-m_2) = 0$$

En efecto, si existe una probabilidad distinta de cero que corresponda a un punto o a un conjunto de puntos cuyas coordenadas no satisfacen la ecuación de la recta (XII.13), la forma cuadrática no puede ser nula. La potencia de los puntos con respecto a la recta no es nula, y como está elevada al cuadrado, multiplicada por la densidad de probabilidad e integrada, sin que existen términos negativos que puedan compensar su contribución, la forma cuadrática tendría un valor positivo en contradicción con la hipótesis.

Si la variable aleatoria bidimendional no está degenerada sobre una recta, la forma cuadrática (XII.12) será, por lo tanto, d e f i n i d a p o s i t i v a, pudiendo anularse unicamente si $u=v=0$ y la investigación del problema de la degeneración lineal se reduce a la discusión clásica de las formas cuadráticas.

Si σ_1^2 es distinto de cero, es posible completar un cuadrado en (XII.12):

$$(12.14)\qquad \sigma_1^2u^2 + 2\mu'_{11}vu + \sigma_2^2v^2 = \left(\sigma_1 u + \frac{\mu'_{11}}{\sigma_1}v\right)^2 + v^2\left(\sigma_2^2 - \frac{\mu'^2_{11}}{\sigma_1^2}\right)$$

Si u y v no son simultáneamente nulos, el segundo término de (12.14) puede anularse unicamente si el último paréntesis es nulo. Como σ_1^2 es siempre positivo, se tiene que la condición necesaria y suficiente para que la forma cuadrática sea definida

positiva es que el determinante de la matriz de sus coeficientes (llamada matriz de covarianza):

(12.15) $$\sigma_1^2 \sigma_2^2 - \mu_{11}'^2 \neq 0$$

sea distinto de cero.

Llamando $\rho^2 = \frac{\mu_{11}'^2}{\sigma_1^2 \sigma_2^2}$ (12.15) se transforma en:

(12.16) $$\sigma_1^2 \sigma_2^2 - \mu_{11}'^2 = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$$

con lo que resulta que la variable aleatoria bidimensional degenera en una variable aleatoria lineal si el cuadrado de ρ (llamado coeficiente de correlación) es igual a la unidad. Si σ_1^2 es nulo, la variable aleatoria se concentra en una recta vertical de abcisa m_1.

XII.5.- DEPENDENCIA LINEAL. Si el coeficiente de correlación es igual a la unidad es evidente que al concentrarse la ley de probabilidad sobre una recta, existe una dependencia lineal estricta entre las variables aleatorias definidas por las leyes marginales.

Como a su vez, por definición $\mu_{11}' = 0$ si las variables marginales son independientes, se tiene que el coeficiente de correlación, cuyo campo de variación está limitado entre +1 y -1 define en cierta manera el grado de dependencia. Si toma los valores extremos, habrá una relación funcional estricta, si toma el valor cero, habrá independencia, y para valores intermedios una dependencia menos rígida, que aumentará a medida que se acerca a 1 ó a -1.

XII.6.- REGRESION LINEAL. Si se considera que X es la variable independiente, tiene interés el precisar la discusión anterior hallando la ecuación de una recta que cumpla con la condición de que la esperanza matemática de los cuadrados de la diferencias de las ordenadas de los puntos de la variable bidimensional, con respecto a las ordenadas de dicha recta, sea mínima. En otros términos, que sea mínima la suma de los cuadrados de las diferencias, ponderadas por sus probabilidades.

La expresión que hay que hacer mínima es:

(XII.7) $$E(Y - \alpha - \beta X)^2$$

Procediendo a centrar las variables marginales, para lo cual hay que restar los primeros momentos, se tiene:

(12.18) $$E(Y - \alpha - \beta x)^2 = E[(Y - m_2) - \beta(X - m_1) + (\alpha + m_2 - \beta m_1)]^2 =$$

$$= \sigma_2^2 - 2\mu_{11}'\beta + \sigma_1^2\beta^2 + (m_2 - \alpha - \beta m_1)^2$$

para lo que primeramente se desarrolló el cuadrado como si constara únicamente de dos

términos: las variables aleatorias agrupadas y las constantes. El doble producto es nulo (porque las variables estan centradas), y el cuadrado de las variables aleatorias agrupadas, se volvió a desarrollar.

El mínimo de la expresión corresponde a los valores de α y de β que anulen las derivadas primeras con respecto a los mismos:

(12.19) $$\alpha = m_2 - \beta m_1 \quad ; \quad \beta = \frac{\mu_{11}}{\sigma_1^2} = \rho \frac{\sigma_2}{\sigma_1}$$

con lo que se tiene finalmente:

(12.20) $$Y = m_2 + \rho \frac{\sigma_2}{\sigma_1}(X - m_1)$$

El coeficiente que da la pendiente de la recta, se llama c o e f i c i e n t e d e r e g r e s i ó n d e y e n x, y la recta, r e c t a d e r e g r e s i ó n d e y e n x.

Es evidente que en los casos de una dependencia no estricta entre x e y (valores de ρ cercanos a +1 ó -1), la recta de regresión puede representar esa dependencia, en un sentido que se precisará en los capítulos XIV y XV, donde también se indicaran algunos problemas de aplicación nada sencillos, que hacen que este método de regresión, de gran importancia teórica, tenga un significado práctico muy dudoso en muchos casos.

Es posible también determinar la regresión de x en y, resultando la recta:

(12.21) $$X = m_1 + \rho \frac{\sigma_1}{\sigma_2}(Y - m_2)$$

XII.7.- APLICACIONES DE LA DISTRIBUCION CONJUNTA. En el capítulo V(párrafo V) se aplicó el artificio de introducir una distribución conjunta sobre el espacio producto, a fin de estudiar la suma de dos variables aleatorias. El mismo procedimiento tiene importancia para determinar la ley del cociente de dos variables aleatorias independientes, que tiene gran importancia por aparecer repetidamente en estadística.

Sea, por ejemplo, hallar la ley de probabilidad el cociente de dos variables aleatorias independientes entre si:

(12.22) $$z = \frac{y}{x}$$

En el plano (x,y), dada la independencia supuesta por hipótesis, la densidad de probabilidad de un punto será:

(12.23) $$f(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

y para hallar la densidad de probabilidad de z habrá que integrar sobre todos los puntos de la recta cuya pendiente es precisamente z(ver figura 3).

Para dicha integración hay que hacer un cambio de variables, pasando a coordenadas polares, o tambien el más sencillo:

(12.24) $$xz = y \quad ; \quad dy = x\,dx\,dz$$

con lo que el elemento diferencial de la distribución conjunta será:

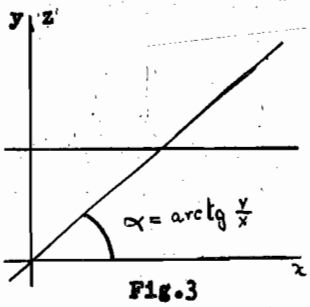


Fig.3

$$(12.25) \qquad f(x,z)dxdz = \frac{1}{2\pi} e^{-\frac{(1+z^2)x^2}{2}} xdxdz$$

La recta de pendiente z se ha transformado en otra recta de ordenada z, y la ley de probabilidad del cociente es entonces la ley marginal de z, o sea que hay que integrar (12.25) sobre todo el campo de existencia de x:

$$(12.26) \qquad f(z)dz = \frac{dz}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{(1+z^2)x^2}{2}} xdx$$

Si se toma como variable de integración $x^2 = u$, $2xdx = du$, la integral se reduce a una gamma generalizada de orden 1 y parámetro $\frac{(1+z^2)}{2}$ (z permanece constante en la integración) cuyo valor se conoce (ver párrafo x.9):

$$(12.27) \qquad \frac{1}{2}\int_{-\infty}^{+\infty} e^{-\frac{(1+z^2)u}{2}} du = \int_0^{\infty} e^{-\frac{(1+z^2)u}{2}} du = \frac{2}{(1+z^2)}$$

y por lo tanto la ley de probabilidad buscada es:

$$(12.28) \qquad f(z) = \frac{1}{\pi} \frac{1}{(1+z^2)}$$

o sea la ley de Cauchy (ver párrafo X.13)

Este método, combinado con la reducción de la integral a una gamma generalizada, permite resolver los problemas más importantes que se presentan en estadística, como se verá en el Capítulo XV.

XII.8.- INTERPRETACION GEOMETRICA DE LA LEY DE JI CUADRADO. Utilizando la distribución conjunta de n variables aleatorias normales, reducidas e independientes entre sí, es posible dar una interpretación geométrica a la distribución de Ji cuadrado, que no sólo permite simplificar algunos problemas de aplicación importantes, sino que también tuvo históricamente un papel destacado en el desarrollo de la estadística matemática.

En el espacio n-dimensional, la densidad de probabilidad correspondiente a la distribución conjunta de n variables aleatorias normales, reducidas e independientes entre sí, será:

$$(12.29) \qquad f(r) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{x_1^2+x_2^2+\cdots+x_n^2}{2}} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{r^2}{2}}$$

de donde r es la distancia al origen.

La probabilidad de superar el cuadrado de una distancia dada al origen, será la que corresponde a los puntos que estan fuera de la esfera de radio igual a esa distancia. La densidad de probabilidad correspondiente podrá hallarse integrando (12.29) sobre la superfice de la esfera de radio r, lo que se reduce en realidad a multiplicar (12.29),

que es constante, por dicha superficie.

Segun la fórmula de LUDWIG SCHAFLI, la superficie de la esfera de radio r en el espacio de n dimendiones es:

$$S = \frac{\pi^{\frac{n}{2}} \cdot 2r^{n-1}}{\Gamma\left(\frac{n}{2}\right)}$$

Multiplicando por (12.29)

(12.30) $$f(r) = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}} \cdot 2r^{n-1}}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{r^2}{2}}$$

y tomando como variable $y = r^2$ (para lo cual hay que recordar la transformación del elemento diferencial $dy = 2rdr$), se tiene:

(12.31) $$f(y) = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}$$

que es precisamente la ley de Ji cuadrado de orden n, hallada en el capítulo X (párrafo X.12) en forma puramente análítica .

XII.9.- Tanto la superficie como el volumen de la esfera en n dimensiones se usan constantemente en física estadística, y tiene cierto interés mostrar como se deducen, también mediante el empleo de integrales gamma.

Recordando la expresión de la integral de Gauss:

(12.32) $$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$$

integral similar en el espacio de n dimensiones, será:

(12.32) $$I_n = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{-\frac{x_1^2+x_2^2+\cdots+x_n^2}{2}} dx_1 dx_2 \ldots dx_n = (2\pi)^{\frac{n}{2}}$$

Pero la integral de Gauss en n dimensiones se puede transformar pasando a coordenadas polares:

(12.33) $$I_n = \int_0^{\infty} e^{-\frac{r^2}{2}} r^{n-1} dr \int d\Omega$$

en donde $\int d\Omega$ es la integral múltiple extendida a la función angular que resulta de la transformación de las variables y del Jacobiano.

La integral con respecto al radio vector se reduce a una gamma, mediante la transformación evidente:

(12.34) $$\tfrac{1}{2}r^2 = x \quad ; \quad rdr = dx$$

con lo que se tiene:

(12.35) $$I_n = 2^{\frac{n-2}{2}} \Gamma\left(\frac{n}{2}\right) \int d\Omega$$

e igualando (12.35) y (12.32):

(12.36) $$(2\pi)^{\frac{n}{2}} = 2^{\frac{n-2}{2}} \Gamma\left(\frac{n}{2}\right) \int d\Omega \quad \therefore \quad \int d\Omega = \frac{2\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}$$

pero esta última integral es la superficie de la esfera de radio unitario. Haciendo $n=2$, se tiene:

(12.37) $$\frac{2\pi}{\Gamma(1)} = 2\pi$$

que es la longitud del círculo de radio unitario; y haciendo $n=3$:

(12.38) $$\frac{2\pi^{\frac{3}{2}}}{\Gamma\left(\frac{3}{2}\right)} = \frac{2\pi^{\frac{3}{2}}}{\frac{1}{2}\Gamma\left(\frac{1}{2}\right)} = 4\pi$$

que es la fórmula de Arquímides para la superficie de la esfera de radio unitario.

Si el radio no es unitario, las superficies estarán en relación igual a la potencia n-1 del radio, con lo que la fórmula general de la superficie de la esfera de radio n en el espacio de n dimensiones será:

(12.39) $$\frac{2\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} r^{n-1}$$

e integrando con respecto al radio se tendrá el volumen:

(12.40) $$\frac{2\pi^{\frac{n}{2}}}{n\,\Gamma\left(\frac{n}{2}\right)} r^{n} = \frac{\pi^{\frac{n}{2}} r^{r}}{\Gamma\left(\frac{n}{2}+1\right)}$$

Después de lo visto en este capítulo se justifica la afirmación de que las integrales gamma son el instrumento analítico, por excelencia, de la estadística matemática.

C U A R T A P A R T E

I N T R O D U C C I O N A L A I N F E R E N C I A E S T A D I S T I C A .

CAPITULO XIII. ESTIMACION PUNTUAL

XIII.1.- Una vez desarrollados, en forma deductiva a partir de una axiomática, los elementos del cálculo de probabilidades, corresponde tratar la aplicación de los resultados obtenidos a la interpretación de las observaciones y mediciones, tema que se identifica con la designación de i n f e r e n c i a e s t a d í s t i c a .

En efecto, como se verá a continuación, la interpretación de las observaciones y mediciones se realiza mediante la introducción de una ley de probabilidad. Pero mientras que en el cálculo de probabilidades propiamente dicho se postulaba la forma de la ley y se estudiaban sus propiedades, en la inferencia estadística hay que conjeturar racionalmente cual es la ley que sigue el fenómeno aleatorio, e inferir a partir de las observaciones los posibles valores de los parámetros que la identifican completamente.

XIII.2.- Segun palabras de R.A.FISHER: "...el objeto de los métodos estadísticos es la "reducción de los datos. Una cantidad de datos, que debido a su magnitud es usualmente "imposible de ser retenida por la mente, debe ser reemplazada por relativamente pocas "cantidades que deberan representar adecuadamente el conjunto,o sea, en otras palabras, "deberan contener tanta información como sea posible, idealmente el total, de la contenida en los datos originales."(a)

El proceso de reducción racional de los datos se realiza suponiendo que el conjunto de las observaciones se puede identificar como una muestra de una cierta ley de probabilidad. Conocida la forma de la ley, o sea su expresión analítica (normal, de Poisson, de Cauchy, etc) y determinado el valor numérico de los parámetros que dentro de cada forma individualizan una ley en particular, el proceso de reducción queda idealmente terminado, y es posible realizar predicciones sobre el resultado de nuevas observaciones.

En la práctica, el esquema anterior adopta una de las dos siguientes formas (b):

(1) Una serie de n observaciones que representan mediciones o determinaciones de una misma característica, r e a l i z a d a s c a d a u n a s o b r e u n o b j e t o d i s t i n t o. Este es el problema típico de la e s t a d í s t i c a, y el significado de la ley de probabilidad que eventualmente se determina es describir la

---

(a) R.A.FISHER: "On the mathematical foundations of theoretical statistics". Philosphical Transactions of the Royal Society (London): Serie A, vol. 222 (1922), pags. 309-368.
(b) Manual de ASTM referido en (30) pag.2.

variabilidad del caracter que se observa, en una población ideal de magnitud infinita. Su estudio sistemático es una de las contribuciones del siglo XX.-

Es evidente la importancia, tanto práctica como teórica, de este caso de aplicación del cálculo de probabilidades. Desde el punto de vista práctico, significa no solo la posibilidad de obtener en forma económica (con un mínimo de observaciones) un conocimiento sobre las características de una población que de otra manera debería estudiarse en su totalidad, sino también, y como se verá en los capítulo siguientes, un complemento esencial del método experimental. Desde el punto de vista teórico representa el primer ejemplo, fuera de la aritmética, de un método coherente de i n d u c c i ó n lo que tiene gran significado para la filosofía de las ciencias.

(2) Una serie de n observaciones, que representan mediciones de una misma característica, r e i t e r a d a s s o b r e u n m i s m o o b j e t o. Este es el caso clásico de la teoría de los errores de observacion, y una relación preestablecida entre la ley de probabilidad y el verdadero valor de la magnitud observada permite hacer una determinación de esta última a partir de un grupo de observaciones afectadas por errores accidentales. Este caso de aplicación fué largamente estudiado por LAPLACE y por GAUSS.

Las dos interpretaciones distintas que se acaban de hacer del proceso de reducción de los datos, n o p r e s e n t a n d i f e r e n c i a s d e s d e e l p u n t o d e v i s t a d e l c á l c u l o d e p r o b a b i l i d a d e s que aplica a ambas los mismos métodos básicos, siendo la inferencia estadística un nuevo ejemplo de la utilización de un mismo modelo matemático para el estudio de fenómenos distintos, como se vió anteriormente (por ejemplo, en el caso del paseo al azar y de los juegos, señalado en el párrafo V.13). Debido a esta similitud de tratamiento, el desarrollo de la inferencia estadística en el siglo XX, ha tenido por consecuencia una revisión profunda y una considerable mejora de eficiencia en los métodos de la teoría de los errores construída en el siglo XIX, que se encuentra reseñada en algunos textos de física, y que hoy no aparece como aceptable.

XIII.3.- METODO DE LAPLACE. A fin de precisar los diversos problemas que se presentan en la inferencia estadística, resulta conveniente discutir un ejemplo de considerable interés histórico, siguiendo las ideas desarrolladas por LAPLACE en su gran tratado (a) y aplicando el teorema de BAYES (ver cap. III, párrafos 19 y 20).

Sea un conjunto de n observaciones de un fenómeno aleatorio de alternativa simple y probabilidad constante, en las que r observaciones corresponden a una alternativa y n-r a la complementaria. Para fijar ideas, se supondrá que el fenómeno

(a) Libro II, Cap. VI.

estudiado es la extracción de bolillas blancas y negras de un bolillero o urna, con reemplazo.

Si existen m bolilleros de composición $p_1, p_2, \ldots, p_m$ conocida, pero sin señales exteriores que identifiquen de que bolillero se ha extraído las n bolillas, aplicando (3.51) se tendrá que la probabilidad condicional de que el bolillero tenga la composición p_i dado el resultado de r bolillas de un color en n extraccciones es

$$P(p_i/r) = \frac{p^r(1-p)^{n-r}}{\sum p^r(1-p)^{n-r}} \tag{13.1}$$

en donde se haa supuesto que todas las probabilidades "a priori" son iguales (por aplicación del principio de indiferencia), y se han simplificado los coeficientes combinatorios.

La idea de LAPLACE consiste en generalizar este resultado para resolver el problema de determinar la composición de un bolillero a partir de n observaciones, suponiendo que la circunstancia de encontrarse frente a un bolillero determinado es consecuencia de una elección entre infinitos bolilleros (o más correctamente, fenómenos al aleatorios de alternativa simple), que poseen todos los valores de p, desde cero hasta 1 inclusive, e igual probabilidad "a priori".

Con este supuesto, la suma del denominador de (13.1) se convierte en una integral entre 0 y 1:

$$f(p_i/r) = \frac{p_i^r(1-p_i)^{n-r}}{\int_0^1 p^r(1-p)^{n-r}dp} \tag{13.2}$$

La expresión (13.2), fijados n y r, es una función de p, que es en realidad una densidad de probabilidad, y el problema consiste en encontrar la forma de utilizar esa función de densidad de probabilidad para deducir el valor más conveniente que hay que atribuir a p, y que sentido preciso tiene la expresión "mas conveniente". El problema así caracterizado es el de la estimación puntual.

XIII.4.- A fin de poder operar convenientemente con (13.2), es útil observar que el denominador es una función euleriana de primera especie $B(r+1, n-r+1)$, que se reduce fácilmente, si como ocurre en el problema considerado $(r+1)$ y $(n-r+1)$ son enteros, a una combinación de factoriales (de este párrafo lo que interesa es el resultado final (13.9)).

