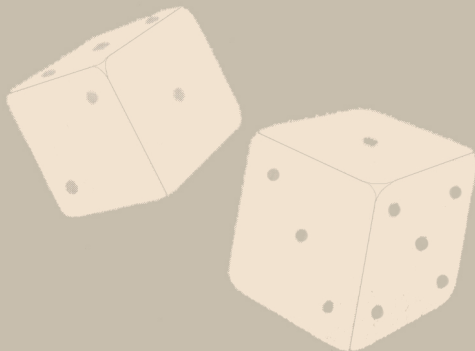
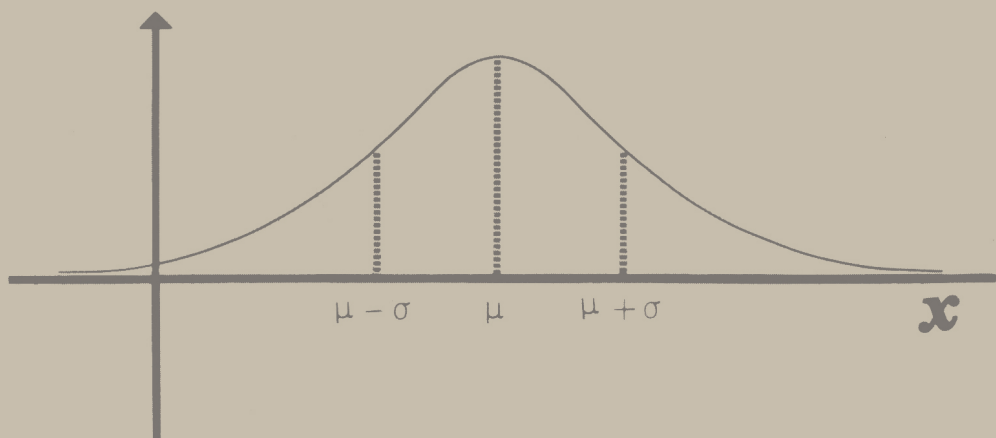


PROBABILIDAD E INFERENCIA ESTADÍSTICA

Programa Regional de Desarrollo Científico y Tecnológico
Departamento de Asuntos Científicos
Secretaría General de la
Organización de los Estados Americanos



PROBABILIDAD E INFERENCIA ESTADISTICA

por

Luis A. Santaló

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Buenos Aires, ARGENTINA

Programa Regional de Desarrollo Científico y Tecnológico

Departamento de Asuntos Científicos

Secretaría General de la

Organización de los Estados Americanos

Washington, D. C.

© Copyright 1970 by
The General Secretariat of the
Organization of American States
Washington, D. C.

Derechos Reservados, 1970
Secretaría General de la
Organización de los Estados Americanos
Washington, D.C.

Primera edición,	1970
Segunda edición, corregida y actualizada	1975

Esta monografía ha sido preparada para su publicación en el
Departamento de Asuntos Científicos de la Secretaría General
de la Organización de los Estados Americanos.

Editora: Eva V. Chesneau

Revisor Técnico: Prof. Evelio Fabbroni
Instituto Interamericano de Estadística
Organización de los Estados Americanos
Washington, D.C.

A los lectores

El programa de monografías científicas es una faceta de la vasta labor de la Organización de los Estados Americanos, a cargo del Departamento de Asuntos Científicos de la Secretaría General de dicha Organización, a cuyo financiamiento contribuye en forma importante el Programa Regional de Desarrollo Científico y Tecnológico.

Concebido por los Jefes de Estado Americanos en su Reunión celebrada en Punta del Este, Uruguay, en 1967, y cristalizado en las deliberaciones y mandatos de la Quinta Reunión del Consejo Interamericano Cultural, llevada a cabo en Maracay, Venezuela, en 1968, el Programa Regional de Desarrollo Científico y Tecnológico es la expresión de las aspiraciones preconizadas por los Jefes de Estado Americanos en el sentido de poner la ciencia y la tecnología al servicio de los pueblos latinoamericanos.

Demostrando gran visión, dichos dignatarios reconocieron que la ciencia y la tecnología están transformando la estructura económica y social de muchas naciones y que, en esta hora, por ser instrumento indispensable de progreso en América Latina, necesitan un impulso sin precedentes.

El Programa Regional de Desarrollo Científico y Tecnológico es un complemento de los esfuerzos nacionales de los países latinoamericanos y se orienta hacia la adopción de medidas que permitan el fomento de la investigación, la enseñanza y la difusión de la ciencia y la tecnología; la formación y perfeccionamiento de personal científico; el intercambio de informaciones, y la transferencia y adaptación a los países latinoamericanos del conocimiento y las tecnologías generadas en otras regiones.

En el cumplimiento de estas premisas fundamentales, el programa de monografías representa una contribución directa a la enseñanza de las ciencias en niveles educativos que abarcan importantes sectores de la población y, al mismo tiempo, propugna la difusión del saber científico.

La colección de monografías científicas consta de cuatro series, en español y portugués, sobre temas de física, química, biología y matemática. Desde sus comienzos, estas obras se destinaron a profesores y alumnos de ciencias de enseñanza secundaria y de los primeros años de la universitaria; de éstos se tiene ya testimonio de su buena acogida.

Este prefacio brinda al Programa Regional de Desarrollo Científico y Tecnológico de la Secretaría General de la Organización de los Estados Americanos la ocasión de agradecer al doctor Luis A. Santaló, autor de esta monografía, y a quienes tengan el interés y buena voluntad de contribuir a su divulgación.