Aplicando reiteradamente la transformación de Gauss, ya empleada en el cap.X, párrafo 8 para calcular $\Gamma(\frac{1}{2})$ se tiene:

$$\Gamma(a) = 2\int_0^\infty x^{2a-1}\, e^{-x^2}dx \quad ; \quad \Gamma(b) = 2\int_0^\infty y^{2b-1}\, e^{-y^2}dy \tag{13.3}$$

Efectuando el producto de ambas gammas:

$$\Gamma(a)\,\Gamma(b) = 4\int_0^{\infty}\int_0^{\infty} x^{2a-1}\, y^{2b-1}\, e^{-(x^2+y^2)}\,dx\,dy \tag{13.4}$$

Pasando a coordenadas polares $x=\rho\cos\varphi$; $y=\rho\,\mathrm{sen}\,\varphi$; $J=\rho$:

$$\Gamma(a)\,\Gamma(b) = 4\int_0^{\infty}\int_0^{\frac{\pi}{2}} \rho^{2(a+b)-1}\, e^{-\rho^2}\, \cos^{2a-1}\varphi\,.\,\mathrm{sen}^{2b-1}\varphi\; d\varphi\, d\rho \tag{13.5}$$

Como los términos en ρ pueden separarse de los términos en φ, la integral doble puede a su vez separarse en el producto de dos integrales simples:

$$\Gamma(a)\,\Gamma(b) = 4\int_0^{\infty} \rho^{2(a+b)-1}\, d\rho\,.\int_0^{\frac{\pi}{2}} \cos^{2a-1}\varphi\,\mathrm{sen}^{2b-1}\varphi\, d\varphi \tag{13.6}$$

La integral en ρ es precisamente $\Gamma(a+b)$ con lo que se tiene:

$$\Gamma(a)\,\Gamma(b) = \Gamma(a+b).\,2\int_0^{\frac{\pi}{2}} \cos^{2a-1}\varphi\;\mathrm{sen}^{2b-1}\varphi\, d\varphi \tag{13.7}$$

La integral trigonométrica se calcula mediante la substitución $\cos^2\varphi = z$, $dz = -2\cos\varphi\,\mathrm{sen}\,\varphi\, d\varphi$:

$$\Gamma(a)\,\Gamma(b) = \Gamma(a+b)\int_0^1 z^{a-1}\,(1-z)^{b-1}\,dz \tag{13.8}$$

y la integral del segundo miembro es precisamente $B(a,b)$, con lo que se ha obtenido la fórmula buscada:

$$B(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)} = \frac{(a-1)!\,(b-1)!}{(a+b-1)!} \tag{13.9}$$

XIII.5.- Con la transformación demostrada en el párrafo anterior, (XIII.2) se convierte en :

$$f(p/r) = \frac{p^r(1-p)^{n-r}}{B(r+1,\,n-r+1)} = \frac{(n+1)!}{r!(n-r)!}\; p^r(1-p)^{n-r} \tag{13.10}$$

Dada en forma explícita la ley de probabilidad "a posteriori" de p_1, un método posible para atribuir un valor conveniente a p_1 sería elegir aquel para el cual (13.10) alcanza un máximo. Pero un instante de reflexión demuestra que la elección no es muy justificada si dicho máximo no es tan pronunciado como para que la integral de (13.10) en un intervalo muy pequeño en torno a él sea muy cercana a la unidad (lo que garantizaría que valores alejados del elegido serían muy poco probables, criterio que se precisará en el capítulo XV).

En vez de analizar detalladamente la forma de la curva representativa de (13.10), la ocurrencia ingeniosa de LAPLACE equivale a hacer n variable e investigar si para n creciendo indefinidamente la densidad de probabilidad se concentra también indefinidamente en torno a un cierto valor. Es evidente que sólo de una manera similar

a ésta se puede fundar un método de estimación puntual del parámetro p, pues sino se correría el riesgo de que la precisión obtenida no fuera la necesaria en una aplicación particular, ya que en este caso la precisión se mediría por la probabilidad de que el valor de p se hallara en un intervalo de longitud dada, lo que obliga a buscar un método asintótico.

El valor en torno al cual se concentra la probabilidad al crecer n, de tal manera que para un intervalo de longitud fija que lo comprenda, la integral de (13.10) tiende al límite 1 cuando n tiende a $+\infty$, es precisamente el primer momento o esperanza matemática de (13.10). La demostración consiste en calcular dicho primer momento y estudiar la distribución de (13.10) en torno a él por medio de la desigualdad de TCHEBICHEFF (ver párrafo IV.12).

$$(13.11)\quad m = \int_0^1 pf(p/r)dp = \frac{(n+1)!}{r!(n-r)!}\int_0^1 p^{r+1}(1-p)^{n-r}dp = \frac{(n+1)!}{r!(n-r)!}B(r+2,\ n-r+1) =$$

$$= \frac{(n+1)!}{r!\,(n-r)!}\cdot\frac{(r+1)!\,(n-r)!}{(n+2)!} = \frac{r+1}{n+2}$$

La varianza se obtiene calculando el segundo momento al origen y restando el primero momento elevado al cuadrado (ver párrafo (IV.11)):

$$(13.12)\quad \mu_2 = \int_0^1 p^2 f(p/r)dp = \frac{(n+1)!}{r!(n-r)!}\int_0^1 p^{r+2}(1-p)^{n-r}dp = \frac{(n+1)!}{r!\,(n-r)!}B(r+3,\ n-r+1)$$

$$= \frac{(n+1)!}{r!\,(n-r)!}\cdot\frac{(r+2)!\,(n-r)!}{(n+3)!} = \frac{(r+1)(r+2)}{(n+2)(n+3)}$$

$$(13.13)\quad \sigma^2 = \mu_2 - m^2 = \frac{(r+1)(r+2)}{(n+2)(n+3)} - \frac{(r+1)^2}{(n+2)^2} = \frac{(r+1)(n-r+1)}{(n+2)^2\,(n+3)}$$

La expresión final de la varianza es una función de n que evidentemente tiende a cero cuando n tiende a $+\infty$, lo que tambien puede verse introduciendo la expresión (13.11) de m:

$$(13.14)\quad \sigma^2 = \frac{m}{(n+3)}\,\frac{(n-r+1)}{n+2}$$

Sumando y restando 2 dentro del paréntesis:

$$(13.15)\quad \sigma^2 = \frac{m(1-m)}{n+3} \leqslant \frac{1}{4(n+3)}$$

porque m es un número positivo inferior a la unidad, y el producto por su complemento a la unidad está acotado superiormente por 1/4.

La expresión (13.11), como estimación de la composición desconocida del bolillero, o en terminología mas precisa, como estimación del parámetro desconocido de una ley bi-

nomial, es la famosa l e y d e s u c e s i o n e s de LAPLACE.

XIII. 6.- DISCUSION DEL METODO DE ESTIMACION DE LAPLACE. Este método ha sido muy discutido, especialmente en las aplicaciones prácticas que de él hicieron LAPLACE y POISSON (hay que tener presente que la terminología empleada de inferencia, estimación, etc., es moderna y no se encuentra en los autores del siglo XIX), como puede verse, por ejemplo, en USPENSKY (22) Cap. IV, Sec.7.

Dado el caracter dudoso del principio de indiferencia empleado, es clàro que su validez está limitada al caso en que efectivamente estan definidas las probabilidades "a priori" y son conocidas, lo que restringe considerablemente su aplicación. Pero en su construcción estan en realidad dados muchos elementos de utilidad para una formulación distinta, libre de las objeciones señaladas.

En efecto, lo que queda principalmente en claro es que la elección de la función de los valores de la muestra que se utiliza para estimar el parámetro desconocido, depende del estudio de una ley de probabilidad que atribuye a dicha función una propiedad que hace razonable adoptarla en lugar del parámetro desconocido. Esto representa un progreso considerable, que fué sin embargo olvidado por los autores posteriores, que muchas veces confundieron la función de la muestra con el verdadero valor del parámetro, con lo que los problemas de inferencia estadística resultan completamente oscurecidos.

XIII.7.- METODO DE LA MAXIMA VEROSIMILITUD. En 1922, R.A.FISHER pudo establecer por fin, en una memoria célebre, las bases de una teoría coherente de la inferencia estadística (a), que permite una interpretación en términos de frecuencias relativas, aprovechando la circunstancia de que una vez resuelto el problema de especificación, toda función de valores de la muestra, o e s t i m a d o r, que se utilice para calcular un valor "conveniente" para emplear en lugar del parámetro desconocido, resulta ser una variable aleatoria cuya ley se deduce de la ley del fenómeno aleatorio en estudio.

En dicha memoria se distinguen claramente las siguientes cuestiones, que se citan textualmente:

(1) P r o b l e m a s d e E s p e c i f i c a c i ó n. Surgen en la elección de la forma matemática de la población; (2) P r o b l e m a s d e E s t i m a c i ó n. Implican la elección de los métodos de cálculo a partir de una muestra, de lo que llamaremos "estadísticos" (segun FISHER), los que se utilizan para estimar los valores de los parámetros de la población hipotética; (3) P r o b l e m a s d e d i s t r i b u -

(a) Ver nota del párrafo XIII.2.

c i ó n. Incluyen la discusión de la distribución de los "estadísticos" derivados de las muestras, o en general cualquier función de cantidades cuya distribución es conocida.

Sólo en casos particulares los problemas de especificación pueden recibir una solución directa, y ello ocurre cuando la forma de la ley de probabilidad se puede deducir de un conocimiento del mecanismo del fenómeno aleatorio. Por ejemplo, cuando se trata del caso del bolillero, resulta evidente que es fenómeno aleatorio de alternativa simple, individualizado por un sólo parámetro constante.

En otros casos, es necesario formular hipótesis para deducir la forma de la ley de probabilidad, y luego será necesario verificar esas hipótesis mediante un procedimiento sistemático que se discutirá en el capítulo siguiente. El ejemplo más famoso de este procedimiento es, precisamente la teoría de los errores. La hipótesis básica es: "El error cometido con un instrumento es el resultado de un número muy grande de "pequeños errores independientes entre sí, tales que cada uno de ellos no contribuye "más que a una pequeña parte del resultado"(a) . En forma matemática, estas son las hipótesis del teorema central del límite, (ver Cap. VIII, párrafos 12 y 13) por lo que sujeto a verificación posterior, resulta razonable considerar que el problema de especificación se resuelve en este caso con la elección de la ley normal.

Los problema de estimación y de distribución estan intimamente relacionados entre sí, como se verá en lo sucesivo. En este capítulo nos limitaremos a considerar principalmente el problema de la estimación, dejando para los dos siguientes un estudio mas detallado del de distribución.

Segun Fisher, el estimador que deberá emplearse para obtener una e s t i m a c i ó n (término con el que se substituirá a "estadístico", suceptible de confusiones) debe vincularse con el verdadero valor del parámetro mediante las siguientes condiciones:

(1) C o n s i s t e n c i a: Para un número elevado de observaciones, la ley de probabilidad del estimador debe concentrarse en torno al verdadero valor del parámetro, (la formulación original de Fisher es algo distinta) y equivale a establecer la convergencia en probabilidad (ver Cap. IV.14) del estimador, en función del número n de observaciones;

(2) E f i c i e n c i a: De todos los estimadores posibles, corresponde elegir el de varianza mínima. Esta condición es de fundamental importancia práctica, porque a menudo es posible construir distintos estimadores consistentes para un mismo parámetro. Por ejemplo, en el caso de una aplicación de la ley de sucesiones de LAPLACE, a un fe-

(a) H.POINCARE: "Calcul des Probabilités", París, G.Carré, 1896, pag. 181.-

nómeno aleatorio de alternativa simple también sería posible emplear el teorema de Bernouilli, teniéndose por lo tanto dos estimadores consistentes, como se verá en el párrafo siguiente;

(13.16) $$\bar{m} = \frac{r}{n} \quad ; \quad \bar{m} = \frac{r+1}{n+2}$$

para los que se ha empleado $\bar{m}$ para indicar que se trata de una estimación, y no de un verdadero valor del parámetro.

Pero además de esta discusión teórica, lo verdaderamente importante es que FISHER dió un método efectivo para construir estimadores a la vez consistentes y eficientes, y que se funda en construir la llamada f u n c i ó n d e v e r o s i m i l i t u d, o sea el producto de los valores que se obtienen al reemplazar en la ley de probabilidad los valores de la muestra, y hallar luego el valor del parámetro, en función de las observaciones, que hace máximo ese producto (a). Es de notar que este método es a s i n t ó t i c o, es decir, válido solamente en el límite para n indefinidamente creciente.

XIII.8.- Sea, por ejemplo, el caso señalado anteriormente, de un fenómeno aleatorio de alternativa simple y probabilidad constante, para el cual y de acuerdo con lo visto hasta ahora, se dispone de dos estimadores distintos en el caso de aplicar tambien la ley de sucesiones de LAPLACE, (debiendo estudiarse para este último, la vinculación con el verdadero valor segun el método de FISHER).

El primero de estos estimadores, indicados en (13.16), es evidentemente consistente, porque por aplicación del teorema de Benouilli (ver Cap.IV, párrafo 14) converge en probabilidad al verdadero valor p. Pero el segundo también lo es, dado que converge en probabilidad al mismo límite, y por lo tanto una posible selección debe fundarse en el criterio de eficiencia.

Construyendo las expresiones de la varianza de ambos estimadores, y estudiando su cociente, resulta que este último tiende al límite uno, de manera que también son asintoticamente equivalentes. Estudiando con más detalle el problema, se puede probar sin embargo que la varianza del primero es siempre inferior a la del segundo, de manera que entonces se tendría una razón para preferir el estimador fundado en el teorema de Bernouilli al dado por la ley de sucesiones, que queda relegado al papel de ejemplo histórico.

(a) La demostración rigurosa de que efectivamente por aplicación de este método se obtienen estimadores consistentes y eficientes (y tambien suficientes, condición auxiliar que no se estudiará en este curso), es muy posterior a la publicación de R.A.FISHER, y se debe a D.DUGUE:"Application des propietés de la limite au sens du calcul de probabilités a l'étude de diverses questions d'estimation." Journal de l'Ecole Polytechnique", vol. 3, (1937),pág. 305; tambien CRAMER (2), Cap.32, da una demostración rigurosa.

Pero es claro que todavía queda abierta la cuestión de si no existe un tercer estimador más eficiente que cualquiera de los dos dados. E evidente la dificultad del problema, que implica, por ejemplo, la delicada cuestión de si existe un estimador eficiente, es decir, si las varianzas de todos los posibles estimadores tienen, asintóticamente, un límite inferior, o en caso contrario, si existen infinitos estimadores que pueden ser ordenados en una sucesión tal que la varianza de cada uno es superior a la de todos los siguientes a partir de un cierto n.

Es precisamente en este caso donde resulta claro el poder del método de la máxima verosimilitud, que permite dar una respuesta mediante la aplicación de un procedimiento analítico simple.

La función de verosimilitud, conforme a su definición general y a la especificación en este caso particular de un fenómeno aleatorio de alternativa simple y probabilidad constante, es:

(13.16) $$L = p^r(1 - p)^{n-r}$$

Para hallar el valor de p en función de los valores de la muestra (en este caso r) que hace máximo a L, conviene pasar al logaritmo:

(13.17) $$\log L = r \log p + (n-r)\log(1-p)$$

Derivando con respecto a p:

(13.18) $$\frac{\partial \log L}{\partial p} = \frac{r}{p} - \frac{(n-r)}{1-p}$$

y haciendo esta derivada igual a cero, se tiene una condición que tiene que cumplirse cuando L es máxima:

(13.19) $$\bar{p} = \frac{r}{n}$$

en donde, como en (13.16), se indica con la barra que no se trata del valor exacto de p sino de una estimación.

Para probar que se trata efectivamente de un máximo, habría que calcular la derivada segunda y mostrar que es negativa, lo que no presenta dificultades.

El método de la máxima verosimilitud permite, pues, no sólo decidir entre los dos estimadores de (13.16), sino que cierra la cuestión, pues sería inútil buscar otro estimador, ya que no podría mejorarse en este caso el resultado obtenido mediante el teorema de Bernouilli.

XIII.9.- APLICACION DEL METODO DE LA MAXIMA VEROSIMILITUD A LA LEY NORMAL. Dada la frecuencia con que se presenta en la práctica la ley normal, tiene interés estudiar detalladamente la aplicación a dicho caso del método de la máxima verosimilitud.

La función de verosimilitud será:

(13.20) $$L = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_i-m)^2}{2\sigma^2}}$$

y su logaritmo será:

(13.21) $$\log L = \sum_{i=1}^{n}\left(\log \frac{1}{\sigma\sqrt{2\pi}} - \frac{(x_i-m)^2}{2\sigma^2}\right) = \sum\left(-\log\frac{1}{\sigma} - \lg\sqrt{2\pi} - \frac{(x_i-m)^2}{2\sigma^2}\right)$$

Los estimadores consistentes y eficientes de m y de σ en función de las observaciones x_i (supuestas independientes entre sí) serán precisamente las combinaciones d de las x_i que hagan máxima (13.20), o lo que es equivalente pero más sencillo de calcular, (13.21).

La condición necesaria que debe cumplirse es la anulación simultánea de las derivadas parciales de (13.21) con respecto a m y a σ:

(13.22) $$\frac{\partial \log L}{\partial m} = \frac{\sum(x_i - m)}{\sigma} = 0$$

$$\frac{\partial \log L}{\partial \sigma} = \sum\left(-\frac{1}{\sigma} + \frac{(x_i - m)^2}{\sigma^3}\right) = 0$$

Utilizando la primera derivada, se tiene:

(13.23) $$\bar{m} = \bar{x} = \frac{\sum x_i}{n}$$

en donde $\bar{m}$ indica que no es el verdadero valor sino una estimación, para la que se adopta usualmente la notación del promedio, indicada por x con una barra.

Reemplando m en la segunda derivada por su expresión (13.23), y teniendo en cuenta que para uniformar denominadores hay que multiplicar numerador y denominador del primer término por σ^2, resulta:

(13.24) $$\bar{\sigma}^2 = s^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

en donde nuevamente $\bar{\sigma}^2$ indica que se trata de una estimación de la varianza y no de su verdadero valor, siendo la convención más usual emplear en su lugar s^2 para evitar confusiones.

Naturalmente, para demostrar completamente que dichos estimadores hacen máxima (13.21), habría que calcular las derivadas segundas y hacer la discusión conocida, lo que no presenta dificultades.

XIII.10.- El método de la máxima verosimilitud confirma los métodos tradicionales de estimación empleados para la ley normal, que consisten en utilizar los momentos de las muestras (calculados en forma similar a los momentos de la ley, pero reemplazando la densidad de probabilidad por 1/n, y la integral por una suma), para estimar los momentos de la ley, que en este caso son los parámetros que la determi

nan completamente, de manera que el método de la máxima verosimilitud no parece agregar nada nuevo, salvo la distinción entre estimación y verdader valor y el tipo de relación entre los dos.

Pero en realidad, y como se observó anteriormente para el fenómeno aleatorio de alternativa simple, además de confirmar los métodos tradicionales, cierra la discusión del problema, por que hace claro que no es posible mejorar el resultado empleando un estimador distinto (salvo una corrección que se introducirá en el siguiente párrafo).

En efecto, cualquier otro estimador consistente no podrá tener una precisión mayor que (13.23) para estimar el primer momento.

Sea por ejemplo el estimador m e d i a n a de la muestra. Si se supone que la muestra tiene un número impar de observaciones $n = 2p+1$, se llama mediana de la misma a la observación que ocupa el lugar $p+1$, una vez que se ordenan todas segun su magnitud. Nuevamente, es una función de los valores de la muestra, cuya ley de probabilidad se puede determinar, y también cumple con la condición de consistencia, pero su varianza es mayor que (13.23), en la relación $\left(\frac{\pi}{2}\right)^2$, o sea que la dispersión de la mediana es 1,57 veces mayor que la del promedio(a).

En otros términos, si se utiliza la mediana, se precisaran 1.570 observaciones para obtener la misma precisión (medida en términos de la dispersión) que se alcanza con sólo 1.000 observaciones si se utiliza el promedio.

Es indudable que esta diferencia es suficiente para descartar el uso de la mediana, a pesar de que es más sencilla de determinar que el promedio.

Esta propiedad de eficiencia del promedio permite precisar la noción de "información contenida en la muestra" que utiliza Fisher al hacer el planteo de los problemas de inferencia estadística que se reprodujo en (XIII.7). Como no es posible mejorar el resultado obtenido con el promedio, se puede decir que esta agota la información contenida en la muestra acerca del verdadero valor del parámetro a estimar, mientras que la mediana, que en 1.570 observaciones obtiene el mismo resultado que el promedio con 1.000 desperdicia en realidad una cantidad de información contenidad en 1.570 observaciones, para extraer sólo la correspondiente a 1.000. Esta diferencia en la obtención de la información se puede medir por el cociente de las dispersiones (b).

XIII.11.- El uso práctico de los estimadores (13.23) y (13.24) obliga a una discusión

---

(a) Para ampliar este ejemplo, y en general todo el tratamiento del método de la máxima verosimilitud, es interesante la exposición de divulgación de G.DARMOIS: "L'Emploi des Observations Statistiques. Methodes d'Estimation", París, 1936, Hermann et Cie. (Actualités Scientifiques et Industrielles, Nº 356).
(b) Este concepto de información no es el mismo que el utilizado por la teoría de la información de Shannon, aludida en los capítulos III y IX.

adicional. En efecto, el método de la máxima verosimilitud se funda en teoremas asintóticos, es decir, válidos para un número de observaciones indefinidamente crecientes, pero en la realidad se trata siempre de un número finito de observaciones, y se presenta un problema de tipo parecido al discutir el uso de la fórmula de Laplace (ver Cap.VI párrafo 7).

Una indicación importante del tipo de error cometido al emplear una expresión finita para aproximar un límite se obtiene determinando las esperanzas matemáticas de (13.23) y (13.24).

Dada la propiedad aditiva de la esperanza matemática (ver Cap. V, párrafo 3), se tiene:

(13.25) $$E(\bar{x}) = E\frac{\sum x_i}{n} = \frac{1}{n}E(\sum x_i) = \frac{1}{n}nm = m$$

o sea que la esperanza matemática del estimador de máxima verosimilitud del primer momento de la ley normal es exactamente dicho primer momento.

Por el contrario, el estimador de máxima verosimilitud de la varianza no posee esa propiedad.

Para probar la afirmación anterior, conviene modificar la expresión (13.24), desarrollando el cuadrado:

(13.26) $$s^2 = \frac{\sum(x_i - \bar{x})^2}{n} = \frac{\sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2}{n}$$

y teniendo en cuenta que $\sum x_i = n\bar{x}$, se tiene finalmente:

(13.27) $$s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n}$$

Aplicando el operador esperanza matemática:

(13.28) $$E(s^2) = E\left(\frac{\sum x_i^2 - n\bar{x}^2}{n}\right) = \frac{1}{n}\left[E(\sum x_i^2) - nE(\bar{x}^2)\right]$$

La esperanza matemática del cuadrado de una variable aleatoria, si se supone que la variable está centrada, lo que no origina problemas, es precisamente la varianza, de manera que la ~~primer~~ 2ª esperanza matemática será la varianza de la ley de probabilidad del promediode la suma de n variables aleatorias centradas e iguales a independientes entre sí, la que teniendo en cuenta las propiedades aditivas de la ley normal (ver Cap.X párrafo 11) es simplemente $\frac{\sigma^2}{n}$. Aplicando estos resultados a (13.28):

(13.29) $$E(s^2) = \frac{1}{n}\left(n\sigma^2 - \sigma^2\right) = \frac{n-1}{n}\sigma^2$$

~~La esperanza matemática del cuadrado de una variable aleatoria, si se supone que la variable está centrada, lo que no origina problemas, es precisamente la varianza, de~~

la esperanza matemática del estimador de la varianza de una ley normal es igual a la varianza multiplicada por una constante positiva inferior a la unidad.

Este resultado no es incompatible con la condición de consistencia enunciada anteriormente (párrafo XIII. 7), porque el límite de la constante multiplicativa para n indefinidamente creciente es la unidad, pero tiene la importante consecuencia de que si se emplea sistemáticamente el estimador (13.24), en general se subestimará las varianzas porque el promedio de las estimaciones, aplicando una ley de los grandes números, converge en probabilidad al promedio de las esperanzas matemáticas, que es menor que el promedio de la suma de las varianzas.

El defecto que se acaba de señalar para el estimador (13.24) de la varianza, que no existe en el (13.23) del primer parámetro, se soluciona facilmente multiplicandolo por la constante correctora $\frac{n}{n-1}$, con lo que se obtiene un estimador no viciado o tambien insesgado de la varianza (denominaciones que provienen a su vez de las de vicio o sesgo para la diferencia entre la esperanza matemática del estimador y el verdadero valor del parámetro):

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \tag{13.30}$$

Actualmente se ha generalizado en los libros de estadística el uso del estimador insesgado (13.30) con la notación s^2 pero hay también textos que continuan empleando (13.24), por lo que hay que tener cuidado en verificar la notación que emplea cada autor, a fin de evitar confusiones cuya importancia se hará aparente en el capítulo XV.

En lo que sigue, se usará la notación s^2 para el estimador insesgado, siguiendo la práctica corriente.

Cuando se emplea el estimador corregido, ya no cumple con la condición de eficiencia y por lo tanto la elección entre emplear o no la corrección depende de la importancia que tenga el sesgo.

Para las muestras de pequeño número de observaciones (caso general de las mediciones de laboratorio), es evidente que el error debido al sesgo es predominante, y habrá que emplear la corrección. Cuando n es grande la influencia de la corrección es muy pequeña, y puede prescindirse de ella.

El límite práctico está dado por la tablas en cuyo uso que tiene importancia la forma exacta de s^2, como se verá en el Capítulo XV, y que fijan n máximo para la corrección en 30, en los casos usuales.

XIII.12.- DISPOSICION DE LOS CALCULOS. Aunque la forma de los estimadores (13.23) y (13.24) es sencilla, su cálculo efectivo puede complicarse debido no sólo al número po -

sible de observaciones sino también a la presencia de comas y fracciones.

A fin de reducir las operaciones a un mínimo, y entre números simples, conviene multiplicar por una constante para eliminar las comas, reemplazar las observaciones restandoles una constante que aparece como una estimación aproximada del promedio y utilizar el desarrollo((13.27) para calcular s^2, como se indica en el cuadro siguiente:

i	x_i	$10x_i$	$y_i = 10x_i - 6$	y_i^2
1	0,2	2	-4	16
2	0,6	6	0	0
3	0,5	5	-1	1
4	1,0	10	4	16
5	0,8	8	2	4
6	0,9	9	3	9
7	0,4	4	-2	4
8	0,7	7	1	1
			$\sum y_i = 3$	$\sum y_i^2 = 51$

$$\bar{y}_i = \frac{3}{8} = 0{,}375 \quad ; \quad s^2 = \frac{51 - 8.0{,}1406}{7} = \frac{51 - 8x0.141}{7} = 7.125$$

$$\bar{y}_i^2 = 0{,}1406 \simeq 0{,}141$$

A las observaciones x_i se la multiplica por 10 para hacer desaparecer las comas, y observando que los extremos son 2 y 10 se adoptó como media de cálculo su promedio 6, obteniéndose la columna de las y_i, de la que resulta inmediatamente la suma 3. Agregando una columna con los cuadrados de las y_i, a la suma de estos cuadrados se le resta n veces el promedio al cuadrado para obtener s^2 dividiendo dicha suma por n-1.

El promedio de las x_i se obtiene sumando la media de cálculo al promedio de las y_i y dividiendo por la constante multiplicativa. El s^2 de las x_i resulta simplemente de dividir el de las y_i por la constante multiplicativa al cuadrado. El resultado final es:

(13.31) $$\bar{x} = \frac{6{,}375}{10} = 0{,}64 \quad ; \quad s^2 = 0{,}071$$

XIII.13.- APLICACION DEL METODO DE LA MAXIMA VEROSIMILITUD A LA LEY DE POISSON. Sea una sucesión de observaciones que se supone que siguen una ley de Poisson, y se trata de construir el estimador consistente y eficiente del parámetro α, que es desconocido

Las observaciones se clasifican en k grupos (en donde k es el número máxi

mo de veces que se ha observado el fenómeno, por ejemplo, emisión de una partícula radioactiva, en un intervalo de longitud fija de un parámetro continuo, que en el ejemplo dado es el tiempo), y si y_j es el número de veces que se ha encontrado un intervalo con j observaciones, se tendrá:

$$\sum_{j=0}^{k} j.y_j = n \quad ; \quad \sum_{j=0}^{k} y_j = m \tag{13.32}$$

en donde n es el total de veces que se ha observado el fenómeno, y m el total de intervalos. Para formar la función de verosimilitud hay que calcular la probabilidad de observar un intervalo con j emisiones, la que conforme con la fórmula (7.4) del Cap. VII, párrafo 2, será

$$\frac{\alpha^j}{j!} e^{-\alpha} \tag{13.33}$$

La función de verosimilitud es:

$$L = \prod_{j=0}^{k} \left(\frac{\alpha^j}{j!} e^{-\alpha} \right)^{y_j} \tag{13.34}$$

y su logaritmo

$$\log L = \sum_{j=0}^{k} y_j (j \log \alpha - \log j! - \alpha) \tag{13.35}$$

El valor atribuído a α que hace máximo (13.34) anula la primera derivada del logaritmo:

$$\frac{\partial \log L}{\partial \alpha} = \sum_{j=0}^{k} y_j \left(\frac{j}{\alpha} - 1 \right) \tag{13.36}$$

de donde resulta:

$$\bar{\alpha} = \frac{\sum j.y_j}{\sum y_j} = \frac{n}{m} \tag{13.37}$$

en la que como siempre se ha utilizado la notación con barra para indicar que no se trata del verdadero valor sino de una estimación y queda confirmado el procedimiento utilizado en (VII.4), que allí se fundó en el teorema de Bernouilli generalizado, es decir, en lo que en inferencia estadística es la condición de consistencia; y la aplicación del método de la máxima verosimilitud permite afirmar que dicho estimador es, además, asintóticamente eficiente.

XIII.14.- OTRAS APLICACIONES DEL METODO DE LA MAXIMA VEROSIMILITUD. El hecho de que tambien en la ley de Poisson el momento de la muestra haya resultado el estimador más conveniente no significa que pueda hablarse en general de que los momentos son siempre utilizables para estimar parámetros. Precisamente el método de la máxima verosimilitud ha permitido poner en claro en qué casos conviene utilizar dichos momentos, y cuando es preciso emplear otras funciones de los valores de la muestra, a fin de obtener estimado-

res consistentes y eficientes.

El el capítulo X (párrafo 13), se determinó la forma analítica de una ley de probabilidad que tiene interés para algunos problemas de física y de estadística, la ley de Cauchy, suponiendo conocida la posición del centro de emisión de partículas con respecto a la pantalla. En el caso de que se conozca solamente la distancia del centro a la pantalla, pero no la ubicación del pié de la perpendicular, y se adopta un origen arbitrario de coordenadas sobre la pantalla, la función de densidad de probabilidad se modifica:

(XIII.38) $$f(x) = \frac{1}{\pi} \frac{1}{1 + (k - m)^2}$$

expresión en la que, como en el capítulo X, la distancia del centro a la pantalla se ha tomado como unidad y el parámetro m, como resulta evidente, indica la posición del pié de la perpendicular trazada por el centro de emisión, en el sistema arbitrario de coordenadas. Por consiguiente, la estimación del parámetro y la ubicación del centro de emisión constituyen el mismo problema.

Siguiendo la práctica tradicional de utilizar el primer momento de la muestra, dado su éxito en la ley normal y en la de Poisson, no se llega en este caso a ningun resultado.

En efecto, la ley de probabilidad del promedio de los valores de las observaciones de una ley de Cauchy, es precisamente la misma ley (ver CRAMER (2)), y por lo tanto tiene el mismo efecto calcular el promedio para estimar m que utilizar una observacion aislada. El procedimiento no tendría sentido, pues precisamente lo que se busca al construir una función de los valores de la muestra es diminuir el campo de variación, e idealmente reducirlo a un punto sí el número de observaciones es suficientemente grande.

En cambio, aplicando el metodo de la máxima verosimilitud es posible encontrar un estimador para m, que resulta ser una función complicada de las observaciones. Si se prefiere una forma más sencilla, es posible emplear la mediana (definida en XIII.10) que en el sentido precisado al discutir su uso en la ley normal, resulta que en este caso utiliza el 80% de las observaciones (a), o sea que con 100 observaciones obtiene la misma precisión que el estimador eficiente con 80.

XIII.15.- METODO DE ESTIMACION DE NEYMAN-MARKOFF. En el párrafo XIII.11 se comprobó que no siempre resulta conveniente utilizar un estimador determinado mediante el método de la máxima verosimilitud, si como ocurre con la varianza de la ley normal, el estimador es sesgado o viciado, es decir, su esperanza matemática no coincide con el verdadero valor del parámetro, por lo que para un número limitado de observaciones se presenta un error

(a) Ver G. DARMOIS (obra citada en XIII.10 pag 18)

sistemático que puede no ser de magnitud despreciable.

Parece natural preguntarse si no es posible encontrar un método de aplicación general, capaz de encontrar estimadores en los que la condición de consistencia asintótica del método de la máxima verosimilitud sea transformada en la condición exacta de que la ley de probabilidad del estimador tenga por esperanza matemática el verdadero valor, cumpliendo además la condición de eficiencia de tener la varianza mínima entre todos los estimadores insesgados que pueden construirse.

JERZY NEYMAN (a), utilizando un teorema de cuadrados mínimos debido a MARKOFF (19) logró establecer un método general, que aplicado a la ley normal permite establecer que el estimador l i n e a l insesgado más eficiente de m es el promedio de los valores de la muestra, y que el estimador c u a d r á t i c o insesgado más eficiente de la varianza es el dado por (13.30).

El interés especial de este método de NEYMAN-MARKOFF es que permite resolver problemas complejos que se presentan en el diseño de muestras que se utilizan para los relevamientos estadísticos (b).

Un ejemplo sencillo de este método es el de la estimación del coeficiente de regresión. (ver capítulo XII, párrafos 5 y 6).

Sea una sucesión de observaciones X_i Y_i, correspondientes cada una a un valor de una variable X_i. Se supone que la variable X_i se mide con exactitud, pero que la determinación de la Y_i incluye un error experimental, que sigue una ley normal. Los errores en cuestión son independientes de X_i. Si entre X_i e Y_i existe una relación funcional de tipo lineal, será preciso calcular los parámetros de la recta eliminando de alguna manera la influencia de los errores.

En este caso, el método de NEYMAN-MARKOFF conduce simplemente a hallar los valores de α y de β que hacen mínima la suma de cuadrados de las diferencias entre la ordenada de la recta correspondiente a un cierto valor de X_i y la observación Y_i:

$$\sum_i = (Y_i - \alpha - \beta X_i)^2 \tag{13.39}$$

Igualando a cero las derivadas parciales con respecto a los parámetros desconocidos se obtiene un sistema de ecuaciones que debe ser satisfecho por los valores de α y β que hacen mínima la suma de cuadrados:

$$\frac{\partial \sum}{\partial \alpha} = -2 \sum (Y_i - \alpha - \beta X_i) = 0$$

(a) J. NEYMAN: "On the two different aspects of the representative methods", Journal of the Royal Statistical Society, vol. 97 (1934), pag 558.

(b) La teoría de los relevamientos estadísticos por muestra ha dado lugar a una extensa literatura especializada. Ver, entre otros P.G. SUKHATME: "Sampling Theory of Surveys with applications", Iowa, Iowa State College Press, 1954 (hay traduccion castellana).

$$\frac{\partial \sum}{\partial \beta} = -2 \sum X_i (Y_i - \alpha - \beta X_i) = 0$$

Reordenando los resultados de las operaciones, se tiene:

$$n\alpha + \beta \sum X_i = \sum Y_i$$

(13.40)

$$\alpha \sum X_i + \beta \sum X_i^2 = \sum X_i Y_i$$

llamadas las e c u a c i o n e s n o r m a l e s, y cuya solución daria los valores buscados de los parámetros.

El procedimiento empleado coincide en este caso simple con el método de cuadrados mínimos, introducido por LEGENDRE en 1806, y descubierto independientemente por GAUSS en 1809, que lo hizo célebre con su teoría de los errores. Pero debe hacerse notar que existen diferencias fundamentales de concepto, porque en el método de Gauss se emplean probabilidades "a priori", que no aparecen en el de NEYMAN-MARKOFF. Por otra parte, dentro de la teoría moderna de la inferencia estadística, y como se verá en el capítulo XV, la interpretación de los resultados y la eficiencia de los métodos son muy superiores a lo que permite la teoría de Gauss, la que por otra parte sería inaplicable de no aceptarse la existencia de probabilidades "a priori" (a)

XIII.16.- DISPOSICION DE LOS CALCULOS. Utilizando algunas simplificaciones convenientes, es posible evitar la solución efectiva del sistema de ecuaciones lineales(13.40).

En efecto, resolviendo la primera ecuación con respecto a α, se tiene:

(13.42) $$\bar{\alpha} = \frac{\sum Y_i}{n} - \frac{\sum X_i}{n} = \bar{y} - \beta \bar{x}$$

Reemplazando el valor de $\bar{\alpha}$ en (13.39) ~~la segunda:~~

(13.43) $$\sum [Y_i - \bar{y} - \beta (X_i - \bar{x})]^2 = \sum (y_i - \beta x_i)^2$$

en donde las letras minúsculas indican que las variables X_i e Y_i han sido c e n t r a d a s, restandoles los respectivos promedios.

Con esta nueva expresión, la derivada respecto de β es:

(13.44) $$\frac{\partial \sum}{\partial \beta} = -2 \sum x_i (y_i - \beta x_i) = 0$$

de donde se despeja

(13.46) $$\bar{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

(a) Algunos ejemplos del uso del teorema de Markoff pueden verse en la obra de DAVID (24)

Teniendo en cuenta el desarrollo de la suma de cuadra os empleado en (13.27) se tiene:

$$\sum x_i^2 = \sum (X_i - \bar{x})^2 = \sum X_i^2 - n\bar{x}^2$$

(13.46)

$$\sum x_i y_i = \sum (X_i - \bar{x})(Y_i - \bar{y}) = \sum X_i Y_i - \bar{x}\sum Y_i - \bar{y}\sum X_i + nxy =$$

$$= \sum X_i Y_i - n\bar{x}\bar{y}$$

Por lo tanto, en un cuadro en el que figuran como dos primeras columnas las correspondientes a X_i e Y_i basta agregar otras dos columnas con los productos $X_i Y_i$ y los cuadrados X_i^2, y efectuar las sumas y promedios para obtener las estimaciones insesgadas de los parámetros:

(13.47) $$\bar{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} \quad ; \quad \alpha = \bar{y} - \beta\bar{x}$$

XIII.15.- TRATAMIENTO DE LAS OBSERVACIONES. Dada una sucesión de observaciones, la reducción de las mismas depende, conforme con lo visto en este capítulo, del tipo de análisis que se quiere realizar.

Si el problema de la especificación está resuelto, por ejemplo, en favor de la ley normal, habrá que calcular los momentos de primer y segundo orden de la serie de observaciones, para obtener estimaciones de los parámetros.

Si se trata de observaciones que siguen la ley de Poisson, bastará con el primer momento, y no tiene sentido calcular el segundo. En cambio, si la solución del problema de especificación indica la ley de Cauchy, bastará con la mediana.

Como se dijo anteriormente, no existen reglas generales para resolver el problema de especificación. Sólo mediante el conocimiento directo del mecanismo del fenómeno aleatorio (casos de la ley de Poisson y de Cauchy), o bien por medio de una teoría especialmente formulada cuyos resultados habrá luego que verificar (caso de la teoría de los errores, en que hipótesis sobre el comportamiento de los errores se traducen por medio del teorema central del límite en la ley normal), será posible orientarse.

Cuando el problema se reduce a hallar una ley de probabilidad de forma sencilla la reducción e interpretación de las observaciones es también simple. Pero cuando la especificación depende de una teoría compleja, sólo parcialmente comprobada, y que da lugar a complicados problemas de estimación, como por ejemplo en el caso de la econometría, que busca interpretar los resultados de las compilaciones de estadísticas eco-

nómicas a la luz de la teoría económica, se convierte en un caso de alta especialización.

En muchos casos, se renuncia a toda tentativa de especificación, y la reducción de las observaciones es nada más que un tratamiento descriptivo experimental, calculando promedios, cuartilos (valores tales que una cuarta parte de las observaciones se hallan por encima, o por debajo), etc., limitandose considerablemente las posibilidades de analisis, y desapareciendo la posibilidad de hacer predicciones.