Septiembre de 1970

INDICE

	Página
A los Lectores	iii
CAPITULO PRIMERO. LA DEFINICION CLASICA DE PROBABILIDAD Y PRIMEROS EJEMPLOS	
1. La Definición Clásica	1
2. Principios Fundamentales	4
3. Probabilidad y Combinatoria	6
4. Probabilidad y Frecuencia	9
5. La Probabilidad Subjetiva o Grado de Creencia ..	9
6. La Probabilidad en Torneos	11
CAPITULO SEGUNDO. DEFINICION AXIOMATICA DE PROBABILIDAD Y PRIMERAS CONSECUENCIAS	
1. Axiomas de la Teoría de Probabilidades	13
2. Consecuencias de los Axiomas	17
3. La Definición Clásica	19
4. Probabilidad Condicional	19
5. Ejemplos	22
CAPITULO TERCERO. VARIABLES ALEATORIAS. FUNCIONES DE PROBABILIDAD	
1. Variables Aleatorias	33
2. Esperanza Matemática	37
3. Momentos de Una Variable Aleatoria	38
4. La Desigualdad de Tchebycheff	39
5. Suma de Variables Aleatorias	40
6. Producto de Variables Aleatorias	42
7. Correlación	44
8. Función Generatriz de Momentos	45
9. Función Característica	47
10. Regresión	47
CAPITULO CUARTO. DISTRIBUCION BINOMIAL. LEY DE LOS GRANDES NUMEROS	
1. Distribución Binomial	49
2. Esperanza Matemática y Varianza de Una Variable Binomial	52
3. Teorema de Bernoulli	53
4. Leyes de los Grandes Números	55

CAPITULO QUINTO. DISTRIBUCION DE POISSON

1. Función de Probabilidad de Poisson	59
2. Distribución Uniforme de Puntos Sobre Una Recta o de Sucesos en el Tiempo	63
3. El Problema de las Filas de Espera	66

CAPITULO SEXTO. DISTRIBUCION NORMAL.
VARIABLES ALEATORIAS CONTINUAS

1. Aproximación de la Distribución Binomial por la Normal	71
2. Variables Aleatorias Continuas	72
3. Función de Densidad de Probabilidad y Función de Distribución Normal	75
4. Teorema Central del Límite	81
5. Otras Funciones de Densidad	85

CAPITULO SEPTIMO. LA FORMULA DE BAYES

1. Inferencia Estadística	87
2. Fórmula de Bayes	88

VI

CAPITULO OCTAVO. ESTIMACION POR PUNTO

1. Muestras	91
2. Media y Varianza de Una Muestra	92
3. Estimadores	94
4. Estimadores Consistentes	96
5. Estimadores Suficientes	97
6. Estimadores Eficientes	99
7. Estimadores de Máxima Verosimilitud	101

CAPITULO NOVENO. ESTIMACION POR INTERVALOS
DE CONFIANZA. VERIFICACION DE HIPOTESIS

1. La Distribución χ^2	105
2. La Distribución t de Student	107
3. Estimación por Intervalos de Confianza	109
4. Verificación de Hipótesis	111
5. Comparación de Distribuciones Experimentales y Teóricas: Test del χ^2	112

APENDICE I. COMBINATORIA	119
--------------------------------	-----

APENDICE II. ALGUNAS FORMULAS DE CALCULO	122
---	-----

APENDICE III. TABLAS	125
----------------------------	-----

Bibliografía	131
--------------------	-----

LA DEFINICION CLASICA DE PROBABILIDAD Y PRIMEROS EJEMPLOS

1. DEFINICION CLASICA

Las teorías matemáticas, sobre todo las que tienen relación con los fenómenos naturales, a los que tratan primero de entender para luego predecir, se construyen siempre a partir de conceptos intuitivos, suficientemente claros para que puedan ser aplicados en las primeras etapas de la teoría, pero no suficientemente rigurosos para que queden a salvo de objeciones cuando la misma alcanza cierto grado de desarrollo. Se hace necesario, entonces, revisar los fundamentos, precisando las definiciones y dándoles, si ello es posible, una construcción axiomática. De esta manera, si alguna vez aparecen resultados contradictorios o paradójicos, se sabrá bien que lo único que deberá ser revisado son los axiomas y a ellos habrá que acudir también siempre que se desee una extensión o una limitación de la teoría.

Sin embargo, para empezar a estudiar una teoría, no siempre es el camino axiomático el más recomendable. Los axiomas son elaborados por quienes conocen muy bien la teoría, y su verdadero sentido y necesidad se comprenden con claridad tan sólo cuando se está ya familiarizado con los elementos y relaciones básicas que ellos tratan de afirmar de manera abstracta y precisa. Es mejor empezar por definiciones tal vez no muy exactas y con ejemplos simples, pero substanciales, para poder comprender luego el verdadero sentido de los axiomas, y para que los mismos aparezcan de manera natural como expresión sintética y firme de conocimientos ya adquiridos.

En el caso de la teoría de las probabilidades vamos a seguir este camino, dando en este primer capítulo la definición clásica y algunos ejemplos, para ver luego cómo, a partir de ellos como modelo, se puede construir una teoría axiomática.

La definición clásica de probabilidades es la que figura por ejemplo en la famosa "Teoría Analítica de las Probabilidades" (Libro II, cap. 1) de P.S. Laplace, obra históricamente fundamental, publicada en 1812. Es la siguiente:

Definición 1.1. *Probabilidad* de un suceso es la razón entre el número de casos favorables y el número total de casos posibles, siempre que nada obligue a creer que alguno de estos casos debe tener lugar de preferencia a los demás, lo que hace que todos sean, para nosotros, igualmente posibles.

Por ejemplo, supuesto un dado ordinario, o sea, un cubo con las caras numeradas del uno al seis, al ser lanzado de manera arbitraria sobre un plano, la probabilidad de que salga el número 2 será $1/6$. En efecto, si el dado está bien construido y el lanzamiento se hace completamente al azar, no hay nada que obligue a creer que una de las caras tenga que salir de preferencia a las demás; las seis caras son, por tanto, igualmente posibles. Existiendo un solo caso favorable y seis casos posibles, la probabilidad será $1/6$, como ya dijimos.