CAPITULO XIV

VERIFICACION DE HIPOTESIS.

XIV.1.- En el capitulo anterior se señaló la necesidad de desarrollar un método de verificación de hipótesis, a fin de comprobar la solución que se de al problema de la especificación.

Este problema aparece no solamente como una etapa previa al de la estimación, sino que tiene interés de por sí, como un segundo procedimiento para interpretar las observaciones.

En efecto, si el problema de la estimación, una vez resuelto satisfactoriamente, permite realizar predicciones sobre el comportamiento futuro (por ejemplo, en un fenómeno cuya probabilidad de aparición depende en forma monótona creciente de una variable continua, como parece razonable en el caso de la emisión de partículas radioactivas, la estimación del parámetro de la ley de Poisson correspondiente permite predecir en un intervalo futuro de longitud determinada, y en términos de probabilidades, cuantas emisiones se podrán observar), es tambien importante poder comprobar si la ley que sigue un fenómeno ha variado, es decir, si dadas dos series de observaciones es posible afirmar que siguen la misma ley de probabilidad o no, lo que puede hacerse comparando las estimaciones correspondientes de los parámetros, como se verá en el capítulo siguiente, o bien directamente sin estimar parámetros.

XIV.2.- En lo que sigue, se entenderá por hipótesis una ley de probabilidad. Si la ley está completamente determinada, es decir, se da su forma analítica y el valor de los parámetros que la individualizan, se dice que se trata de una h i p ó t e s i s s i m p l e. Un ejemplo de hipótesis simple sería una ley normal, con valores numéricos determinados de sus parámetros m y σ.

Pero si la ley está incompletamente determinada (por ejemplo, una ley normal con un valor dado de σ, pero m puede tomar cualquier valor dentro de un intervalo fijado) se tratará de una h i p ó t e s i s c o m p u e s t a, terminología justificada porque una hipótesis compuesta está formada por un conjunto de hipótesis simples (las que resultan de dar al parámetro no fijado, todos los valores compatibles con la hipótesis, o sea en el ejemplo anterior, dando a m un valor comprendido en el intervalo, se obtiene una hipótesis simple integrante de la hipótesis compuesta).

XIV.3.- Un ejemplo que tiene cierto interés histórico servirá para aclarar lo anterior.

En el siglo XVIII, el naturalista BUFFON decidió verificar experimentalmente si una moneda era equilibrada, para lo cual realizó 4.040 tiros, de los que 1.992 mostraron una cara y 2.048 la otra. ¿Como proceder para interpretar esta serie de observaciones, a fin de contestar a la pregunta de BUFFON?.

En la terminología introducida, la hipótesis a verificar es que se trata de un fenómeno aleatorio de alternativa simple, con probabilidad 1/2 . Como la ley de probabilidad está determinada completamente por el parámetro $p=1/2$, se trata de una hipótesis simple.

Una primera solución sería calcular la probabilidad del desvío entre el valor esperado $n.p=4.040 \times 1/2=2.020$ y el observado $x=r - np=2.048 - 2.020=28$.

Utilizando la aproximación normal de la ley binomial que sigue la probabilidad de r alternativas en n (ver capítulos IV y VI), el desvío reducido es

(14.1) $$t = \frac{x}{\sigma} = \frac{r - np}{\sqrt{npq}} = 0,89$$

En la tabla I de CRAMER (3) se puede observar que la probabilidad de que el valor absoluto del desvío t supere 0,9 es aproximadamente 0.36.

(14.2) $$P(|t| \geqslant 0,9) = 2(1 - 0,81592) \cong 0,36$$

Intuitivamente, el resultado no es tan improbable como para desecharlo, pero es evidente la necesidad de dar mayor precisión a la discusión que antecede.

En efecto, si se adopta un cierto límite para el valor de la probabilidad, rechazando la hipótesis si resulta menor que este límite, se puede cometer un error, consistente en r e c h a z a r l a h i p ó t e s i s c u a n d o e s c i e r t a o e r r o r d e p r i m e r a c l a s e, porque por bajo que se haya fijado el límite de la probabilidad, que se conviene en llamar n i v e l d e s i g n i f i c a c i ó n, una probabilidad pequeña no significa imposibilidad.

En la interpretación de las probabilidades como frecuencias relativas, fijar un nivel de significación equivale a establecer un límite para el error de primera clase, si ese nivel se aplica en todos los casos. Un nivel de significación usual como el 0,05 implica que si el procedimiento de verificación de la hipótesis se aplica sistemáticamente, en promedio se rechazará la hipótesis cuando es cierta un 5% de las veces (o un 1% , si el nivel de significación es de 0,01).

Pero también es claro que es posible cometer un e r r o r d e s e g u n d a c l a s e, que consiste en aceptar la h i p ó t e s i s c u a n d o e s f a l s a, y que ocurrirá cuando la probabilidad del desvío reducido que se observa en un ejemplo similar al anterior, es superior al nivel de significación fijado, pero el parámetro es distinto de 1/2.

Mientras que el error de primera clase es fácil de calcular, el error de segunda

clase plantea problemas más complicados, pero es tan importante como el de primera, y es evidente que un criterio racional de verificación de hipótesis debe tener en cuenta ambos tipos de errores.

XIV.3.- Dar un método de verificación de hipótesis significa elegir una ley de probabilidad, y además una región crítica dentro del campo de variación de la variable aleatoria, tal que si el punto representativo de las observaciones cae dentro de esa región se rechaza la hipótesis.

En el ejemplo anterior, la ley de probabilidad era la normal, y la región crítica, si se adopta el nivel de significación 0,05, serían los dos segmentos indefinidos a la izquierda y a la derecha a una distancia 1,96, a contar del origen.

Pero es evidente qu la elección del nivel de significación no identifica una región crítica determinada. Existen muchas regiones a las que la ley asigna la probabilidad 0,05, y si se eligió la región simétrica, fué debido a las siguientes consideraciones: (1) Parece razonable elegir como región crítica la que se aparta más del origen (intuitavamente no se aceptaría como región crítica una que contenga el origen, aunque su probabilidad sea también de 0,05, porque si bien su nivel de significación, o sea la probabilidad del error de primera clase es la fijada, en el caso de que la hipótesis fuera falsa y el parámetro no fuera igual a 1/2 si no por ejemplo a 1/4, habría una probabilidad muy elevada de aceptar la hipótesis cuando es falsa); (2) Las desviaciones a un lado y a otro del valor esperado son igualmente inaceptables.

Para cumplir con la condición de dar adecuada consideración a los dos tipos de errores, habría que fijar no solamente un nivel de significación, sino también la probabilidad del error de segunda clase. Pero esto en general es imposible, por las siguientes razones: (1) Calcular la probabilidad del error de segunda clase, importa fijar una alternativa, transformandose el problema en decidir si una de dos hipótesis es cierta; (2) No sólo puede ocurrir que no exista región crítica que cumpla con las dos condiciones si las alternativas son simples, sino que el caso más frecuente es que se trata de verificar una hipótesis simple contra una hipótesis compuesta (en el ejemplo anterior, verificar si una moneda está equilibrada, consiste en realidad en decidir entre la alternativa $p=1/2$, y la compuesta p igual a cualquier valor del intervalo (0,1), salvo 1/2), y como a una misma región crítica las distintas leyes correspondientes a una hipótesis compuesta asignan distintas probabilidades, el problema así planteado no tendría sentido.

La solución consiste en establecer condiciones menos rigurosas sobre el error de segunda clase. En vez de fijar un valor, se podría investigar si es posible determinar una región crítica tal que la probabilidad de no cometer el error de segunda clase, tam

bién llamada p o t e n c i a o p o d e r s e p a r a d o r, por razones facilmente comprensibles, fuera máxima para cada una de las leyes que integran la hipótesis compuesta alternativa, mientras que se mantiene el nivel de significación. Un método que cumpliera con esta condición se denomina u n i f o r m e n t e m a s p o d e r o s o.

No siempre es posible encontrar una región crítica que cumpla con la condición anteriormente enunciada, y entonces será necesario investigar que otras condiciones se pueden imponer, de tal manera que el procedimiento de verificación de hipótesis resulte dando un peso adecuado a los dos tipos de errores.

Se comprende facilmente que la teoría general de verificación de hipótesis es muy compleja, y que en cuanto los problemas se complican resultan en realidad de la competencia de especialistas (a).

XIV.4.- Sin entrar en el detalle de la teoría, cuyas dificultades se acaban de señalar, conviene discutir el procedimiento empleado en el caso de la experiencia de BUFFON con la moneda, a fin de fijar ideas.

La elección de una región crítica simétrica en torno al valor esperado $np=2.020$ de la hipótesis simple que se trata de verificar, implica rechazar a dicha hipótesis como inexacta cuando el desvío reducido supera en valor absoluto a 1,96. Si en lugar de los desvíos reducidos en torno al valor esperado, se pasa al desvío sin reducir que es igual en valor absoluto a 62,298, la región crítica resulta comprender a las observaciones que en una serie de 4.040 tiros superan a 2.082 o estan por debajo de 1.958 apariciones de una misma cara.

Pero si la moneda es desequilibrada, y el valor del parámetro es en realidad $p = 0,54$, el valor esperado correspondiente sería 2.181,6, y se tendrá un resultado en la región crítica solamente si el desvío no reducido es mayor que -99 (obtenido por diferencia entre el límite superior de la región crítica fijado anteriormente y el nuevo valor esperado, ya que la otra parte simétrica de la región crítica resulta tan alejada que le corresponde una probabilidad nula si la moneda está desequilibrada con $p=0,54$). La probabilidad de rechazo de la hipótesis se obtiene calculando mediante la aproximación normal la probabilidad de caer en la región crítica, la que teniendo en cuenta que $\sqrt{npq} = 31,671$, resulta igual al área de la curva normal comprendida entre $-3,12$ y $+\infty$, que es 0,99910 (empleando una tabla con entrada hasta el centésimo de desvío reducido,

(a) La teoría moderna de la verificación de hipótesis tiene su origen en las siguientes publicaciones: J.NEYMAN and E.S.PEARSON: "On the use and interpretation of certain test criteria for purposes of statistical inference", Biometrika, vol.20 A (1928), pag.175 y 263, y ha recibido desde entonces considerable atención, traducida en casi un centenar de memorias importantes. Un resumen se puede consultar en (25) y en (26).

como la de la colección (27), y reteniendo hasta la quinta cifra decimal). Por lo tanto la probabilidad del error de segunda clase (aceptar la hipótesis cuando es falsa), será en este caso igual a 1-0,99910, o sea 0,00090. En términos de frecuencias relativas significa que si la moneda es desequilibrada con $p=0,54$ y se realizan 4.040 tiros con ella para decidir si es equilibrada, con un nivel de significación de 0,05, se aceptará la hipótesis cuando es falsa en nueve de diez mil series de 4.040 tiros.

Aunque en el ejemplo dado la probabilidad del error de segunda clase es muy pequeña es evidente que para completar la discusión del método empleado para verificar la hipótesis de que la moneda es equilibrada, es necesario calcular la probabilidad del error de segunda clase en función de una sucesión de valores de p, como aparece en la tabla siguiente:

COMPARACION DE LA PROBABILIDAD DE ERROR DE SEGUNDA CLASE Y DE LA POTENCIA (O PODER SEPARADOR) EN EL CASO DE VERIFICAR LA HIPOTESIS $p=0,5$ CON 4.040 OBSERVACIONES Y UN NIVEL DE SIGNIFICACION DE 0,05.

p	np	x=2.082-np	$\sqrt{npq}$	t	Probabilidad del error de segunda clase	Potencia
0,51	2.060,4	21	31,769	0,66	0,74537	0,25463
0,52	2.100,8	-18	31,754	-0,57	0,28434	0,71566
0,53	2.141,2	-59	31,724	-1,86	0,03144	0,96856
0,54	2.181,6	-99	31,671	-3,12	0,00090	0,99910
0,55	2.222,0	-140	31,621	-4,42	0,00000	1,00000

Como es evidente, para valores de p muy cercanos a 1/2, la probabilidad del error de segunda clase es grande (para $p=0,51$ es mayor que 0,74), y disminuye rápidamente al aumentar la diferencia entre los valores del parámetro.

Si se considera que la probabilidad de aceptar la hipótesis cuando es falsa, con $p=0,52$, empleando el mismo procedimiento de realizar una sucesión de tiros y calcular la probabilidad de los desvíos mediante la aproximación normal es excesivamente alta, quedan dos alternativas. La primera es disminuir la amplitud de la r e g i ó n d e a c e p t a c i ó n (segmento complementario de la región crítica), aumentando el nivel de significación. La segunda es conservar el nivel de significación y aumentar el número de tiros.

Se comprende facilmente que si se considerara necesario disminuir el nivel de significación para reducir la probabilidad del error de primera clase, ampliando la región de aceptación, disminuye la región crítica, aumentando la probabilidad del error de se-

gunda clase, y es necesario llegar a un compromiso.

Este tipo de problemas se encuentra en la aceptación de materiales por muestra, en donde al error de primera clase recibe el nombre de r i e s g o d e l p r o d u c t o r, ya que mide la probabilidad de rechazar una partida cuando en realidad está de acuerdo con la especificación, y el de segunda clase se llama r i e s g o d e l c o n s u m i d o r, porque indica la probabilidad de aceptar como conforme a una cierta especificación una partida que en realidad no lo está. En CRAMER (3), párrafo 16.6 se encuentra discutido en forma ilustrativa esta tipo especial de verificación de hipótesis, para el que se han calculado tablas, como la referida en (28), que para determinadas especificaciones (porcentaje de elementos defectuosos), indican planes de muestra (número de elementos a ensayar) con determinados riesgos.

XIV.5.- Una vez fijado el nivel de significación, determinada la región crítica y calculada la función de potencia (o, lo que es equivalente, la probabilidad del error de segunda clase, que tomando como variable el parámetro es la llamada c a r a c t e r í s t i c a o p e r a t i v a), queda terminada la discusión estadística de un procedimiento de verificación de hipótesis.

Es interesante enunciar en forma precisa el significado operacional de toda la discusión anterior. Segun QUENOUILLE (a):"Existe un viejo dicho segun el cual es posible "probar cualquier cosa mediante la estadística. Mientras que esta afirmación aparece co"mo cierta sola para la estadística utilizada incorrectamente, la recíproca se aplica "con igual validez a la estadistica correcta: E s i m p o s i b l e p r o b a r a l "g o m e d i a n t e l a e s t a d í s t i c a, ya que los métodos estadísti-"cos correctamente aplicados nunca permiten probar experimentalmente nada. L o q u e "e s t o s m é t o d o s p e r m i t e n, s i n e m b a r g o, e s a v e r i-"g u a r l a p o s i b i l i d a d d e e r r o r e n u n a a f i r m a-"c i o n, o l a c o n f i a n z a q u e p u e d e d e p o s i t a r s e "e n u n a m e d i c i ó n e x p e r i m e n t a l."

En efecto, y conforme con la discusión anterior, lo que se logra es acotar la probabilidad de los dos tipos de error, en una forma que permite afirmar que si el procedimiento se emplea sistemáticamente, en promedio se rechazará la hipótesis cuando es cierta en un porcentaje de veces igual al nivel de significación, y se aceptará una hipótesis cuando es falsa tambien en un porcentaje de veces que depende del valor que tenga en la realidad el parámetro.

(a) "The Design and Analysis of Experiment", Londres, Griffin & Co. 1953, pag.3. (los subrayados no son del autor).

XIV.6.- EMPLEO DE JI CUADRADO. A menudo el problema de verificar una hipótesis se plantea en una forma más sencilla que en el caso de la inspección para la aceptación de materiales conforme a una especificación. Lo que interesa muchas veces es hacer mínimo el error de primera clase, mientras que el error de segunda clase no se considera perjudicial a menos que los valores del parámetro difieran mucho del que fija la hipótesis.

Un procedimiento que sea de aplicación en una variedad de casos, y que asegure una función de potencia rapidamente creciente sería de uso indicado, y este el método de JI cuadrado, inventado por KARL PEARSON en 1900.

El método consiste en calcular las desviaciones entre el valor esperado y los valores observados elevar al cuadrado esas desviaciones (para evitar la compensación de las desviaciones de distinto signo), dividir a cada uno de los cuadrados por el valor esperado, para tener una medida relativa de la importancia de la desviación, y sumar todas las medidas relativas así obtenidas.

Aplicando al problema de BUFFON, se tiene:

$$(14.3)\quad \chi^2 = \frac{(r_1 - np)^2}{np} + \frac{(r_2 - np)^2}{np} = \frac{(p+q)(r_1 - np)^2}{npq} = \frac{(r_1 - np)^2}{npq} = 0{,}776$$

El resultado es precisamente el cuadrado de una variable que es asintóticamente normal, y por lo tanto tendrá, también asintóticamente, una ley como la designada bajo el nombre de "ley de Ji cuadrado" en el Cap. X, párrafo 7 (la designación de Ji cuadrado corresponde a la letra griega que PEARSON empleó para identificar la medida relativa de la desviación).

Empleando la tabla III de CRAMER (3), se observa que para el cuadrado de una variable normal, el valor 3,841 se supera con una probabilidad igual a 0,05. Si se adopta, como antes, dicha probabilidad como nivel de significación, el resultado es ampliamente satisfactorio, en el sentido de considerar verificada la hipótesis de que la moneda es equilibrada.

Empleando la ley normal, y no su cuadrado, se había llegado antes al mismo resultado, por un procedimiento que requería menos cálculos (no se elevaba al cuadrado), pero la ventaja decisiva en favor del procedimiento del Ji cuadrado es su mayor poder separador En efecto , si se calcula la probabilidad del error de segunda clase, en caso de que el valor exacto del parámetro fuera $p = 0{,}51$, es solamente 0,49093, contra 0,74537 utilizando la normal, y la potencia (o poder separador), naturalmente, es mucho mayor (0,50907 contra 0,25463). Esto se debe a que al emplear la normal se utiliza solamente un desvío, mientras que con el Ji cuadrado se combinan los dos, e indica que en todos los casos en que aparece como utilizable la ley normal para verificar una hipótesis, conviene en realidad emplear el Ji cuadrado.

XIV.7.- El empleo correcto del método del Ji cuadrado requiere analiazar más cuidadosamente el ejemplo de la moneda de BUFFON.

La medida de las desviaciones relativas expresada en (14.3) es una suma de variables, cada una de las cuales es a su vez proporcional a una variable asintóticamente normal. Pero la suma misma no era proporcional sino directamente igual al cuadrado de una variable asintóticamente normal, lo que se mostró analíticamente sin dificultad en la misma expresión (14.4).

Esta propiedad se debe al hecho de la existencia de una relación entre las desviaciones de los valores observados y los valores esperados:

(14.5) $$(r_1 - np) + (r_2 - nq) = (r_1 + r_2) - n(p + q) = 0$$

dado que $r_1 + r_2 = n$, porque tratandose de un fenómeno aleatorio de alternativa simple la suma del número de veces que se presenta cada observación es igual al total de la observaciones, y $p + q = 1$.

Si sobre un diagrama cartesiano se llevan como abcisas las desviaciones positivas, y las negativas como ordenadas, los puntos en los cuales la suma de los cuadrados de las ordenadas no supera un cierto valor con la probabilidad menor que 0,05, se encontrarán dentro de un círculo cuyo radio estará determinado por la ley de probabilidad que siga la suma de los cuadrados, pero dada la existencia de la relación (14.5), los puntos representativos de una experiencia posible no se encuentran en un cuarto de circunferencia sino sobre la recta bisectriz del ángulo (porque $p = q$), y la determinación de la región crítica dependerá de una ley de probabilidad unidimensional, que en este caso resulta ser la de Ji cuadrado de orden uno. Si se atendiera sólo al caracter asintóticamente normal (salvo una constante de proporcionalidad) de cada una de las desviaciones relativas, se llegaría en cambio a la conclusión errónea de que habría que utilizar la ley de Ji cuadrado de orden dos.