En cambio, si en vez de un dado de forma cúbica, suponemos que el mismo se alarga según una de sus dimensiones hasta formar un prisma recto de base cuadrada y altura mayor que los lados de la base, evidentemente las seis caras dejan de ser "igualmente posibles". La probabilidad de que salga una determinada cara ya no es $1/6$, pues si bien continúa habiendo un solo caso favorable y seis posibles, entre estos últimos los hay que están en mejores condiciones para salir que otros; es más probable, en efecto, que salga una cara lateral que no una base y esta probabilidad aumenta al aumentar la altura del prisma.

2

Como segundo ejemplo, supongamos un bolillero o urna con 20 bolillas blancas y 8 rojas. Sacando una bolilla al azar, ¿cuál es la probabilidad de que resulte roja? Hay 28 casos posibles, de los cuales sólo 8 son favorables; luego, la probabilidad será $8/28 = 2/7$. Este resultado supone que todas las bolillas son igualmente posibles; si hubiera algunas con más probabilidad de salir que otras (por ejemplo, si las hubiera de diferentes tamaños) la probabilidad sería distinta.

El inconveniente de la definición anterior radica precisamente en la necesidad de fijar la atención en que todos los casos sean "igualmente posibles". Hay veces, como en los ejemplos que acabamos de mencionar, en que el cumplimiento o no de esta condición salta a la vista, mientras que otras veces tal hecho pasa inadvertido y se necesita mucha atención y "buen sentido" para ponerlo de manifiesto.

Muchas veces, aun cuando los casos no sean igualmente posibles, se puede aplicar la definición anterior, con tal de tener cuidado de contar cada caso tantas veces como posibilidades hay

de que se verifique. Consideremos, siguiendo a Laplace, el ejemplo siguiente:

"Se lanza al aire por dos veces consecutivas una moneda cuyos lados opuestos llamaremos "cara" y "cruz". Se pide la probabilidad de obtener cara por lo menos una vez". A primera vista podría creerse que los casos posibles son: las dos veces cara, una vez cara y otra cruz, y las dos veces cruz. Contando de esta manera nos encontramos con dos casos favorables y tres casos posibles; por tanto la probabilidad sería $2/3$.

Sin embargo, un poco de atención pone de manifiesto que el caso en que sale una vez cara y otra vez cruz es más frecuente que los otros dos; en efecto, puede ocurrir que salga la primera vez cara y la segunda cruz o viceversa. El segundo caso tiene, pues, dos posibilidades y hay que contarlos dos veces, como si se tratara de dos casos diferentes. Representada la cara por C y la cruz por F, los verdaderos casos posibles son (C, C), (C, F), (F, C), (F, F), y por tanto hay tres casos favorables y cuatro posibles, de manera que la solución del problema es $3/4$.

La definición de Laplace supone que el número de casos favorables, y por tanto también el de casos posibles, es finito. En este caso la probabilidad de un suceso es siempre un número real comprendido entre 0 y 1. La probabilidad 0 significa que no hay ningún caso favorable, o sea, que el suceso es *imposible*. La probabilidad 1 significa que el número de casos favorables es igual al número de casos posibles, o sea, que el suceso es *seguro*.

Si hay infinitos casos posibles, la definición anterior debe modificarse, pero ello se hace fácilmente sustituyendo el *número* de casos favorables o posibles, por la *medida* de los mismos. La teoría de probabilidades aparece así íntimamente ligada con la teoría de la *medida* de conjuntos. Supongamos el siguiente problema:

Elegido al azar un punto dentro de un segmento A de longitud a , ¿cuál es la probabilidad de que pertenezca a otro segmento B de longitud b contenido en el primero?

Naturalmente, el número de casos posibles, que es el número de puntos del segmento A, lo mismo que el número de casos favorables, que es el número de puntos del segmento B, son ambos infinitos. Sin embargo, considerando los puntos como igualmente posibles, basta tomar la longitud de cada segmento (o, con mayor generalidad, su medida, si se tratara de conjuntos de puntos que no llenaran todo el segmento), en vez del número de puntos, para que

la definición sea aplicable, y la probabilidad resulta ser igual a b/a .

Obsérvese en este ejemplo, que si en vez del segmento B considerásemos un conjunto finito de puntos del segmento A, la probabilidad resultaría ser igual a 0, sin que ello signifique imposibilidad. Análogamente, si en vez de un segmento contenido en A tomamos un conjunto de puntos B que sea el mismo A, excepto un número finito de puntos, como la medida de B es igual a la de A, resulta que la probabilidad es 1, sin que ello signifique certeza. O sea: el hecho de que la probabilidad 0 implique imposibilidad y la probabilidad 1 certeza no es válido si se trata de la probabilidad de conjuntos infinitos.

2. PRINCIPIOS FUNDAMENTALES

Probabilidades Totales. Hemos visto que la probabilidad de que al lanzar un dado al azar salga el número 2 es $1/6$. Supongamos que se busca la probabilidad de que "salga el 2 o el 5". Tenemos en este caso dos casos favorables, el 2 y el 5, dentro de los mismos seis casos posibles. Por tanto, la probabilidad es $2/6 = 1/3$.

4

Supongamos, como segundo ejemplo, una urna con 20 bolillas blancas, 16 rojas y 5 azules; en total 41 bolillas. La probabilidad de que una bolilla sacada al azar sea blanca es $P(\text{blanca}) = 20/41$, puesto que hay 20 casos favorables y 41 posibles. Análogamente la probabilidad de que la bolilla sacada al azar sea roja es $P(\text{roja}) = 16/41$. Supongamos que se quiera hallar la probabilidad de que la bolilla extraída sea "blanca o roja". Los casos favorables son $20 + 16 = 36$, y los posibles son los mismos 41 de antes. Por tanto la probabilidad buscada será $P(\text{blanca o roja}) = (20 + 16)/41 = 20/41 + 16/41 = P(\text{blanca}) + P(\text{roja})$.