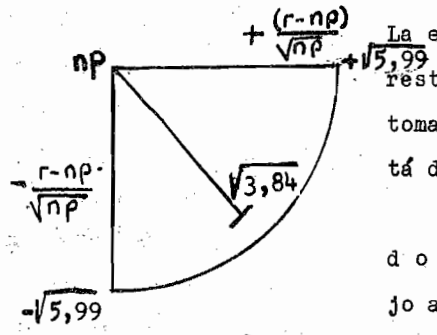


Fig. 3

La existencia de la relación (14.5) se interpreta como una restricción a la libertad de elección de valores que pueden tomar las desviaciones, pues una vez fijada una desviación está determinada la otra.

Consistentemente con esta denominación, se llama g r a d o d e l i b e r t a d de un fenómeno aleatorio complejo al número de clases en que se dividen las observaciones menos uno, y la cuestión tiene importancia, porque es posible demostrar que la generalización de la medida relativa de la desviación entre los valores observados y los esperados, del tipo de la (14.3) sigue una ley de Ji cuadrado de orden igual al de grados de libertad (a).

(a) Para una demostración rigurosa, que es un tanto complicada, ver CRAMER (2) Cap. 30.

Sea, por ejemplo, verificar la hipótesis de que un dado es equilibrado (o sea que cada una de sus caras tiene probabilidad 1/6), utilizando los resultados de 120 tiros. Si f_i es el número de veces que se ha observado el resultado i (el índice i varía de 1 a 6), se forma la diferencia entre f_i y su valor esperado de acuerdo con la hipótesis sis: $np_i = 120 \times 1/6 = 20$, se eleva al cuadrado, se divide por 20 y se suma para todo i. Como el número de grados de libertad es igual al de clases menos uno, en el caso $6 - 1 = 5$, y una vez elegido el nivel de significación 0,05, conforme a la tabla III de CRAMER (3) se aceptará la hipótesis si

$$\chi^2 = \sum_i \frac{(f_i - 20)^2}{20} \leqslant 11{,}070 \tag{14.6}$$

dando por sentado que el poder separador del método es suficiente. Pero si hubiera un interés muy grande en asegurarse que una cierta hipótesis no da lugar a un error de segunda clase con una probabilidad elevada, habría que estudiar la función de potencia antes de considerar que una serie de observaciones de 120 tiros es decisiva.

XIV.8.- APLICACION DEL METODO DEL JI CUADRADO A LA VERIFICACION DE HIPOTESIS INCOMPLETAMENTE DETERMINADAS. Una ventaja importante del método introducido por Pearson es su flexibilidad de aplicación, que permite resolver problemas de verificación aun cuando la ley de probabilidad esté incompletamente determinada.

Sea por ejemplo, verificar la hipótesis de que una serie de datos es una muestra extraída de una ley nórmal. La determinación completa de la ley requeriría fijar los valores de los parámetros m y σ, pero estos son desconocidos. Sin embargo, utilizando las estimaciones dadas por (13.23) para m y por (13.30) para σ^2, es posible dividir el intervalo de variación en segmentos, a cada uno de los cuales y por medio de una tabla de la integral de la ley normal (como la primera columna de la TABLA I de CRAMER (3)) resulta atribuída una probabilidad p_i, después de lo cual puede procederse como en el ejemplo dado, construyendo la expresión

$$\chi^2 = \sum \frac{(r - np_i)^2}{np_i} \quad ; \quad \sum p_i = 1 \tag{14.7}$$

Es necesario tener muy presente que el número de frados de libertad no es en este caso el de clases menos uno, sino que hay que restar tambien el número de parámetros estimados por medio de las observaciones. En efecto, si se representa la serie de observaciones por medio de un punto en el espacio de n dimensiones (ver Cap. XII, párrafo 8) la ley del Ji cuadrado de orden n indica la probabilidad de que el punto caiga fuera de una hiperesfera de radio χ. La limitación en los grados de libertad debido a la división en clases resta, como se señaló anteiormente, una unidad, pero la

circunstancia de que todas las muestras que dan lugar a, por ejemplo, un valor de la estimación de m, estan ubicadas sobre el hiperplano de ecuación $\sum \frac{x_i}{n} = \bar{x}$ restringe a los grados de libertad en una unidad adicional. Similarmente, el estimador de σ^2 determina una hipersuperficie sobre la que deben hallarse todos los puntos representativos de muestras que dan lugar a un mismo valor de s^2, por lo que la imposición de estas tres limitaciones hace que en realidad se trate de una ley de Ji cuadrado de orden n-3 (a). En el caso de tratarse de una ley de Poisson, las restricciones son dos, debido a la clasificación en grupos y a la estimación de un parámetro. En este caso, los grupos estan dados por el número de observaciones en un cierto intervalo de la variable continua, y para que la limitación en los grados de libertad sea efectiva, hay que incluir en un sólo grupo a los intervalos en que las observaciones superan un cierto número k, al que se le asigna la probabilidad $\sum_{j=k}^{\infty} \frac{\bar{\alpha}^j}{j!} e^{-\bar{\alpha}}$, de tal manera que $\sum p_i = 1$.

Aplicando el procedimiento a la discusión de los datos de W.A.WALLIS, sobre las vacantes en la Corte Suprema de los Estados Unidos (ver Cap. VII, párrafo 5), se tiene:

(14.8) $$\chi^2 = \frac{(0,8)^2}{58,2} + \frac{(2,1)^2}{29,1} + \frac{(1,7)^2}{7,3} + \frac{(0,4)^2}{1,4} = 1,033$$

En la tabla III de CRAMER se puede observar que para un número de grados de libertad igual a 2, y un nivel de significación 0,05, el límite inferior de la región crítica es 5,991 de manera que se puede considerar que las diferencias entre las frecuencias observadas y las teóricas no son significativas, y la hipótesis de que las vacantes se producen siguiendo una ley de Poisson con parámetro $\alpha = 0,5$ es aceptable.

El detalle de aplicación de este procedimiento de verificación de hipótesis incompletamente determinadas requiere algunas precauciones adicionales. La agrupación de las observaciones debe realizarse de tal manera que cada clase incluya por lo menos 10 (para asegurar una razonable convergencia a la normal de la ley de las desviaciones), y cuando la variable es continua (caso de la ley normal) la necesidad de utilizar intervalos dados por la aproximación máxima de las mediciones obliga a introducir correcciones especiales (es muy conveniente estudiar detalladamente los ejemplos de CRAMER (2), párrafo 30.4).

XIV.9.- VERIFICACION DE LA HIPOTESIS DE LA HOMOGENEIDAD. A menudo se presenta el caso de que la hipótesis a verificar no es una ley de probabilidad completa o incompletamente determinada, sino que se trata ya sea de asegurar que la ley no ha variado en dos series de observaciones, ya sea que se trate de comprobar que "las demás circunstancias

(a) Una demostración de que la estimación de parámetros restringe los grados de libertad en la forma sugerida por el razonamiento geometrico puede verse en CRAMER (2) Cap. 30. Es de notar que originariamente Pearson dió una demostración falsa de que la estimación de parámetros no influía. La corrección se debe a R.A. FISHER.

"se mantienen constantes" (hipótesis básica del método experimental), o que la variacion de un factor específico tiene influencia.

La disposición de las observaciones en un cuadro de doble entrada (tabla de contingencia), cuando el problema se refiere a un caso de alternativa simple, permite clasificarlas en cuatro grupos. Si las columnas del cuadro indican las dos alternativas, y las filas las distintas oportunidades en que se realizaron las observaciones (o la variacion de otro factor) como se indica simbólicamente:

ν_{11}	ν_{12}	ν''_1
ν_{21}	ν_{22}	ν''_2
ν'_1	ν'_2	

la llamda h i p ó t e s i s n u l a (inexistencia de diferencias) implica suponer que las probabilidades correspondientes a cada una de las cuatro clasificaciones del cuadro resultaría ser el producto de las probabilidades m a r g i n a l e s de filas y columnas (si no hay influencias en el fenómeno, las observaciones agrupadas en el total se deberían comportar de la misma manera que las observaciones clasificadas segun el posible factor actuante).

Las probabilidades de la hipótesis nula se estiman dividiendo las frecuencias marginales por el número total de observaciones

(14.9) $$p_i = \frac{\nu''_i}{n} \quad ; \quad p_j = \frac{\nu'_j}{n}$$

de manera que la expresión del Ji cuadrado sería:

(14.10) $$\chi^2 = \frac{\sum (\nu_{ij} - np_ip_j)^2}{np_ip_j} = \frac{\sum \left(\nu_{ij} - \frac{\nu''_i\,\nu'_j}{n}\right)^2}{\frac{\nu''_i\,\nu'_j}{n}} = \sum_{i,j} \frac{\left(\nu_{ij} - \nu'_j \cdot p_i\right)^2}{\nu'_j \cdot p_i}$$

El número de estimaciones independientes es 2, porque una vez estimada una de las probabilidades marginales, queda también determinada la complementaria y la división en clases resta otro grado de libertad, de manera que habría que utilizar la tabla de χ^2 correspondiente al orden $4 - 2 - 1 = 1$.

El interés de este caso es evidentemente muy grande en todo trabajo de investigación como ya se señaló en la introducción (Cap. 1, parrafo 5) y se generaliza sin dificultad cuando hay r filas y s columnas. Las estimaciones independientes de las probabilidades marginales de las filas son $r-1$, y las de las columnas $s-1$, y el número de clases rs, de manera que se tiene un número de grados de libertad

(14.11) $$\nu = rs - (r + s - 2) - 1 = (r - 1)(s - 1)$$

XIV.10.- DIAGRAMAS DE CONTROL. En los últimos veinte años ha adquirido gran difusión

un método simplificado de verificación de la hipótesis de homogeneidad, de aplicación principalmente en el control de procesos industriales, y tambien en trabajos de laboratorio.

Dado el gran número de factores que influyen en un proceso industrial los resultados del mismo (por ejemplo. dimensiones de una pieza de máquina) son consecuencia de una superposición de efectos que no son totalmente controlables en forma directa. Pero si se admite que estos efectos actúan al azar (siguiendo una ley de probabilidad), el resultado final seguirá también otra ley de probabilidad correspondiente a la suma. Sería posiblemente aventurado afirmar que se puede emplear el teorema central del límite para asignar una ley normal al resultado, pero sí puede pensarse que la ley tendrá primer y segundo momento, y la experiencia indica generalmente una concentración elevada, por lo que puede estudiarse la variación de los promedios de las muestras (grupos de mediciones) que se toman periódicamente, como una medida de la estabilidad de las co condiciones en que se realiza el proceso.

El procedimiento tiene la ventaja de ser expeditivo, y aplicable por personal sin mayor especialización, y consiste en calcular el primer momento de las muestras y llevar su resultado a un gráfico, marcando un punto para cada muestra. Previamente, se ha hecho una estimación del primer momento de la ley, promediando a su vez los promedios de numerosas muestras (o lo que es equivalente, promediando el total de gran número de observaciones) y se han establecido límites para la variación del promedio de la muestra en función de su tamaño, recordando que la varianza del promedio es $\frac{\sigma^2}{n}$ (la varianza se ha estimado también mediante la primer serie prolongada de observaciones), y multiplicando la dispersión estimada (raíz cuadrada de la estimación de la varianza) por 3. Los límites de control para el promedio serían entonces $\bar{\bar{x}} \pm 3 \frac{s}{\sqrt{n}}$ (donde la letra con doble barra indica promedio de promedios, o promedio general). Si todos los puntos representativos de las muestras que integran el grupo de observaciones utilizado para hacer las estimaciones caen dentro de los límites, se dice que el proceso está en condiciones de control (la variación de las mediciones sigue una ley de probabilidad), porque si fuera una ley normal la probabilidad de obtener un punto fuera de los límites cuando el proceso se desarrolla en condiciones de homogeneidad (error de primera clase) sería 0,00067. Si no hay ninguna indicación sobre la forma de la ley, pero se admite que los momentos de las muestras son estimaciones válidas de los momentos de la ley, la desigualdad de TCHEBICHEFF permite asegurar un nivel de significación de 1/9, y si la ley es aproximadamente normal el nivel de significación será siempre muy inferior a esta cota.

En lo sucesivo, se toman periódicamente muestras, y se las representa en el gráfico. Cuando un punto cae fuera de los límites, se investiga la posible causa de varia-

ción, para restablecer el nivel anterior de homogeneidad, o sea, mantener la c a l i d a d (fijada por la especificación de que las variaciones de medición no deben superar ciertos límites) (a).

El mismo procedimiento puede aplicarse en un laboratorio de mediciones, en donde posiblemente el personal está mejor preparado que en una fábrica para utilizar un procedimiento estadístico más satisfactorio pero no resulta recomendable aplicarlo si no hay motivos para pensar que es necesario. El caso típico sería el de controlar la precisión de las mediciones con un cierto aprato o con un procedimiento de análisis. En este caso, el diagrama de control se refiere a las variaciones del s^2, y cuando un punto cae fuera de los límites, se hace una investigación más precisa para constatar si se está manteniendo la precisión deseada.

XIV.11.- El problema más general de la verificación de una hipótesis n o p a r a m é t r í c a (es decir, de la verificación de hipótesis en las que en forma directa o indirecta no intervienen valores de parámetros) ha recibido considerable atención en los últimos años, habiendose logrado desarrollar un importante método práctico para la verificación de hipótesis de homogeneidad por SMIRNOFF y KOLMOGOROFF (b).

(a) Existe una extensísima bibliografía sobre el control de calidad industrial mediante los gráficos descriptos, que a veces es de dudosa solvencia científica. Un texto muy completo, que cubre también el problema de la aceptación de materiales por muestra indicado en capítulos anteriores, es el de E. L. GRANT:"Statistical Quality Control", New York, McGraw-Hill, 1952.
El manual de las ASTM requerido en (31) da normas para la utilización de los gráficos de control en diversos casos, tablas auxiliares, etc. Una aplicación muy interesante a las técnicas de un laboratorio de radioisótopos se describe en R.A.FAIRES y B. H.PARKS: "Radioisótopos", Buenos Aires, Eudeba, 1960.
(b) Ver al texto de LINDGREEN y Mc ELRATH referido en (25), cap. 8.

C A P I T U L O X V

E S T I M A C I O N P O R I N T E R V A L O S Y P R O B L E M A S C O N E X O S.

XV.1.- En el capítulo XIII se explicaron las bases de los dos métodos de estimación puntual más usuales, los que consistían en la selección de estimadores fundandose en las propiedades de su ley de probabilidad.

Tanto en el método de la máxima verosimilitud, de R.A.FISHER, como en el de NEYMAN-MARKOFF, no se determinaba explicitamente la ley de probabilidad del estimador, sino que se construía a este último mediante un procedimiento que aseguraba que su ley de probabilidad satisfacía las condiciones fijadas de consistencia y eficiencia.

Pero como se verá enseguida, hay casos en los cuales es necesario determinar la ley de probabilidad del estimador, es decir, hay que resolver el tercero de los problemas indicados por FISHER en la memoria citada en el Cap. XIII, párrafo 6, o sea el de d i s t r i b u c i ó n.

La deducción formal de dicha ley no presenta dificultades en los casos de estimación de parámetros más usuales, pues apoyandose en las propiedades de la ley que se supone que siguen las observaciones, es posible deducir su forma exacta, como ocurre con el promedio de las muestras extraídas de una ley normal, que sigue otra ley normal con el mismo parámetro m y una varianza $\frac{\sigma^2}{n}$, o bien la tendencia a un límite, como también ocurre en el caso del promedio de una muestra extraída de una ley con momento de segundo orden, porque entonces es válido el teorema de LINDEBERG-LEVY (caso especial del teorema central del límite aplicable cuando las leyes de las variables sumandos son idénticas, demostrado en el Cap. XI, párrafo 3), y el estimador sigue a s i n t ó t i c a m e n t e una ley normal (a).

Pero cuando se trata de pasar del conocimiento de la forma exacta o asintótica al uso de la ley del estimador, surge la grave dificultad que puede ocurrir que dependa de los parámetros que se trata de estimar, que son desconocidos, y que es precisamente lo que sucede con el promedio de las muestras extraídas de una ley normal.

Es evidente, por lo visto, la importancia de hallar estimadores cuya ley de probabilidad no dependa de los parámetros a estimar, y si los estimadores que se desea usar por ser consistentes y eficientes, tienen esa propiedad inconveniente, habrá que construir funciones de los mismos que a su vez tengan una ley de probabilidad independiente de los parámetros.

XV.2.- En el caso de la ley normal, único que se estudiará, el promedio de la muestra,

(a) CRAMER (2) Cap. 27 y 28, estudia el problema en general.

que es el estimador óptimo tanto para el método de FISHER como para de de NEYMAN-MARKKOFF, tiene precisamente el inconveniente señalado, y por lo tanto, toda vez que haya de hacerse un estudio más detallado que el que permite la determinación de un solo punto como estimación óptima, deberá utilizarse una función del mismo, cuya construcción se verá más adelante.

El estimador de la varianza posee la importante propiedad de que la ley de probabilidad de una función sencilla del mismo, es independiente de los parámetros, lo que se demuestra haciendo uso de las propiedades de las funciones características estudiadas en el Cap. X.

En efecto, si la suma de cuadrados de las diferencias entre cada una de las observaciones x_i y m se descompone sumando y restando $\bar{x}$, se tiene

$$(15.1)\qquad \sum(x_i - m)^2 = \sum[(x_i - \bar{x}) + (\bar{x} - m)]^2 = \sum(x_i - \bar{x})^2 + n(\bar{x} - m)^2 - 2(\bar{x} - m)\sum(x_i - \bar{x}) = (n-1)s^2 + n(\bar{x} - m)^2$$

porque al resolver la suma que figura en el doble producto el término se anula, ya que $\sum x_i = n\bar{x}$ y el primer sumando es el numerador de (13.29).

Dividiendo ambos miembros por σ^2:

$$(15.2)\qquad \frac{\sum(x_i - m)^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} + \frac{n(x - m)^2}{\sigma^2}$$

El primer miembro es una suma de cuadrados de variables normales reducidas, y sigue una ley de Ji cuadrado de orden n (ver Cap. X, párrafo 4).

En el segundo miembro, el primer sumando es una función del estimador s^2 y σ^2, cuya ley no se conoce, y el segundo es nuevamente el cuadrado de una ley normal reducida (recordando que el promedio x sigue una ley normal de parámetros m y $\frac{\sigma^2}{n}$), que sigue una ley de Ji cuadrado de orden 1.

Si se recuerda la propiedad de las funciones características de la suma de variables aleatorias independientes (para lo cual habría que probar previamente la independencia de los sumandos del segundo miembro, lo que no ofrece dificultades), se tiene:

$$(15.3)\qquad \varphi(u) = \varphi_1(u), \varphi_2(u)$$

Tanto $\varphi(u)$ como $\varphi_2(u)$ son conocidas (ver Cap. X, párrafo 12):

$$(15.4)\qquad \varphi(u) = \frac{1}{(1 - 2iu)^{\frac{n}{2}}} \quad ; \quad \varphi_2(u) = \frac{1}{(1 - 2iu)^{\frac{1}{2}}}$$

y por lo tanto $\varphi_1(u)$ será igual a su cociente:

$$(15.5)\qquad \frac{\varphi(u)}{\varphi_2(u)} = \frac{1}{(1 - 2iu)^{\frac{n-1}{2}}}$$

Pero este cociente resulta ser la función característica de una ley de Ji cuadrado de orden $(n-1)$, de manera que la función de densidad de probabilidad de la función de s^2

y σ^2 es finalmente:

$$f\left(\frac{(n-1)s^2}{\sigma^2}\right) = \frac{\left(\frac{1}{2}\right)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{(n-1)s^2}{\sigma^2}\right)^{\frac{n-3}{2}} \cdot e^{-\frac{(n-1)s^2}{2\sigma^2}} \tag{15.6}$$

La circunstancia de que el denominador del estimador insesgado sea igual al orden del Ji cuadrado no es casual. R.A.FISHER probó que cuando existe una relación lineal entre las variables de una suma de cuadrados (caso del promedio), se reduce en uno el número de cuadrados independientes que se suman (ver CRAMER (2), Cap 27), reducción que tambien queda en evidencia al calcular la esperanza matemática (ver Cap. XIII, párrafo 11). De este hecho se deduciran importantes aplicaciones en lo sucesivo.