El hecho es general. Supongamos que el número total de casos posibles de un determinado suceso es n , dentro de los cuales hay m casos en que se produce el suceso A y r casos en que se produce el suceso B. Será $P(A) = m/n$, $P(B) = r/n$. Si A y B no pueden tener lugar simultáneamente, o sea los m casos correspondientes a A son distintos de los r casos correspondientes a B (se dice que A y B son *incompatibles*), entonces el número de casos en que puede tener lugar A o B será $m + r$ y por tanto

$$P(A \text{ o } B) = \frac{m+r}{n} = \frac{m}{n} + \frac{r}{n} = P(A) + P(B). \quad [1.1]$$

Este hecho constituye el clásico *Principio de las Probabilidades Totales*, que se enuncia:

Si dos sucesos A y B son incompatibles, la probabilidad de que ocurra "A o B" es igual a la suma de la probabilidad de A más la probabilidad de B.

La condición de que A y B sean incompatibles, es decir, que no puedan ocurrir simultáneamente, es indispensable. Consideremos, por ejemplo, el mismo caso de un dado numerado de 1 a 6 y supongamos que A representa el suceso de que salga el 2 y B el suceso de que salga un número par. Se tiene así $P(A) = 1/6$ y $P(B) = 3/6 = 1/2$ (puesto que entre 1 y 6 hay tres números pares). La probabilidad de "A o B", o sea, la que salga el 2 o salga un número par no es, sin embargo, $P(A) + P(B) = 2/3$, puesto que si ocurre A, también ocurre B, por ser 2 un número par. Los sucesos no son incompatibles y por tanto la relación [1.1] no es válida. La probabilidad en este caso es simplemente igual a la de B o sea $1/2$.

Obsérvese que el razonamiento puede repetirse con un número de sucesos mayor que dos. Siempre que ellos se excluyan mutuamente, es decir, sean incompatibles entre sí, la probabilidad de que ocurra uno u otro es igual a la suma de las probabilidades respectivas.

Probabilidades Compuestas. Consideremos el siguiente problema:

"Se lanzan simultáneamente dos dados, uno rojo y otro blanco, y se busca la probabilidad de que el rojo salga 2 y el blanco 5."

Los casos posibles son ahora $6 \cdot 6 = 36$, pues en el dado rojo puede salir cualquier número del 1 al 6 y, por cada caso, en el dado blanco puede también salir cualquier valor entre 1 y 6. De entre estos 36 casos posibles, únicamente hay un caso favorable, y por tanto la probabilidad buscada es $1/36$. Se tiene así que, mientras que la probabilidad de que en el dado rojo salga 2 es $1/6$ y la de que en el dado blanco salga 5 es también $1/6$, la probabilidad de que ocurran a la vez los dos sucesos es igual al producto $(1/6) \cdot (1/6) = 1/36$.

Este hecho es general. Supongamos que para un suceso A hay m_1 casos favorables entre n_1 posibles y que para otro suceso B hay m_2 casos favorables entre n_2 casos posibles. Entonces se tiene $P(A) = m_1/n_1$ y $P(B) = m_2/n_2$. Consideremos ahora el conjunto de pares de casos posibles y favorables. Es claro que habrá $n_1 n_2$ pares de casos posibles y, entre ellos, habrá $m_1 m_2$ casos favorables (en que tienen lugar A y B simultáneamente). Por tanto $P(A \text{ y } B) = m_1 m_2 / n_1 n_2 = P(A) \cdot P(B)$.

Podemos, por tanto, enunciar el siguiente *Principio de las Probabilidades Compuestas*:

Si dos sucesos son independientes entre sí, la probabilidad de que ocurran A y B a la vez, es igual al producto de la probabilidad de A por la probabilidad de B, o sea,

$$P(A \text{ y } B) = P(A) \cdot P(B). \quad [1.2]$$

La condición de que A y B sean *independientes* es fundamental. En caso contrario, la igualdad [1.2] deja de ser válida. Supongamos, por ejemplo, que se lanza un dado y se busca la probabilidad de que el número resultante sea "par y menor que 4". Si A indica el suceso "par" y B el suceso "menor que 4", resulta $P(A) = 1/2$ y $P(B) = 3/6 = 1/2$. El único caso favorable es el 2 y por tanto $P(A \text{ y } B) = 1/6$, mientras que $P(A) \cdot P(B) = 1/4$. Esto nos dice que los sucesos "par" y "menor que 4" no son independientes: si se sabe ya que es par, es menos probable que sea menor que 4 que si no se supiera.

Muchas veces no es fácil averiguar *a priori* si dos sucesos son o no independientes. Por esto, veremos en el capítulo siguiente que es mejor tomar la condición [1.2] como *definición* de independencia, es decir, se dice que dos sucesos A y B son independientes si se cumple [1.2], y no lo son en caso contrario.

6

3. PROBABILIDAD Y COMBINATORIA

Mediante la definición de Laplace y los principios fundamentales anteriores, y siempre que el número de casos posibles sea finito (éstos se llaman *probabilidades finitas*), se pueden resolver muchos problemas. En realidad son problemas de análisis combinatorio, en los cuales la idea de probabilidad únicamente sirve para darles un enunciado atractivo, pero cuya mayor o menor dificultad sólo estriba en los razonamientos y fórmulas de tipo combinatorio que es necesario emplear. Por esto hemos reunido en el Apéndice I las fórmulas necesarias. Vamos a ver algunos ejemplos.

Problema 1.1. Una urna contiene bolillas numeradas del 1 al 5. Se sacan sucesivamente al azar las 5 bolillas (sin reposición). Se busca la probabilidad de que juntando los números de cada bolilla según el orden de extracción resulte el número 21345.

Solución. Los casos posibles son las permutaciones que pueden formarse con las 5 bolillas, o sea, $P_5 = 5! = 120$. Puesto que hay un solo caso favorable, el resultado es $1/120$.

Problema 1.2. Una urna contiene 10 bolillas numeradas del 0 al 9. Se sacan sucesivamente al azar 5 bolillas (sin reposición).

Se busca la probabilidad de que juntando los números de cada una en el orden de extracción resulte el número 80314.

Solución. El número de casos posibles es el número de permutaciones o variaciones de 5 elementos elegidos entre 10 (véase Apéndice 1), o sea, $P_{10,5} = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 30240$. Luego la probabilidad buscada es $1/30240$.

Problema 1.3. Una urna contiene a bolillas blancas y b bolillas negras. Al sacar al azar r bolillas de una vez (suponiendo $r \leq a$), ¿cuál es la probabilidad de que todas ellas sean blancas?

Solución. El número de casos posibles es igual al número de combinaciones de r bolillas tomadas entre las $a + b$ existentes, o sea, es igual a $C_{a+b,r}$. El número de casos favorables es igual al número de combinaciones de r bolillas tomadas entre las a blancas, o sea, $C_{a,r}$. Por tanto, la probabilidad buscada es:

$$P = \frac{C_{a,r}}{C_{a+b,r}} = \frac{a!(a+b-r)!}{(a-r)!(a+b)!}.$$

Problema 1.4. Se tiene 5 pares de zapatos mezclados y cada par es distinto de los demás. Si se eligen 2 zapatos al azar, ¿qué probabilidad hay de que correspondan a un mismo par?

Solución. El número de casos posibles es igual al de combinaciones de 2 objetos elegidos entre 10, o sea, $C_{10,2} = 45$. Los casos favorables son 5, igual en número al de los 5 pares existentes. Por tanto la probabilidad buscada es $P = 5/45 = 1/9$.

Problema 1.5. Se tiene una baraja de 40 cartas, donde hay 4 ases. Se reparten éstas entre 4 personas, de manera que toquen 10 cartas a cada una. ¿Qué probabilidad hay de que a cada una toque un as?

Solución. Los casos posibles son las particiones de 40 en conjuntos de 10, o sea

$$P_{10,10,10,10}^{40} = \frac{40!}{(10!)^4}.$$

Para contar los casos favorables, observemos que prescindiendo de los ases, las 36 cartas restantes se pueden distribuir de $P_{9,9,9,9}^{36} = (36!)/(9!)^4$ maneras y luego los 4 ases de $4!$ maneras. Por tanto, el número de casos favorables es el producto de estos dos números y la probabilidad buscada resulta $P = 0,109$.

Problema 1.6. Si n personas se sientan al azar en una fila, ¿cuál es la probabilidad de que dos de ellas determinadas queden una al lado de la otra?

Solución. El número de casos posibles es $n!$ Para contar los favorables, se suponen las $(n - 2)!$ maneras en que pueden sentarse las restantes personas y se intercalan sucesivamente las dos que se quieren vecinas, desde el principio al final. Hay $n - 1$ posibilidades, y alternando las dos personas quedan, en cada caso, dos posibilidades más. Por tanto el resultado es $P = 2(n - 1) \cdot (n - 2)!/n! = 2/n$.

Problema 1.7. Se elige al azar un número de 6 cifras. Hallar la probabilidad de que todas las cifras sean diferentes.

Solución. Los números de 6 cifras son los comprendidos entre 100.000 y 999.999, o sea, los casos posibles son $9 \cdot 10^5$. Los casos favorables se calculan de la siguiente manera: la primera cifra puede ser cualquiera de las 1, 2, ..., 9 (o sea, hay 9 posibilidades); la segunda cifra puede ser cualquiera de las 0, 1, 2, ..., 9 menos la ya elegida (o sea, hay 9 posibilidades); la tercera puede ser cualquiera de las 0, 1, 2, ..., 9 menos las dos ya elegidas (o sea, hay 8 posibilidades). Así sucesivamente, se ve que los casos favorables son $9 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 = 152.880$. La probabilidad buscada resulta igual a $P = 0,151$.

Problema 1.8. En una urna hay 20 bolillas numeradas de 1 a 20. Si se van sacando al azar una a una sin reposición, ¿cuál es la probabilidad de que la bolilla número 8 salga precisamente en la octava extracción?

Solución. La probabilidad de que la bolilla No. 8 no salga la primera vez es $19/20$; la probabilidad de que tampoco salga la segunda vez es $18/19$ y así sucesivamente. La probabilidad de que sí salga la octava vez es $1/13$, puesto que quedan 13 bolillas. Luego, por el principio de las probabilidades compuestas, la probabilidad buscada es:

$$P = \frac{19}{20} \cdot \frac{18}{19} \cdot \frac{17}{18} \cdots \frac{13}{14} \cdot \frac{1}{13} = \frac{1}{20}.$$

Obsérvese que este resultado es independiente del número 8; es decir, la probabilidad es la misma para cualquier bolilla y cualquier extracción, lo que es evidente *a priori*, pues cualquier bolilla tiene la misma probabilidad de salir en cualquier extracción.

4. PROBABILIDAD Y FRECUENCIA

En los casos de lanzamiento de dados o extracción de bolillas, bajo ciertas condiciones hasta ahora no muy precisas de "igual posibilidad", la probabilidad se puede calcular *a priori*. En realidad, como hemos visto, se trata de problemas de análisis combinatorio, disfrazados con lenguaje probabilístico.

Hay otro tipo de problemas en que la probabilidad aparece como resultado de muchos ensayos o pruebas, sin que se pueda pensar en calcularla de antemano, ya sea por desconocerse la manera de actuar de las causas que originan el fenómeno, ya sea por ser éstas demasiado numerosas o complicadas. Supongamos, por ejemplo, que en una determinada población se ha observado que en los últimos 100.000 nacimientos, 51.600 de los nacidos han sido varones. Se dice entonces que la "probabilidad" de que un nuevo nacimiento sea de varón es igual a $51.600/100.000 = 0,516$. Esta probabilidad se llama *probabilidad experimental, probabilidad estadística o frecuencia*. Se obtiene también como cociente al dividir el número de casos favorables por el número de casos posibles. Se diferencia de los problemas del número anterior en que allí los casos favorables se calculaban a partir de ciertas hipótesis, sin necesidad de ensayos, en tanto que ahora son resultados de la experiencia.