XV.3.- Es conveniente interpretar geométricamente el procedimiento utilizado. La suma del primer miembro se representa como un punto del espacio n-dimensional, a una distancia del origen igual a $\sqrt{\frac{\sum(x_i-m)^2}{\sigma^2}}$, de manera que la ley de Ji cuadrado correspondiente permite medir la probabilidad de que el punto se halle fuera de una esfera de radio dado si considera la función de distribución, y la función de densidad de probabilidad corresponde a los puntos que se hallan sobre la superficie de la hiperesfera (ver Cap. XII, párrafo 8) de radio $r^2 = \frac{\sum(x_i-m)^2}{\sigma^2}$.

Pero si el promedio de la muestra es igual a $\bar{x}$, todas las muestras con el mismo promedio se hallaran sobre un hiperplano de ecuación $\frac{\sum x_i}{n} = \bar{x}$, de manera que en realidad el lugar geométrico de los puntos representativos de muestras con un $\bar{x}$ y un s dados, será una hiperesfera en el espacio (n-1) dimensional (si $n = 3$, esta segunda hiperesfera se reduce a un círculo), y el centro de esta última hiperesfera será el punto de coordenadas $x_i = \bar{x}$.

Es claro que la función de densidad de probabilidad del conjunto de puntos representativos de muestras con un mismo $\bar{x}$ y s^2 es la de la superficie de la hiperesfera de centro $x_i = \bar{x}$, pero lo que no es inmediato es que sea también una ley de Ji cuadrado de orden (n-1), lo que se prueba hallando su función característica, como se hizo.

XV.4.- ESTIMACION DE LA VARIANZA. Como la ley de Ji cuadrado no depende de ningun parámetro, se podrá formar inmediatamente las siguientes expresiones:

$$P\left(\chi^2_{p_1} \geqslant \frac{(n-1)s^2}{\sigma^2} \geqslant \chi^2_{p_2}\right) = 1 - (p_1 + p_2) = 1 - p \tag{15.7}$$

$$P\left(\frac{(n-1)s^2}{\chi^2_{p_1}} \leqslant \sigma^2 \leqslant \frac{(n-1)s^2}{\chi^2_{p_2}}\right) = 1 - p \tag{15.8}$$

donde $\chi^2_{p_1}$ es el valor que el Ji cuadrado de orden (n-1) supera con probabilidad $1-p_1$,

el que se obtiene de la tabla correspondiente (tabla III de CRAMER (3)).

La varianza es una constante caracteristica de la ley normal en estudio, la que resulta recubierta por un segmento aleatorio (porque sus extremos son variables aleatorias) con probabilidad 1-p, de manera que mediante el artificio de construir una función del estimador con ley de probabilidad independiente de los parámetros (en este caso la ley independiente es la de Ji cuadrado de orden (n-1)) se ha logrado completar la estimación puntual estableciendo la precisión de la misma, o sea un intervalo aleatorio con una cierta probabilidad de recurrir el verdadero valor (a).

XV.4.- La probabilidad p desempeña el papel de un nivel de significación, que en este caso mide la probabilidad de cometer el error de que el segmento no recubra el verdadero valor. Como se señaló al discutir el nivel de significación en el Cap. XIV, párrafo 3, dado un nivel existen infinitos intervalos que se pueden construir con el mismo, y la selección del intervalo más adecuado no depende en este caso de un error de segunda clase que no puede definirse, sino de la necesidad de darle la longitud más pequeña posible.

Como la ley de Ji cuadrado, especialmente para pequeños valores de n es muy asimétrica, la determinación de un intervalo de longitud mínima no es problema sencillo, pero en general se obtiene un intervalo de longitud cercana al óptima adoptando $p_1 = 0$ o sea $x^2_{p_1} = +\infty$ con lo que se tiene con probabilidad 1-p, una cota superior de la varianza, o también se puede elegir $p_1 = p_2 = \frac{1}{2}p$.

XV.5.- LA DISTRIBUCION t DE "STUDENT". En el caso del promedio no es tan sencillo como en el de la varianza encontrar una función simple del estimador, cuya ley de probabilidad resulte independiente de los parámetros. En realidad, hay que combinar la variable aleatoria $\bar{x}$, previamente centrada, con otra variable aleatoria, que es el estimador de su varianza s^2, obteniendose por cociente (15.10).

En lo sucesivo se llamará v a la variable aleatoria centrada $(\bar{x} - m)$.

(a) Esta método de estimación por intervalos se debe a NEYMAN, y se publicó por primera vez en forma bibliograficamente accesible en la memoria citada en el Cap. XII, párrafo 3, que conjuntamente con la de FISHER referida en Cap. XIII, párrafo 2, y las de NEYMAN y PEARSON en Cap. XIV, párrafo 3, constituyen los orígenes de la teoría moderna de la inferencia estadística.

Una presentación detallada del método de estimación por intervalos se encuentra en el siguiente artículo de J.NEYMAN: "Outline of a theory of satistical estimation based on the clasical theory of probability", Philosophical Transactions of the Royal Society (London) vol.236 (1937), pag. 333 ya citada en el apéndice al Cap. III (párrafo III. A. 2). Este artículo se complementa utilmente con la respuesta de NEYMAN a ciertas objeciones de R.A.FISHER: "Fiducial argument and the theory of confidence intervals", Biometrika, vol.32 (1941), pag. 128.

Para hallar la ley de probabilidad que sigue la variable aleatoria "t", hay que proceder como en el Cap. XII, párrafo 7 para encontrar la ley del cociente de dos variables normales, es decir, hay que formar la distribución conjunta de s y de v y luego, mediante un cambio de variables conveniente, integrar sobre todos los puntos que corresponden a pares de valores de v y de s que hacen constante a t.

La función de densidad de probabilidad de v, expresada en forma diferencial, es:

$$f(v)dv = \frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma} e^{-\frac{nv^2}{2\sigma^2}} dv \tag{15.10}$$

o sea una ley normal centrada con varianza $\frac{\sigma^2}{n}$.

La función de densidad de probabilidad de s hay que deducirla de la de s^2, ya encontrada en el párrafo XV.4, y que es, tambien en forma diferencial:

$$f\left(\frac{(n-1)s^2}{\sigma^2}\right)d\left(\frac{(n-1)s^2}{\sigma^2}\right) = \frac{(\frac{1}{2})^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{(n-1)s^2}{\sigma^2}\right)^{\frac{n-3}{2}} \cdot e^{-\frac{(n-1)s^2}{2\sigma^2}} d\left(\frac{(n-1)s^2}{\sigma^2}\right) \tag{15.11}$$

Resolviendo el diferencial y el paréntesis:

$$f(s)ds = \frac{2(\frac{1}{2})^{\frac{n-1}{2}}(n-1)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)\sigma^{n-1}} s^{n-2} e^{-\frac{(n-1)s^2}{2\sigma^2}} ds \tag{15.12}$$

con lo que se tiene la función de densidad de probabilidad de s.

Si las variables aleatorias v y s son independientes, la función de densidad de probabilidad de su distribución conjunta será el producto de sus funciones de densidad de probabilidad (ver Cap. XII, párrafo 3):

$$f(s,v)dsdv = \frac{K}{\sigma^n} s^{n-2} e^{-\frac{(nv^2+(n-1))s^2}{2\sigma^2}} dsdv \tag{15.13}$$

habiéndose agrupado los términos constantes que no intervendran en lo sucesivo en K, que resulta igual a

$$K = \frac{2(\frac{1}{2})^{\frac{n-1}{2}}(n-1)^{\frac{n-1}{2}}\sqrt{n}}{\Gamma\left(\frac{n-1}{2}\right)\sqrt{2\pi}} \tag{15.14}$$

Los valores de s y v que hacen constante a t satisfacen la relación $\sqrt{n}\,v = ts$. Si se representan graficamente (como en la figura 3, del Cap. XII, párrafo 7), llevando a s sobre el eje de abcisas (únicamente sobre la parte positiva, porque no tienen sentido las dispersiones negativas), y a v sobre el eje de ordenadas (de $-\infty$ a $+\infty$), los puntos con t constante estaran representadas en la semirrecta de abcisas positivas que pasa por el origen y tiene una pendiente $tg\alpha = t$.

Para hallar la función de densidad de probabilidad de t, hay que integrar sobre todos los puntos de la semirrecta correspondiente, para lo cual habría que pasar a coorde

nadas polares, o bien efectuar el cambio de variables aún más sencillo empleado en el Cap. XII, párrafo 7, que en este caso consiste en reemplazar v por t dejando a s constante, e integrar luego sobre todo el campo de existencia de s (o sea de 0 a $+\infty$). En este último cambio de variables se tiene $dsdv = \frac{sdsdt}{\sqrt{n}}$, por lo que (15.13) se transforma en:

$$f(s,t)dsdt = \frac{K}{\sqrt{n}\,\sigma^n}\, s^{n-1}\, e^{-\frac{[t^2+(n-1)]s^2}{2\sigma^2}}\, ds\,dt \tag{15.15}$$

Integrando con respecto a s:

$$f(t)dt = \frac{K\,dt}{\sqrt{n}\,\sigma^n}\int_0^{+\infty} s^{n-1}\, e^{-\frac{(t^2+(n-1))s^2}{2\sigma^2}}\, ds\,dt \tag{15.16}$$

Si en la integral definida se efectúa el cambio de variables $s^2 = w$, $2sds = dw$, se tiene:

$$f(t)dt = \frac{K\,dt}{2\sqrt{n}\,\sigma^n}\int_0^{+\infty} w^{\frac{n-2}{2}}\, e^{-\frac{(t^2+(n-1))w}{2\sigma^2}}\, dw \tag{15.17}$$

con lo que la integral se ha transformado en una gamma generalizada de orden $n/2$ y parámetro igual al factor de w en la exponencial. Estas integrales se calculan facilmente (ver Cap. X, párrafo 9), siendo iguales al cociente de la gamma del mismo orden por el parámetro elevado a una potencia igual a dicho orden, con lo que se tiene una expresión en la que σ^n que figura en el denominador de (15.17), se simplifica:

$$f(t)dt = \frac{K\,dt}{2\sqrt{n}\,\sigma^n}\;\frac{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}\,\sigma^n}{\left[t^2+(n-1)\right]^{\frac{n}{2}}} = \frac{K\,\Gamma\left(\frac{n}{2}\right)2^{\frac{n-1}{2}}}{\sqrt{n}\left[t^2+(n-1)\right]^{\frac{n}{2}}}\,dt \tag{15.18}$$

Reemplazando K por su expresión (15.14), simplificando, y sacando (n-1) factor común del paréntesis, se tiene finalmente la función de densidad de probabilidad buscada, que es i n d e p e n d i e n t e d e l o s p a r á m e t r o s de la ley normal(a):

$$f(t)dt = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)}\cdot\frac{1}{\sqrt{n-1}}\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}}\,dt \tag{15.19}$$

XV.6.- Dada la importancia que tiene en la práctica, esta función de densidad de probabilidad ha sido tabulada (ver tabla IV de CRAMER (3)). El número de grados de libertad es el de la ley de Ji cuadrado, que sigue el denominador.

(a) El importante descubrimiento de la ley de probabilidad de t se debe a "Student" (seudónimo con que efectuaba sus publicaciones W.S.GOSSETT, encargado del laboratorio de una cervecería de Dublin y uno de lo más notables especialistas es estadística del siglo XX):" The probable error of a mean", Biometrika, vol.6 (1908),pag. 1. R.A.FISHER revisó posteriormente la deducción de Student, probando rigurosamente la independencia de v y s :"Applications of Student's distribution" Metron, vol. 5 (1925), pag. 90.

Cuando $\nu = n-1 = 1$, o sea $n = 2$, reemplazando en (15.19) se tiene

(15.20) $$f(t)dt = \frac{1}{\pi} \frac{1}{1+t^2} dt$$

que es la ley de Cauchy, y por lo tanto es simétrica y no tiene ningun momento.

Para valores de $\nu > 1$, la ley continúa siendo simétrica y posee momentos, verificándose que en el límite para $\nu \to +\infty$ (15.19) se transforma en la normal reducida, siendo la aproximación bastante aceptable para $\nu \geqslant 30$, y valores grandes del desvío, lo que puede comprobarse comparando las tablas II y IV de CRAMER (3), para una probabilidad 0,05. A la ley normal le corresponde 1,96 y al "t" de Student 2,04 o sea una diferencia poco superior al 4%.

Hay que tener presente, antes de utilizarla, cual es la expresión del s^2 con que se trabaja. En la deducción se ha seguido con la convención adoptada en el capítulo XIII de denominar s^2 al estimador i n s e s g a d o de la varianza, (ver Cap. XIII, párrafo 11) pero si, como por ejemplo hace CRAMER, se entiende por s^2 al estimador de máxima verosimilitud, la expresión de t cambia:

(15.21) $$t = \sqrt{n-1}\,\frac{y}{s}$$

Se ve inmediatamente que (15.20) y (15.9) son numéricamente iguales, dando en cada caso la correspondiente interpretación a s.

En cuando al cáculo numérico, es posible utilizar los procedimientos recomendados en el Cap. XIII, párrafo 10, sin necesidad de ninguna corrección al final, porque tanto numerador como denominador resultan afectados en igual forma por la constante multiplicativa, y por lo tanto el t calculado con los números modificados es el mismo que el se calcularía con los números tal como se obtienen de las observaciones.

XV.7.- APLICACION A LA ESTIMACION DE m. Procediendo en la misma forma que en el caso de la varianza, se tiene que

(15.22) $$P(-t_p \leqslant \sqrt{n}\,\frac{\bar{x} - m}{s} \leqslant +t_p) = 1 - p$$

de donde se deduce

(15.23) $$P(\bar{x} - t_p \frac{s}{\sqrt{n}} \leqslant m \leqslant \bar{x} + t_p \frac{s}{\sqrt{n}}) = 1 - p$$

La interpretación de (15.23) es que el segmento aleatorio tiene una probabilidad $1-p$ de recubrir el verdadero valor, y por consiguiente una probabilidad p de no recubrirlo. Si el procedimiento se emplea sistemáticamente, en un porcentaje p de veces, no recubrirá el verdadero valor.

Como la ley del t de Student es simétrica, una vez fijado p el segmento de menor longitud al que corresponde una probabilidad $1-p$ es el simétrico son respecto a $\bar{x}$, por lo que en este caso el método de estimación por intervalos es de aplicación más

simple que para la varianza.

XV.8.- Si en la expresión (15.9) del t, se confunde la estimación s con el verdadero valor σ, se tiene evidentemente la ley normal. Este es el método de la teoría clásica de los errores, al que se agregaba el concepto de error probable, definido por el intervalo 0,6745. A este intervalo le corresponde una probabilidad $p=\frac{1}{2}$, de manera que si se admite que los errores accidentales se distribuyen normalmente en torno al verdadero valor, a los valores que quedaban fuera del intervalo les resulta atribuída una probabilidad "a posteriori" menor que 1/2, porlo que se los descartaba, mientras que al centro del intervalo, le correspondía la máxima probabilidad "a posteriori".

Pero como observa NEYMAN (en la memoria citada en III.A.2 y en XV.3) resulta dificilmente justificable hablar de probabilidades "a priori" y "a posteriori" y emplear el teorema de Bayes, cuando se trata de mediciones reiteradas sobre un mismo objeto, porque el verdadero valor de la medida que es el m de la ley normal, es una constante y no una variable aleatoria. Los que son variables aleatorias son los extremos del intervalo confidencial y dentro de el no hay ningun valor mas conveniente que otro. Además en una interpretación de frecuencias relativas, es razonable disminuir la probabilidad de error (que tiene el sentido de que el segmento no recubra el verdadero valor), por lo que conviene disminuir p, adoptandose en la práctica los niveles 0,05 y 0,01, en vez de $p=\frac{1}{2}$, que daría errores demasiado frecuentes.

El estimador s^2 converge en probabilidad al verdadero valor σ^2, pero si el número de observaciones es pequeño, no sólo sus fluctuaciones no son despreciables, sino que la ley que sigue t es la de Student y no la normal. Para un nivel $p=0,05$ y 3 observaciones (dos grados de libertad), el valor de t, segun la tabla IV de CRAMER (3) es 4,303, considerablemente mayor que el 1,960 de la ley normal, de manera que la confusión entre la estimación y el verdadero valor conduce a errores muy importantes. Como 1,96 supera a 1,886 que en la tabla de t corresponde a $p=0,20$ y es bastante menor que 2,920, que corresponde a $p=0,10$, el uso equivocado de ley normal conduciría en realidad a una frecuencia relativa de error poco inferior al 20%, mientras que el nivel de significación aparente es de 5%.

Por esta causa, ya desde 1937 la American Society for Testing Materials, cuyas normas son amplimente reconocidas, recomienda el uso del t de Student y la presentación de resultados de mediciones en forma de intervalos confidenciales ("ASTM Manual on Presentaction of Data", cuya actualizacion es la primera parte del manual referido en (31)), dando una tabla muy práctica (pag. 43) para los niveles $p_1=0,10$, $p_2=0,05$ y $p_3=0,01$ en la que se debe entrar por el número de observaciones (y no por el de grados de libertad), y en la que se $a_p=\frac{t_p}{\sqrt{n-1}}$, de manera que (15.23) se simplifica:

(15.24) $$P(x - a_p s \leqslant m \leqslant \bar{x} + a_p s) = 1 - p$$

disminuyendo el número de operaciones numéricas a realizar (en este caso $s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$).

Sin embargo, en una teoría general de la estimación, que se ocupe de otros problemas que los en realidad muy restringidos que présenta la teoría de los errores, y en los que podría presentarse validamente una ley de probabilidades "a priori" queda abierta la cuestión de si la aplicación del método de intervalos confidenciales no puede conducir a resultados contradictorios con las probabilidades "a posteriori". Segun VON MISES (a), no habría posibilidad de inconsistencias.

XV.9.- APLICACION DEL t DE STUDENT A PROBLEMAS DE REGRESION. En el Cap. XIII, párrafo 13 se consideró el problema de una dependencia aproximadamente lineal entre dos variables (X_i, Y_i), de las cuales se consideraba que X_i se podía determinar con precisión (por ejemplo, es un tiempo), y que Y_i era el resultado de la superposición de una función lineal de las X_i y de una variable aleatoria normal, centrada en la ordenada de la recta con una varianza constante e independiente de X_i. Este esquema puede representar un simple problema de teoría de errores, si las variables normales son los errores accidentales de medición, o bien referirse al cuadro más general de un fenómeno aleatorio normal cuyo primer parámetro se desplaza siguiendo una función lineal, como se supone que ocurre en algunos fenómenos económicos y de otro orden.

De los dos parámetros que individualizan la recta, tiene especial interés el correspondiente a la pendiente, o c o e f i c i e n t e d e r e g r e s i ó n, siguiendo la terminología del esquema aún más general descripto en el Cap. XII, párrafo 6 y R.A.FISHER descubrió que la ley de probabilidad del estimador correspondiente

(15.25) $$\bar{\beta} = b = \frac{\sum x_i y_i}{\sum x_i^2}$$

deducida ya en el Cap. XIII párrafo 14 es normal, y podrá aplicarse el t de Student, con un número de grados de libertad igual al de observaciones menos dos.

XV.10.- Para demostrar esta importante aplicación del t de Student basta simplemente desarrollar la expresión del estimador (15.26).

El promedio y de las observaciones Y_i, que se útiliza para transformarlas en las observaciones centradas y_i del numerador de (15.25), es, conforme al planteo del problema:

(15.26) $$\bar{y} = \frac{\sum Y_i}{n} = \frac{\sum (\alpha + \beta x_i + e_i)}{n} = \alpha + \bar{e}$$

(a) "On the foundations of Probability and Statistics", Annals of Mathematical Statistics, Vol. 12 (1941), pag.191.

en donde ya las x_i han sido centradas por su promedio y e es la variable normal superpuesta a la función lineal. Al resolver la suma, el término en β se elimina porque resulta factor común de una suma de desviaciones con respecto a un promedio, la que es nula, y aparece el promedio de los valores de las variables normales e.