Otro ejemplo clásico es el siguiente. Supongamos que se toma nota del número de personas de una población que fallecen a los x años. Si \bar{d}_x es el número de personas que fallecen a los x años de edad, dentro de un número total l_x de personas consideradas, se puede decir que la "probabilidad de vida" a los x años (o sea, la probabilidad de que una persona de x años llegue a cumplir $x + 1$ años) es $(l_x - \bar{d}_x)/l_x$. Estos datos figuran en las llamadas tablas de mortalidad, que varían con la población de que se trata y con el tiempo, y en ellas se basan los seguros de vida.

La importancia práctica fundamental del cálculo de probabilidades estriba en que la probabilidad experimental y la probabilidad teórica, cuando ésta se puede calcular y bajo condiciones muy amplias, tienden a coincidir cuando el número de experimentos es grande. Es decir, si en un problema de lanzamiento de dados o de extracción de bolillas se repite el experimento un gran número de veces y se calcula la frecuencia de cierto suceso, ella tiende a coincidir (en el sentido que precisaremos más adelante) con la probabilidad teórica calculada de antemano.

5. LA PROBABILIDAD SUBJETIVA O GRADO DE CREENCIA

Tanto las probabilidades que pueden ser calculadas de antemano, como las que resultan como frecuencia de cierto proceso

experimental, son probabilidades objetivas, que pueden expresarse en números y ser sometidas al cálculo matemático. Son las únicas de que trata el cálculo de probabilidades.

Hay otros casos, sin embargo, en que la palabra "probabilidad" se usa en un sentido menos preciso, casos en que, si bien es posible una evaluación más o menos grosera de la misma, no se pueden dar reglas para su determinación precisa, y por tanto escapa al tratamiento matemático. Se trata, más bien, de un "grado de creencia" acerca de que tengao no lugar un determinado hecho. Supongamos el siguiente caso:

"Dos países A y B están en guerra, ¿cuál es la probabilidad de que triunfe el país A?"

Evidentemente, no es lícito decir: como hay dos casos posibles, la derrota o la victoria, y uno solo favorable, la probabilidad es $1/2$. En efecto, puede muy bien ser que los dos casos no sean igualmente posibles y hay que juzgar teniendo en cuenta la potencialidad bélica, las condiciones geográficas, la ciencia militar de los estados mayores, etc. para evaluar la mayor o menor posibilidad de cada caso. Se comprende que es una cuestión compleja y que no cabe pensar en una evaluación exacta de la probabilidad. Tampoco se puede acudir a la probabilidad experimental o estadística, pues la guerra en cuestión es un caso único que no se puede repetir sucesivamente, ni se dispone de un gran número de casos análogos, como en la probabilidad de vida o muerte de una persona.

10

Sin embargo, tampoco es exacto decir que no tiene sentido hablar de probabilidad en este caso, pues no hay duda de que es posible cierta evaluación de la misma. Podría, por ejemplo, preguntarse la opinión de 100 personas capacitadas e imparciales de un país neutral y dividir por 100 el número de respuestas favorables a la victoria del país A. Es muy probable que si el resultado arrojase un saldo favorable del 90%, o sea, una probabilidad $P = 0,9$, este resultado no diferiera mucho del que se obtendría si en vez de 100 personas se consultaran a 1.000, igualmente capacitadas e imparciales.

De todas maneras este tipo de problemas no puede considerarse incluido en la esfera de la probabilidad experimental, puesto que el resultado no proviene de una estadística de casos idénticos o muy análogos, sino que se basa en opiniones subjetivas de un caso único, opiniones que, aun suponiendo que hayan sido emitidas con conocimiento de las condiciones bélicas de los dos países, están influenciadas por la apreciación subjetiva de las mismas (valentía, capacidad directiva, voluntad de lucha, ...) que no puede "medirse" ni graduarse en una escala imparcial y objetiva.

Preguntas del mismo tipo son: ¿Cuál es la probabilidad de que haya vida en Marte?, ¿cuál la de que Napoleón muriera envenenado?, ¿cuál la de que el equipo x gane el campeonato?

Todos estos problemas quedan fuera de la teoría de las probabilidades en el sentido usual, que es el que vamos a seguir en los capítulos siguientes, si bien se han hecho y se siguen haciendo interesantes tentativas para que ellos puedan tratarse también por métodos matemáticos.

6. LA PROBABILIDAD EN TORNEOS

El triunfo en las competencias deportivas depende en gran medida de la influencia del azar, tal vez más de lo que parece a primera vista. Consideremos dos ejemplos:

1. Sea un juego entre dos jugadores o dos equipos, A y B. Supongamos que la probabilidad de ganar A sea 0,6 y la de ganar B sea 0,4. La probabilidad de empate se supone nula, puesto en el caso de empatar, la partida se repite hasta que gana los jugadores.

Si juegan una sola partida, la probabilidad de resultar campeón B, que es el que juega peor, es 0,4, que es bastante grande. Se puede disminuir esta probabilidad si se conviene en jugar 3 partidas y declarar campeón al que gane por lo menos 2 de ellas. Entonces la probabilidad de que B resulte campeón es $p_3 = 3(0,4)^2 \cdot 0,6 + (0,4)^3 = 0,35$, que es inferior a la probabilidad anterior. Si se juegan 5 partidas y se declara campeón al que gane por lo menos 3 de ellas, la probabilidad de que el campeón sea B es $p_5 = 10 \cdot (0,4)^2(0,6)^2 + 5(0,4)^3 \cdot 0,6 + (0,4)^5 = 0,317$. Si se juegan 7 partidas y se declara campeón al que gane por lo menos 4, la probabilidad de que resulte campeón B es $p_7 = 0,290$. Vemos que la probabilidad va disminuyendo, pero con lentitud, de manera que aun en un número relativamente grande de partidas, la probabilidad de que resulte campeón el peor es bastante significativa (aun que por cierto siempre es inferior a 0,5).