Las y_i centradas serán

(15.28) $$y_i = Y_i - \bar{y} = \alpha + \beta x_i - e_i - (\alpha + \bar{e}) = \beta x_i + e_i - \bar{e}$$

de donde se deduce inmediatamente

(15.29) $$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i(\beta x_i + e_i - \bar{e})}{\sum x_i^2} = \beta + \frac{\sum x_i e_i}{\sum x_i^2}$$

porque el número e es factor comun de una suma que es nula. Si se pasa β al primer miembro, en el segundo queda sólo una combinación lineal de variables aleatorias normales independientes y centradas (porque los coeficientes son números y no variables aleatorias), que es tambien normal y centrada, con lo que se ha probado que las desviaciones entre los valores del estimador y el verdadero valor siguen una ley normal centrada, conviniendo en pasar al primer miembro $\sum x_i^2$:

(15.29) $$(b - \beta)\sqrt{\sum x_i^2} = \frac{\sum x_i e_i}{\sqrt{\sum x_i^2}}$$

de modo que la varianza de esta ley normal es la misma que la de las e_i:

(15.30) $$E\left(\frac{\sum x_i e_i}{\sqrt{\sum x_i^2}}\right)^2 = \frac{\sum x_i^2 \sigma^2}{\sum x_i^2} = \sigma^2$$

Para obtener un estimador de la varianza, hay que calcular la suma de los cuadrados de las diferencias entre los valores de Y_i y las ordenadas de la recta en cuya ecuación los parámetros han sido reemplazados por las estimaciones (13.47), o sea proceder com en (13.43) y desarrollar el cuadrado:

(15.31) $$\sum (Y_i - bx_i)^2 = \sum y_i^2 + b^2 \sum x_i^2 - 2b \sum x_i y_i = \sum y_i^2 - b^2 \sum x_i^2$$

porque $\sum x_i y_i = b \sum x_i^2$.

Para obtener un estimador insesgado de la varianza, hay que calcular la esperanza matemática de la suma de cuadrados, la que después de algunas transformaciones resulta ser igual a (n-2), con lo que la expresión final será:

(15.32) $$s^2 = \frac{\sum y_i^2 - b^2 \sum x_i^2}{n - 2}$$

y como precisamente el número de grados de libertad de t está dado por el denominador, la expresión a emplear será, finalmente

(15.33) $$t = \frac{(b - \beta)\sqrt{\sum x_i^2}}{s}$$

con $\nu = n - 2$, mediante la cual podran deducirse intervalos confidenciales para estimar β:

(15.34) $$b - t_p \frac{s}{\sqrt{\sum x_i^2}} \leq \beta \leq b + t_p \frac{s}{\sqrt{\sum x_i^2}}$$

Para el cálculo numérico, conviene seguir la disposición indicada en el Cap. XIII, párrafo 14, a la que habrá que agregar una columna con los cuadrados de las Y_i, para determinar $\sum y_i^2$ que figura en s^2. De esta manera, puede utilizarse tambien un control de las operaciones, ya que se han calculado todos los términos de $(x_i + y_i)^2$, el que puede agregarse como cuarta columna que debe ser igual a la suma de los tres primeras.

XV.11.- APLICACION DEL t DE STUDENT A LA VERIFICACION DE HIPOTESIS SOBRE m. Es evidente que la teoría de estimación por intervalos está muy ligada a la teoría de verificación de hipótesis. En el caso corriente, en que mediante la realización de un conjunto reducido de mediciones se trata de verificar si una cierta magnitud (longitud, peso, composición química, etc.) es la que establece una cierta especificación, y se admite que los errores accidentales se distribuyen normalmente en torno al verdadero valor, basta formar la expresión del t, determinando p la región crítica para un nivel de significación que mide la probabilidad del error de primera clase (rechazar la hipótesis cuando es cierta, segun lo definió en el Cap. XIV, párrafo 3), la que corresponde a los valores de t mayores en valor absoluto que t_p:

(15.35) $$\left| \sqrt{n} \frac{\bar{x} - m}{s} \right| > t_p$$

Introduciendo el m de la especificación, y el $\bar{x}$ de las observaciones, se rechazará la hipótesis cuando se verifica la desigualdad anterior.

Pero a diferencia con lo que ocurre en el caso de la estimación por intervalos, en la verificación de hipótesis existe el error de segunda clase (probabilidad de aceptar la hipótesis cuando es falsa), y se han construído gráficos de uso sencillo, que permiten establecer el número de observaciones necesario para detectarlo con una probabilidad dada, en función de dos niveles de significación ($p = 0{,}05$ y $0{,}01$) y de la diferencia entre m y otro m', como el gráfico 10 de la colección (27).

XV.12.- SIGNIFICACION DELA DIFERENCIA DE DOS PROMEDIOS. A menudo ocurre que no se trata de verificar si el resultado de un conjunto de mediciones es compatible con la hipótesis de que m tiene un cierto valor, sino si es compatible con el m determinado mediante otro conjunto de observaciones, el que ya no sería un número exacto, sino un intervalo. Este problema, en realidad, es un ensayo de homogeneidad que podría enunciarse como la verificación de la h i p ó t e s i s n u l a (no hay diferen -

cias entre los primeros parámetros de las dos leyes normales que en cada caso se supone que siguen las observaciones). En el caso de un laboratorio químico, consistiría en determinar si dos muestras de un cierto producto presentan la misma proporción de un cierto componente, y se trata de distinguir si una diferencia entre los promedios se debe simplemente al hecho de que el promedio de las observaciones es una variable aleatoria (diferencia no significativa), o si existe una diferencia de composición (diferencia significativa).

En el caso en que se puede admitir que el σ^2 de las dos leyes normales es igual el problema se reduce a una aplicación simple del t de Student (mediciones realizadas con el mismo procedimiento y en las mismas condiciones).

Aunque no se conozca exactamente el valor de m, la hipótesis nula permite a afirmar que es el mismo, y hay que formar el t correspondiente a la diferencia de las dos leyes normales, la que tendrá un $m' = 0$ y una varianza igual a la suma de las varianzas (la diferencia de dos leyes normales es un caso especial de la suma, tratada en el Cap. X, párrafo 11) en que los m se restan pero la elevación al cuadrado mantiene la suma de las varianzas, como resulta de (10.38)).

Por lo tanto, la expresión homóloga de (15.10) tendrá en el numerador la diferencia $(x_1 - x_2)$, porque por hipótesis $m_1 - m_2 = 0$. El estimador de la varianza será un poco mas complicado. Si en el primer caso se han realizado n_1 observaciones, y n_2 en el segundo, la varianza de la diferencia de los promedios será igual a la suma de las varianzas, que en la hipótesis de igualdad de estas últimas es:

$$\left(\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right) = \left(\frac{n_1 + n_2}{n_1 n_2}\right)\sigma^2 \tag{15.36}$$

Para construir un estimador insesgado de σ^2, conviene utilizar todas las observaciones, con lo que se disminuye la varianza y se tiene una estimación más precisa. Si S_1^2 es la suma de los cuadrados de las diferencias entre las observaciones y su promedio en la primera serie, y S_2^2 es la suma de los cuadrados homólogos de la segunda, aplicando el mismo procedimiento que en el Cap. XIII, párrafo 11, se tiene:

$$E(S_1^2 + S_2^2) = (n_1 - 1)\sigma^2 + (n_2 - 1)\sigma^2 = (n_1 + n_2 - 2)\sigma^2 \tag{15.37}$$

Teniendo en cuenta (15.36), la expresión final será:

$$t = \sqrt{\frac{n_1 n_2}{n_1 + n_2}(n_1 + n_2 - 2)}\,\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_1^2 + S_2^2}} \tag{15.38}$$

la que corresponden $(n_1 + n_2 - 2)$ grados de libertad, lo que se demuestra mediante el teorema de FISHER citado en (XV.8), o bien directamente por las propiedades aditivas del ji cuadrado que siguen las sumas de cuadrados (Cap. X, párrafo 12 y Cap. XV, párrafo

2). Si el valor absoluto de (15.38) supera el t_p fijado mediante el nivel de significación y los grados de libertad, se rechazará la hipótesis nula, en cuyo caso se dirá que la diferencia entre los promedios es significativa. En caso contrario, se acepta la hipótesis nula, y la diferencia no es significativa.

XV.13.- Como ejemplo de aplicación de (15.38) es interesante discutir la serie de análisis realizados por Lord Rayleigh a fines del siglo pasado.

Determinando la densidad del nitrógeno proveniente de compuestos químicos, obtuvo la siguiente serie de resultados:

Fecha	Compuesto de origen	Número de determinaciones	$\bar{x}$
Nov.Dic.1893	Oxido Nítrico	4	2,30007
Dic. 1893	Oxido Nitroso	2	2,29904
Enero 1894	Nitrito de Amonio	2	2,29943

En todos los casos se utilizó como agente reductor el hierro y las dos primeras series de mediciones eran ostensiblemente compatibles, superponiéndose los intervalos de variación entre las mediciones extremas de cada serie, de manera que no había motivo aparente para no adoptar el promedio general $m_1 = 2,29956$ como una determinación mas precisa.

Pero con anterioridad (agosto y setiembre de 1892), había realizado otra determinación de la densidad del nitrógeno partiendo del aire y utilizando como agente reductor el cobre, la que dió un resultado que aparecía como mayor $m_2 = 2,31025$, y la serie de cuatro mediciones entonces realizada no se superponía con el conjunto de las mediciones posteriores.

Ante esta dificultad, Lord Rayleigh, hizo una nueva determinación a partir del aire en diciembre de 1893, pero usando esta vez el hierro como agente reductor, la que dió un valor $m_3 = 2,31004$, y en enero de 1894 la última a partir de un compuesto químico, la que era compatible con las dos anteriores. Como la última determinación a partir del aire era superior al promedio de las tres determinaciones químicas $m_0 = 2,29947$. Se consideró significativa la diferencia, y se iniciaron entonces las investigaciones que condujeron al descubrimiento de los gases raros.

Modernamente, un analista que utilizara procedimientos estadísticos, interpretaría los resultados mediante el t , en el supuesto que admitiera que las determinaciones a partir de compuestos químicos y la última a partir del aire tenían el mismo error experimental.

Multiplicando por 100.000 los valores de las observaciones para hacer desaparecer

las comas, y calculando las sumas de cuadrados de las diferencias con los promedios, se tiene:

(15.39) $$S_o^2 = 133.152 \quad ; \quad S_2^2 = 538$$

y el t sería:

(15.40) $$t = \sqrt{\frac{32}{8+4}(8+4-2)}\ \frac{1,057}{\sqrt{133,690}} = 14,9$$

Utilizando la tabla IV de CRAMER (3) para 10 grados de libertad, se puede comprobar que el valor excedido con $p = 0,001$ es 4,587, muy inferior al obtenido, y la diferencia entre los promedios es altamente significativa.

XV.14.- VERIFICACION DE HIPOTESIS SOBRE EL COEFICIENTE DE REGRESION. ANALISIS DE VARIANZA. En el espquema descripto en (XV.9), si el coeficiente de regresión es nulo, se puede admitir que las leyes normales no dependen de un parámetro (por ejemplo el tiempo), lo que puede constituir una conclusión de la mayor importancia en numerosos campos de investigación.

Una manera de verificar la hipótesis nula en este caso sería calcular el intervalo de confianza del coeficiente de regresión, mediante (15.35), aceptandola si incluye el cero, y rechazandola si no lo incluye.

Pero tanto para el coeficiente de regresión como para el coeficiente de correlación (en este último caso interesan además los valores +1 y -1 como índices de dependencia), si el número de observaciones es reducido la ley de probabilidad de los estimadores es muy abierta, y es posible llegar facilmente a conclusiones falsas (aceptando la hipótesis de independencia o de dependencia cuando son falsas).

La discusión desde el punto de vista de la función de potencia es complicada, y por otra parte conviene encontrar procedimientos que sean también aplicables a casos más complejos (como podría ser, en el ejemplo del nitrógeno que se acaba de discutir, verificar la homogeneidad de las tres determinaciones a partir de compuestos químicos). En este orden de ideas, es de fundamental importancia el método del a n á l i s i s d e v a r i a n z a inventado por R.A.FISHER, y que consiste en comparar entre sí diferentes estimaciones de la varianza, realizadas de tal manera que, en caso de existir, denuncian los efectos que se investigan.

El método depende de la circunstancia de que el cociente de dos estimaciones de una una misma varianza:

(15.41) $$F = \frac{S_1^2}{S_2^2}\ ,$$

es, recordando el resultado obtenido en (XV.2), igual al cociente de dos variables aleatorias Ji cuadrado, de orden igual al de observaciones menos uno que intervienen en cada

estimación, y divididas cada una por ese mismo orden:

$$F = \frac{\frac{\chi^2_{\nu_1}}{\nu_1}}{\frac{\chi^2_{\nu_2}}{\nu_2}} \tag{15.42}$$

La ley de probabilidad de este cociente, se calcula de la misma manera que en el caso de la ley de Cauchy y del t de Student, formando la función de distribución conjunta e integrando a lo largo de la recta sobre la cual F=cte., y resulta evidentemente independiente de los parámetros de la ley normal, por lo que es una ley exacta sucepti ble de ser empleada para un número reducido de observaciones, y ha sido tabulada (ver CRAMER (3), tabla V).

El procedimiento consiste en obtener dos estimaciones distintas de la varianza, descomponiendo sistemáticamente la suma de cuadrados. Por ejemplo, en el caso de verificar si las tres series de determinaciones de la densidad del nitrógeno a partir de compuestos químicos son homogéneas, se forma la suma de los cuadrados de las diferencias de las observaciones con el promedio general:

$$\sum_i \sum_j (x_{ij} - \bar{x})^2 \tag{15.43}$$

en donde i indica la serie de observaciones, y j las observaciones dentro de una serie.

Esta suma de cuadrados se descompone introduciendo los promedios $\bar{x}_i$ de cada serie, lo que teniendo en cuenta que los dobles productos se anulan, conduce a

$$\sum_{i,j} (x_{ij} - \bar{x})^2 = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 + \sum n_i (\bar{x}_i - \bar{x})^2 \tag{15.44}$$

La primer suma se utiliza para estimar el error experimental, pues consiste en diferencias de observaciones con el promedio de su serie, pero la segunda es una estimación de la varianza de los promedios, o sea, de la posible falta de homogeneidad de las series.

El número de los grados de libertad de estas dos estimaciones resulta de la propiedad aditiva, ya empleada en la significación de la diferencia de dos promedios, calculando primero los grados de libertad del total (n-1), luego los de los promedios (número de grupos menos 1) y finalmente los del error experimental por diferencia:

Grados de libertad de la diferencia entre series(diferencia de promedios)	2
Grados de libertad dentro de las series (error experimental)	5
Grados de libertad de la suma total	8 - 1=7

Este método, en cuyos detalles no podemos entrar, ha dado lugar a la extensa e importante teoría del p l a n e o d e e x p e r i m e n t o s, o sea al estudio de los procedimientos para disponer las experimentaciones a fin de poner en evidencia la posible existencia de efectos variados (a).

XV.NUMEROS DE AZAR. Los métodos de inferencia estadística brevemente reseñados en los últimos capítulos parten del supuesto esencial de que las observaciones siguen una ley de probabilidad.

A menudo, como en el ejemplo reiteradamente usado de la emisión de partículas radio activas, el análisis del fenómeno permite conjeturar racionalmente la validez del supuesto.

Pero también es frecuente el caso en que las observaciones dependen de un proceso de elección en el cual interviene el observador, como ocurre cuando se trata de extraer una muestra reducida de una población más amplia, en la inspección de materiales o en los relevamientos estadísticos (ver Cap. IV, párrafo 8).

Los probabilistas de los siglos XVIII y XIX, perfectamente cosncientes de esta dificultad, la resolvían estudiando solamente las extracciones de bolillas de urnas o bolilleros, segun ciertas reglas que garantizan que las observaciones siguen una ley de probabilidad (mezclado de la urna, etc.), y en cuanto a las aplicaciones prácticas, suponían que podían asimilarse a las extracciones de bolilleros,

Pero la forma efectiva de realizar esta asimilación puede ser muy difícil. Si se trata de inspeccionar una partida de materiales, es muy claro que sería absurdo colocar todos los integrantes de la partida dentro de un bolillero para hacer extracciones. Más racional parece asignar un número a cada elemento, y emplear un bolillero con bolillas numeradas. La extracción de una bolilla con un número implica la elección del elemento correspondiente.

Este método tampoco es muy operativo, y la solución se encontraría si se dispusiera de una tabla en la que los números se encontraran "al azar". El sentido preciso de esta expresión sería el de que los dígitos se encontraran en la tabla con frecuencias relativas compatibles con la hipótesis de cada uno tiene igual probabilidad de aparición.

La confección de una tabla de esta tipo no es tampoco problema simple. Una primera tentativa, bastante exitosa, se debe al estadístico inglés L.H.TIPPETT, el que emplea

---

(a) Además de la sucinta exposición del párrafo 16.4 de CRAMER (3), es sumamente interesante el libro de R.A.FISHER: "The Design of The Experiment" del que existe una traducción castellana (B.I.E.T.A. Rosario).
Una introducción bastante amplia a esta teoría se puede ver en R. ANDERSON & T.A. BANCROFT:"Statistical Theory in Research", New York, McGraw Hill, 1952. Tambien se puede consultar la obra de H. de QUENOUILLE,, citada en el párrafo XIV.5.

números cuyo orden de aparición no guardaba aparentemente, ninguna relación, como ocurría con las columnas de números de estadísticas comerciales del Reino Unido, y confeccionó así una tabla de 40.000 dígitos. La idea puede ser utilizada en caso de que no se disponga de una tabla mejor, empleando por ejemplo las dos columnas centrales del grupo de cuatro que en una guía telefónica indican el número de un abonado, pues la ordenación alfabética rompe, aparentemente, cualquier regularidad en la atribución de dichos números (no conviene utilizar la última columna, pues los conmutadores se indican con unos pero incluyen líneas a las que corresponden otros dígitos, los que serían así menos frecuentes, y en la primera, en algunos sistemas telefónicos los números que comienzan con cero se reservan para lineas de prueba, de comunicación interna, etc, que no figuran en guía).

Tablas más seguras son las de KENDALL y BABBINGTON SMITH, referidas en (29), las que son más extensas que las de TIPPETT (contienen 100.000 dígitos), han sido comprobados mediante cuatro métodos distintos de verificación de hipótesis, y su mecanismo de confección parece dar mayor seguridad de que estan realmente al azar (en efecto, una sucesión de números puede dar resultado favorable a un método de verificación de hipótesis, pero no a otro, y de hecho los cuatro millares que no resultaron satisfactorios en la tabla de KENDALL verifican satisfactoriamente algunos de los métodos y otros no).

Estas tablas, debido a los diversos procedimientos de verificación empleados, deben utilizarse solamente leyendo los números en el mismo sentido que si fuera un texto escrito, y tampoco deben romperse, casos de usar grupos de dos o de cuatro dígitos, los que estan indicados.

Si de una población de 50 elementos se desea extraer al azar una muestra de cinco elementos, se asigna a cada elemento un número. A cada número de orden se le asigna a su vez un grupo de dos dígitos (hay 100 grupos de dos dígitos, desde el 00 al 99), y se le eligen los elementos que corresponden a los cinco primeros grupos de dos dígitos.

Por ejemplo, los cinco primeros pares de dígitos de la tabla son:

23 15 75 48 59

Haciendo la convención de asignar a cada número de orden los que corresponden a su doble menos uno y el inmediato inferior, estos cinco primeros grupos de dos dígitos corresponden a:

12 8 38 25 30

por lo que se eligiran como elementos de la muestra aquellos a los cuales correspondió esa identificación.

En las tablas se dan las referencias relativas al método de construcción, el tipo de verificación de hipótesis empledos, los millares excepcionales, etc., de todo lo cual conviene enterarse antes de usarlas.

En caso de usar sistemáticamente la tabla para extraer muestras al azar, conviene en lo posible no repetir los números anteriormente empleados, para evitar posibles correlaciones (de ahí el interés de que la tabla sea extensa).

Tambien se han construído tablas de números de azar, en las que los dígitos estan distribuídos normalmente, para simular una ley normal (referidas en (30)), las que son de uso constante en una variedad de problemas estadísticos y de investigación operativa.

REFERENCIAS BIBLIOGRAFICAS

I. CALCULO DE PROBABILIDADES

(1) CANTELLI: "Calcolo delle Probabilità" (vol. 1), Bologna, N. Zanichelli, 1957.

(2) H.CRAMER:"Mathematical Methods of Statistics", Princeton, Princeton University Press, 1946.

(3) H.CRAMER:"The Elements of Probability Theory and some of its Applications", New York, John Wiley, 1955.