2. Cuando se celebran torneos eliminatorios, o sea sin que jueguen todos contra todos, la influencia del orden en las eliminatorias es mucha. Vamos a dar un ejemplo de tres jugadores A, B y C, tales que A juega mejor que B, B juega mejor que C y, sin embargo, C juega mejor que A. Al decir "A juega mejor que B" quiere decir que la probabilidad de que A gane a B es superior a la de que B gane a A, o sea que en un número grande de partidas entre estos jugadores, el número de las partidas ganadas por A supera al de las ganadas por B.

Supongamos que cada jugador dispone de un dado especial, cuyas caras están numeradas según indica la tabla siguiente:

A	1	2	2	3	3	3
B	1	1	2	2	6	6
C	1	1	2	3	5	5

El juego consiste en que cada jugador lanza su propio dado y gana el que saca el número mayor. Si sale el mismo número, hay empate, y se repite el juego.

Si se llama $p(X, Y)$ a la probabilidad de que X gane a Y (siendo $X, Y = A, B, C$), por simple recuento de casos posibles y favorables, se encuentran las siguientes probabilidades:

$$p(A, B) = 16/30, \quad p(B, A) = 14/30, \quad p(C, A) = 16/29$$

$$p(A, C) = 13/29, \quad p(B, C) = 16/30, \quad p(C, B) = 14/30$$

12

de manera que

$$p(A, B) > p(B, A), \quad p(B, C) > p(C, B)$$

y sin embargo

$$p(A, C) < p(C, A)$$

lo cual parece paradójico.

2

DEFINICION AXIOMATICA DE PROBABILIDAD Y PRIMERAS CONSECUENCIAS

1. AXIOMAS DE LA TEORIA DE LAS PROBABILIDADES

En los problemas de probabilidad que hemos visto en el capítulo anterior se notan los siguientes elementos: a) el conjunto de los casos posibles; b) el conjunto de los casos favorables; c) la probabilidad de cada caso favorable, que es un número real comprendido entre 0 y 1. Queremos ahora precisar las definiciones de estos conjuntos y dar forma axiomática a la teoría de la probabilidad.

El conjunto de "casos posibles" puede ser cualquier conjunto abstracto, con la condición de que sus elementos aparezcan como resultado de un experimento (lanzar un dado, sacar una bolilla de una urna, elegir un número en una tabla, ...). En vez de "conjunto de casos posibles" se ha introducido el nombre de "espacio muestral", cuya definición es la siguiente:

Definición 2.1. Se llama *espacio muestral*, asociado a un experimento, al conjunto de todos los resultados posibles de dicho experimento.

En general representaremos por E a los espacios muestrales. Los experimentos que dan lugar a espacios muestrales se llaman *experimentos aleatorios*.

Para que un conjunto de elementos sea un espacio muestral debe haber, por tanto, un cierto proceso cuyos resultados sean, precisamente, los elementos del conjunto.

Ejemplo 1. Los números 1, 2, 3, 4, 5, 6 constituyen el espacio muestral del experimento aleatorio de "lanzar un dado" cuyas caras estén numeradas del 1 al 6.

Ejemplo 2. Las letras C y F constituyen el espacio muestral del experimento aleatorio de "lanzar una moneda", proceso cuyo resultado puede ser cara (C) o cruz (F).

Ejemplo 3. Los pares (C, C), (C, F), (F, C), (F, F) constituyen el espacio muestral del proceso de "lanzar dos monedas simultáneamente".

Ejemplo 4. Los números 3, 4, ..., 18 constituyen el espacio muestral del experimento aleatorio de "lanzar 3 dados y sumar los números obtenidos".

Ejemplo 5. Si se tiene una urna con N bolillas numeradas de 1 a N y se van sacando sucesivamente todas ellas, anotando ordenadamente los números que van saliendo; entonces las N! permutaciones que se pueden formar con los números 1 a N constituyen el espacio muestral del experimento aleatorio de "sacar todas las bolillas de la urna".

Ejemplo 6. Supongamos un punto que se mueve sobre una circunferencia y que puede detenerse en cualquier punto de la misma; los puntos de la circunferencia constituyen un espacio muestral, en este caso constituido por infinitos elementos.

14 Precisemos ahora la definición de los conjuntos de casos favorables. Desde luego éstos deben ser subconjuntos de E. Cuando E es finito, no hay inconveniente en considerar como posibles conjuntos favorables a todos los subconjuntos de E. Pero cuando E es infinito, la consideración de "todos" los subconjuntos de E puede presentar dificultades. Por otra parte, no es necesario considerar todos los subconjuntos del espacio muestral para poder construir la idea de probabilidad de acuerdo con los ejemplos del capítulo anterior. Lo importante es que si A y B representan conjuntos de casos favorables de ciertas proposiciones, la unión $A \cup B$ debe ser también un conjunto de casos favorables, precisamente el correspondiente a la proposición "A o B". También el conjunto complementario $\bar{A} = E - A$ debe figurar en la clase de conjuntos de casos favorables, puesto que corresponde a la proposición "no se cumple A".

Dentro de la matemática hay precisamente una estructura especial para clases de subconjuntos de un conjunto, llamada σ -álgebra (o *álgebra de Borel* o "*tribu*") que cumple dichas condiciones. Su definición es la siguiente:

Definición 2.2. Una σ -álgebra en un conjunto E es una clase B de subconjuntos A_i de E, tales que se cumplen las dos condiciones siguientes:

- i) Si $A_1, A_2, \dots \in B$, entonces $\cup A_i \in B$,
- ii) Si $A \in B$, entonces $\bar{A} = E - A \in B$.