(4) W.FELLER:"An Introduction to Probability Theory and its Applications", New York, John Wiley.

(5) A.FISHER:"The Mathematical Theory of Probability and its Applications to Frecuency Curves and Statistical Method", New York, Macmillan, 1922.

(6) M.FRECHET:"Recherches Theoriques Modernes sur le Calcul des Probabilités" (vol 1) París, G.Villars, 1937).

(7) M.FRECHET:"Les Probabilités associées a un système d'evenements compatibles et dépendants". Paris, Hermann, 1940-1943.

(8) B.W.GNEDENKO:"Lerbuch der Wahrscheinlichkeitsrechnung", Berlin, Akademie Verlag, 1958.

(9) B.W.GNEDENKO und A.J.CHINTSCHIN:"Elementare Einführung in die Wahrscheinlichkeitsrechnung", Berlin, Veb deutscher Verlag der Wissenschaften, 1955.

(10) B.W.GNEDENKO and A.N.KOLMOGOROV:"Limit Distributions for Sums of Independent Random Variables", Cambridge, Addison Wesley, 1954.

(11) H.JEFFREYS:"Theory of Probability", Oxford, Oxford University Press, 1939.

(12) J.M.KEYNES:"A Treatise on Probability", London, Macmillan, 1921.

(13) A.A.KHINTCHINE:"Asymptotische gesetze der Wahrscheinlichkeitsrechnung" (Ergebnisse der Mathematik und ihrer Grenzgebiete), Berlin, Springer, 1933.

(14) A.N.KOLMOGOROV:"Foundations of the Theory of Probability"? New Yor, Chelsea,1950.

(15) P.S.LAPLACE:"Theorie Analytique del Probabilités", París, 1820 (vol. VII de las obras completas).

(16) P.LEVY:"Calcul des Probabilités", Paris, Gauthier Villars, 1925.

(17) P.LEVY:"Theory de l'Adition des Variables Aleatories", Paris, Gauthier Villars, 1937.

(18) M.LOEVE:"Probability Theory", New York. Van Nostrand, 1955.

(19) MARKOFF: "Wahrscheinlichkeitsrechnung",Leipzig, Teubner, 1912.

(20) R. von MISES:"Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und Theoretische Physik", Leipzig und Wien, Deuticke, 1931.

(21) H.RICHTER:"Wahrscheinlichketstheorie", Berlin, Springer, 1956.

(22) J.V.USPENSKY:"Introduction to Mathematical Probability", New York, McGrawHill, 1937.

(23) I.TODHUNTER:"A History of the Mathematical Theory of Probability", New York (reimpresión), Chelsea, 1949.

II. ESTADISTICA

(24) F.N.DAVID:"Probability Theory for Statistical Methods", Cambridge, Cambridge University Press, 1951.

(25) B.W.LINDGREN and G.W.McELRATH:"Introduction to Probability and Statistics", New York, Macmillan, 1959.

(26) A.M.MOOD:"Introduction to the Theory of Statistics", McGraw Hill, New York,1950.

III. TABLAS Y MANUALES

(27) E.S.PEARSON and H.O.HARTLEY:"Biometrika Tables for Statisticians", Cambridge, Cambridge University Press, 1956.

(28) M.G.KENDALL and B.BABINGTON SMITH:"Tables of Random Sampling Numbers" (Tracts f for Computers, Nº XXIV), Cambridge, Cambridge University Press, 1954.

(29) H.WOLD:"Tables of Random Normal Deviates" (Tracts for Computers, NºXXV), Cambridge, Cambridge University Press.

(30) A.S.T.M.:"Manual on Quality Control of Materials".

(31) H.G.ROMIG:"50-100 Binomial Tables", New York, John Wiley, 1947.

INDICE

Erratas advertidas

Pag.	línea	Dice	Debe decir
7	7	de $\frac{1}{37}$	es de $\frac{1}{37}$
7	18	la situación	la caracterización de la situación
14	7	suma lógica (A|B)	suma lógica (A+B)
14	33	par y multiplo de 3	par y del suceso multiplo de 3
30	2	$+(-1)^n\ p(A_1A_2\ldots A_n)$	$+(-1)^{n+1}\ p(A_1A_2\ldots A_n)$
32	20	y quedando células	y reuniendo células
72	6	$< \frac{K\ \sigma^2}{\varepsilon^2\ n}$	$\frac{K^2\ \sigma^2}{\varepsilon^2\ n}$
74	12	$(r+\frac{1}{2})\ \log(\frac{r}{rp})$	$(r+\frac{1}{2})\ \log(\frac{r}{np})$
75	2	$\left(-\frac{tp}{\sqrt{npq}} - \frac{1}{2}\frac{t^2p^2}{\sqrt{npq}}\right)$	$\left(-\frac{tp}{npq} - \frac{1}{2}\frac{t^2p^2}{n\ pq}\right)$
75	5	aproximada	creciente
78	21	$np = 0,5$	$np = 5$
80	1	$< \frac{r - np}{npq} <$	$< \frac{r - np}{\sqrt{npq}} <$
92	27	$n(ps + q)^{n-1}$	$n(ps + q)^{n-1}p$
92	30	$g'(1) =$	$g'(1) = \alpha$
94	22	conjunto que valores	conjunto de valores
101	14	$\left\lceil \frac{(w - a)^3 p}{3!\ \sqrt{n}\ K} \right\rceil$	$\left\lceil \frac{w^3\ p}{3!\ \sqrt{n}K} \right\rceil$
103	28	VON MISES (21)	VON MISES (20)
105	ultima	p..	p_{ij}
108	17	absoluto a	absoluto a
112	15	en problema	en el problema
116	17	(ver párrafo IX)	(ver párrafo IX,5)
119	24	(o en grupos de letras)	(o grupos de dos letras)
129	8	(9.43)	(9.44)
129	11	(9.42)	(9.43)
134	32	0,999 99	0,999 999
135	8	t-1, t , t 1	t-1, t , t+1

Impreso en los talleres del Departamento de Biblioteca y Publicaciones de la Facultad de Ciencias Exactas y Naturales.

ADENDA A LAS ERRATAS ADVERTIDAS

PAG.	LINEA	DONDE DICE:	DEBE DECIR:
21	6	... se puede formar cuatro	... se puede formar con cuatro
27	16	$P(B+C/A) =$	$P((B+C)/A) =$
	18	$P(B+C).A$	$P((B+C).A)$
28	4*	... de A ,	... con A ,
	5*	... con B o con B ,	... con B o con B^,
		, corresponde al	, corresponde el
29	2	... de A^	... con A^ .
32	12	XLX	XIX
34	6	(n 1)-esima...	(n+1)-ésima...
	15	$p_{n+1} = \frac{-}{1+q_n}$	$p_{n+1} = \frac{1}{1+q_n}$
40	5*	...sea el A_j.	... sea el A_{j_o}.
50	7*	$P_n(r) = \frac{n!}{n!(n-r)!} p^r q^{n-r}$	$P_n(r) = \frac{n!}{r!(n-r)!} p^r q^{n-r}$
	2*	$\sum_{r=0}^{n} \frac{n}{n.(n-r)} = (p+q)^n = 1$	$\sum_{r=0}^{n} \frac{n!}{r!(n-r)!} p^r q^{n-r} = (p+q)^n = 1$
51	6*	$p(n-r) \lesseqgtr (r-1)q$	$p(n-r) \lesseqgtr (r+1)q$
52	2	... en $r_.$, ...	en r_1 , ...
	3*	$= \frac{5!}{5!5!} (\frac{1}{2})^{10} =$	$= \frac{10!}{5!5!} (\frac{1}{2})^{10} =$
56	8*	$= \sum_{r=0}^{n} P(r)(r-k)^2 =$	$= \sum_{r=0}^{n} P_n(r)(r-k)^2 =$
58	7*	$P(\lvert\frac{x}{t} - p\rvert > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} = \frac{pq}{n\varepsilon^2}$	$P(\lvert\frac{r}{n} - p\rvert > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} = \frac{pq}{n\varepsilon^2}$
60	6*	$\frac{1}{n^4\varepsilon^2} = \delta$	$\frac{1}{4n\varepsilon^2} = \delta$
74	2	$P_n(r) = \frac{\ldots\ldots\ldots}{\ldots\sqrt{2\pi n}}$	$P_n(r) = \frac{\ldots\ldots}{\ldots\sqrt{2\pi r}}$
	9	$P_n(r) = \ldots (\frac{np}{q})^{r+\frac{1}{2}} \ldots$	$P_n(r) = \ldots (\frac{np}{r})^{r+\frac{1}{2}}$
75	17*	$x = r-pn$	$\lvert x\rvert = \lvert r-pn\rvert$
78	16	valores elevados de u ,	valores elevados de n ,
	24	$P_{10}(4 \leq t \leq 6) =$	$P_{10}(4 \leq r \leq 6) =$
	5*	$P_{10}(4 \leq t \leq 6) =$	$P_{10}(4 \leq r \leq 6) =$
79	11	... np = 0,5)	... np = 5)
81	10*	$t = \frac{x+0,5}{npq}$	$t = \frac{x+0,5}{\sqrt{npq}}$
	5*	VI.10.-	VI.11.-

PAG.	LINEA	DONDE DICE:	DEBE DECIR:
84	12#	...subintervalos T.	...subintervalos ΔT .
85	1	$P_T(y) = \frac{\dots \dots (n-y-1)}{n^y} \dots$	$P_T(y) = \frac{\dots \dots (n-y+1)}{n^y} \dots$
	10#	o sea que el ... ley: n_1	o sea que es ... ley: $\mu_1 = \alpha$
86	13	, con 113 ...	, con 122 ...
89	3	"Studen"	"Student"
91	5#	$= \sum \frac{y}{y!} \dots$	$= \sum \frac{\alpha^y}{y!} \dots$
92	16#	... establecerse si	... establecerse
	3#	$g''(s) = \sum p_r s^{r-2}.(n-1).n$	$g''(s) = \sum p_r s^{r-2}(r-1)r$
93	19	$\mu_2 = g'''(1+ + g''''(1) - g'(1)$	$\mu_2 = g'''(1) - g''(1) + g'(1)$
96	17	p u n t o c e r o ...	p u n t o u = c e r o ...
97	18	$\varphi^*(w) = \sum p_r e^{i(r-m)w}$	$\varphi^*(u) = \sum p_r e^{i(r-m)u}$
	23	$\dots = (pe^{iqu}+qe^{-pu})^n$	$\dots = (pe^{iqu}+qe^{-ipu})^n$
	29	(8.35) ... = ...	(8.35) ... $\rightarrow$ ...
98	12#	pondiente de (8.37)...	pondiente de (8.38)
100	7#	$\underline{f''(a)r^2(a) - \dots}$	$\underline{f''(a).f(a) - \dots}$ .....
108	17	valor absoluto a ,	valor absoluto a ϵ ,
115	16	(9.5)	(9.6)
	17	(9.5)	(9.6)
117	5	(9.15)	(9.13)
	7#	... y T a la ...	... y T_y a la ...
126	20	... desde a los que su vez...	...desde los que a su vez ...
129	5	$p_{ij}^{(t-m)} < p_j + \dots$	$p_j^{(t-m)} < p_j + \dots$
	8	(9.43)	(9.44)
	11	(9.42)	(9.43)
	13	$p_{ij}^{(t-m)} > p_{ij} - \dots$	$p_j^{(t-m)} > p_j - \dots$
137	8	$\lvert T(t)\ T(t)\ T^t \rvert$	$\lvert T_1(t)\ T_2(t)\ T^t \rvert$
141	3	e s c a l e a .	e s c a l e r a .
142	8	$\frac{L}{\sqrt{2\pi}}$	$\frac{1}{\sqrt{2\pi}}$
143	10	$F(X) = \dots\dots(1-x)^{q-1}$	$F(X) = \dots\dots (1-x)^{q-1}\, dx$
	14#	(10.16) $\Gamma^*(n) =$	(10.16) $\Gamma'^*(n) =$
	8#	(10.17) $\Gamma(n) =$	(10.17) $\Gamma'^*(n) =$
145	2	$\dots = \frac{1}{2\pi\sigma} \dots + \frac{m}{}$	$\dots = \frac{1}{\sqrt{2\pi}\,\sigma} \dots + \frac{m}{\sqrt{2\pi}}$

PAG.	LINEA	DONDE DICE:	DEBE DECIR:
145	9	$\ldots = \sigma^2 \{ -(x-m) \ldots$	$\ldots = \frac{\sigma^2}{\sqrt{2\pi}} \{ -\frac{(x-m)}{\sigma} \ldots$
	15	$\mu_4 = \{ \ldots - \frac{(x-m)^3}{\sigma^2} \ldots$	$\mu_4 = \{ \ldots - \frac{(x-m)^3}{\sigma^3} \ldots$
147	3	$\varphi_{(w)} = e^{imw} \varphi x_{(w)} = e^{imw \frac{\sigma^2 w}{2}}$	$\varphi_{(w)} = imw\ \varphi^{*}_{(w)} = e^{imw - \frac{\sigma^2 w^2}{2}}$
	13	(10.38)	debe mantenerse la notación w en vez de u .
	10#	(10.40)	el parámetro es t y no $n/2$
148	2	$\ldots = \ldots = \frac{ni}{(1-2u)^{\frac{n+2}{2}}}$	$\ldots = \ldots = \frac{ni}{(1-2iu)^{\frac{n+2}{2}}}$
	4	$\ldots = \ldots = \frac{ni}{\ldots}$	$\ldots = \ldots = \frac{n(n+2)i^2}{\ldots}$
151	14#	$\varphi(u) =$	$\varphi(t) =$
	10#	$\varphi(\frac{u}{n}) =$	$\varphi(\frac{t}{n}) =$
	7#	$\{\varphi(\frac{u}{n})\}^n =$	$\{\varphi(\frac{t}{n})\}^n =$
	5#	$\lim_{n \to \infty} \{\varphi(\frac{u}{n})\}^n =$	$\lim_{n \to \infty} \{\varphi(\frac{t}{n})\}^n =$
152	15#	$\ldots = \ldots + o(\frac{u^2}{n})^n\} = \ldots$	$\ldots = \ldots + o(\frac{u^2}{n})\}^n =$
153	8	$t_{2p} = \frac{y_p - 30}{\sqrt{30}} \therefore y_p = \sqrt{30 t_{2p}} + 30$	$t_{2p} = \frac{y_p - 30}{\sqrt{60}} \therefore y_p = \sqrt{60 t_{2p}} + 30$
	9	$t_{2p} = \sqrt{2y_p} - \sqrt{29} \quad y_p = \frac{1}{2}(t_{2p} + \sqrt{29})^2$	$t_{2p} = \sqrt{2y_p} - \sqrt{59} \quad y_p = \frac{1}{2}(t_{2p} + \sqrt{59})^2$
157	11	$E\{u(x-m_1\}+\ldots$	$E\{u(X-m_1)+\ldots$
	4#	$\ldots = \ldots + v^2(\sigma_2^2 - \frac{\mu_{11}'}{\sigma_1})$	$\ldots = \ldots + v^2(\sigma_2^2 - \frac{\mu_{11}'^2}{\sigma_1^2})$
158	3	$\sigma_1^2 \sigma_1^2 - \mu_{11}$	$\sigma_1^2 \sigma_2^2 - \mu_{11}'^2 \neq 0$
	2#	$= \ldots - \ldots + \ldots + (m_2 - \alpha + \beta m_1)^2$	$= \ldots - \ldots + \ldots + (m_2 - \alpha - \beta m_1)^2$
159	2#	$xz = y$; $dxdy = xdxdz$	$xz = y$; $dy = xdxdz$
165	9	se haa	se ha
	20	$= \frac{\ldots}{\ldots(1-n)\ldots}$	$= \frac{\ldots}{\ldots(1-p)\ldots}$
	21	, que es en reali	, y es en reali
167	12	$m = \ldots dr =$	$m = \ldots dp =$
	7#	(13.14) $\sigma^2 = \frac{m \quad m}{\ldots} \ldots$	(13.16) $\sigma^2 = \frac{m}{\ldots}$
169	12#	debe vincular...	debe vincularse...
172	9	...a m y a ;	...a m y a σ ;

PAG.	LINEA	DONDE DICE:	DEBE DECIR:
173	9	qie...	que...
174	7*	...primer esperanza...	...segunda esperanza...
	2*		eliminar los dos últimos renglones
175	1	falta en el texto antes del primer renglon	La esperanza matemática del estimador de la varianza...
	5*	El mímite...	El límite...
	2*	XIII.10.-	XIII.12.-
176	4*	XIII.11.-	XIII.13.-
177	6	en donde m es ...	en donde n es ...
	17	$\ldots = \ldots \left(\frac{1}{\alpha} - 1\right)$	$\ldots = \ldots \left(\frac{1}{\alpha} - 1\right)$
	6*	XIII.12.-	XIII.14.-
178	6	XIII.13.-	XIII.15.-
	6	...En el párrafo XIII se...	...En el párrafo XIII.11 se...
179	10*	$\sum_{j} = \ldots$	$\sum_{i} = \ldots$
180	1	$\frac{\partial \Sigma}{\beta} = \ldots (Y_i - \quad - \quad X_i) = 0$	$\frac{\partial \Sigma}{\partial \beta} = \ldots (Y_i - \alpha - \beta X_i) = 0$
	14*	XIII.14.-	XIII.16.-
	11*	$\alpha = \ldots - \ldots = \ldots - \ldots$	$\alpha = \ldots - \beta \ldots = \ldots - \ldots$
	10*	...en la segunda:	... en (13.39)
181	11	XIII.15.-	XIII.17.-
183	12*	, es decir, se da...	, es decir, si se da...
190	7	(14.4)	(14.3)
191	7*	...frados...	...grados...
193	8	... ...11 ...	... ...12 ...
		... ...	... ...
		... ...	... ...
197	6*	...desea usar	...desean usar
198	12	$\ldots = \ldots = \ldots -$	$\ldots = \ldots = \ldots +$
		$- \ldots = \ldots$	$+ \ldots = \ldots$
	5*	$\ldots = \frac{11}{\ldots}$	$\ldots = \frac{1}{\ldots}$
199	2	$\ldots = \ldots e^{-\frac{(n-1)s}{2\sigma^2}}$	$\ldots = \ldots e^{-\frac{(n-1)s^2}{2\sigma^2}}$
	11	$\ldots \frac{\Sigma (x_{i-m})^2}{\ldots}$	$\ldots \frac{\Sigma (x_i - m)^2}{\ldots}$
	15	idem anterior	idem anterior

PAG.	LINEA	DONDE DICE:	DEBE DECIR:
199	17	$\ldots = x$,	$\ldots = \bar{x}$,
	18	y un s da-	y un s^2 da-
	19	o sea $x^2_{p1} = \ldots$	o sea $\chi^2_{p_1} = \ldots$
	14#	$\frac{s^2}{\sqrt{n}}$	$\frac{s^2}{n}$
202	(15.15)		
	(15.18)		falta en el último miembro dt
	(15.19)		
203	15	$t = \sqrt{n-1}\,\frac{y}{s}$	$t = \sqrt{n-1}\,\frac{v}{s}$
	8#	$\ldots + t_p \frac{ss}{\sqrt{n}}) = \ldots$	$\ldots + t_p \frac{s}{\sqrt{n}}) = \ldots$
	16	(15.9)	no existe
204	2	(15.9)	no existe
	1#	...,y en la que se...	..., y en la es ...
206	1	... y e es la ...	... y e_i es la ...
	4	... normales e .	... normales $\bar{e}$.
	6	$y_i = \ldots = \ldots + \ldots - e_i - \ldots = \ldots$	$y_i = \ldots = \ldots + \ldots + e_i - \ldots = \ldots$
	9	porque el número e ...	porque el número $\bar{e}$ ...
	14	$\sum x_i^2$	$\sqrt{\sum x_i^2}$
	9#	$\sum (Y_i - \ldots$	$\sum (y_i - \ldots$
208	4#	$t = \ldots \frac{\ldots}{\ldots}$	$t = \ldots \frac{\ldots}{\sqrt{\ldots}}$
	2#	ema de ...	lema de...
	1#	i cuadrado ...	Ji cuadrado...
210	15	(15.35)	(15.34)

N O T A : Cuando el número de la línea está afectado por un asterisco # , debe comenzarse a contar desde el pié de la página.