Como es bien sabido, \cup significa la *unión* de conjuntos, y \cap la *intersección* de éstos. Además el conjunto vacío se representará por \emptyset .

Consecuencias. De la definición anterior se deduce:

a) $E = A \cup \bar{A} \in \mathcal{B}$;

b) $\emptyset = \bar{E} \in \mathcal{B}$;

c) Si $A_1, A_2 \in \mathcal{B} \Rightarrow \bar{A}_1, \bar{A}_2 \in \mathcal{B} \Rightarrow \bar{A}_1 \cup \bar{A}_2 = \overline{A_1 \cap A_2} \in \mathcal{B} \Rightarrow A_1 \cap A_2 \in \mathcal{B}$. En general, si $A_1, A_2, \dots \in \mathcal{B}$ entonces la intersección $\cap A_i \in \mathcal{B}$.

Obsérvese, pues, que no ha sido necesario postular que la intersección de conjuntos de una σ -álgebra pertenece a la misma; ello es una consecuencia de las propiedades i) y ii) de la definición.

A partir de esta definición, los clásicos casos favorables se incluyen dentro de los llamados "sucesos", cuya definición es la siguiente:

Definición 2.3. Los elementos de una σ -álgebra \mathcal{B} en un espacio muestral E , se llaman *sucesos*. Los elementos de E se llaman *sucesos elementales*.

Definición 2.4. Se dice que los sucesos A_1, A_2, \dots son *incompatibles* si no existe ningún elemento de E que pertenezca a dos o más de ellos, es decir, si $A_i \cap A_j = \emptyset$ para todo par de índices $i \neq j$.

Ejemplo 1. Al lanzar un dado hemos visto que el espacio muestral es $E = \{1, 2, 3, 4, 5, 6\}$. Un *suceso* puede ser el subconjunto $\{3, 4\}$ y corresponde al hecho de salir "3 ó 4", o sea a la unión de los sucesos elementales $\{3\}$ y $\{4\}$. Otro suceso es $\{1, 2, 5\}$. Ambos sucesos son incompatibles. En cambio no lo son los sucesos $\{3, 4\}$ y $\{1, 4, 6\}$ pues ambos tienen en común el elemento $\{4\}$.

Ejemplo 2. En el espacio muestral $E = \{3, 4, \dots, 18\}$ correspondiente al proceso de lanzar 3 dados y sumar los puntos obtenidos, un suceso es $\{3, 5, 7, 11, 13, 17\}$, que corresponde a la proposición "la suma es un número primo". Otro suceso es el formado por los números pares $\{4, 6, 8, 10, 12, 14, 16, 18\}$, y los dos sucesos son incompatibles.

Tenemos ahora todos los elementos necesarios para poder dar forma axiomática al concepto de probabilidad, tomando como modelo los ejemplos y definición del capítulo anterior, pero dándoles forma abstracta, como corresponde a una teoría matemática.

Sea E un espacio muestral y \mathcal{B} una σ -álgebra en E . Entonces se establece la definición siguiente.

Definición 2.5. Se llama *probabilidad* de un suceso $A \in \mathcal{B}$ a un número real $P(A)$, que cumple los siguientes axiomas:

I. $P(A) \geq 0$

II. Si A_1, A_2, \dots son sucesos incompatibles, entonces

$$P(\cup_i A_i) = \sum_i P(A_i)$$

III. $P(E) = 1$.

Obsérvese que el axioma II no es más que el principio de las probabilidades totales del capítulo 1 (extendido a un número finito o infinito de sucesos), teniendo en cuenta que la unión de varios conjuntos es el conjunto cuyos elementos pertenecen a uno o a otro de los conjuntos.

Toda terna (E, \mathcal{B}, P) que cumpla las condiciones anteriores, se llama un *espacio de probabilidad*. Si el espacio muestral es finito, el espacio de probabilidad se llama también finito. Si E es finito o infinito numerable, el espacio de probabilidad se llama *discreto*.

Convenio. Hemos dado la definición general de probabilidad, válida para espacios muestrales E cualesquiera. Sin embargo, en la presente monografía supondremos siempre, por simplicidad, lo siguiente:

a) Si E es finito o infinito numerable, la σ -álgebra \mathcal{B} será siempre el conjunto de todos los subconjuntos de E .

b) Si E es el conjunto de puntos de la recta (o del plano, o, en general, de \mathbb{R}^n) la σ -álgebra \mathcal{B} será la constituida por todos los conjuntos *medibles* de E . En particular, en el caso de la recta, \mathcal{B} será el conjunto de todos los intervalos abiertos, más los conjuntos formados por un solo punto y los conjuntos que puedan formarse por la unión de un número finito o de una infinidad numerable de estos conjuntos. No se considerarán otros conjuntos más generales.

De acuerdo con este convenio, en general en lo sucesivo no hará falta mencionar la σ -álgebra B , sino tan sólo E y P .

Ejemplo 1. Sea $E = \{a_1, a_2, \dots, a_6\}$ el espacio muestral correspondiente al lanzamiento de un dado. Los elementos a_i son los números $1, 2, \dots, 6$. La σ -álgebra B de los sucesos está constituida por todos los subconjuntos de E . Podemos definir $P(a_i) = 1/6$, para todo i , lo cual equivale a suponer que el dado es perfecto; si no fuera un cubo exacto o no estuviera construido de material homogéneo, la probabilidad de cada cara podría cambiar. Según el axioma II, si $i \neq j$, será $P(a_i \cup a_j) =$ probabilidad de que salga el número i o el $j = P(a_i) + P(a_j) = 2/6 = 1/3$. Análogamente, por el axioma II, si i, j, k son diferentes, es $P(a_i \cup a_j \cup a_k) = P(a_i) + P(a_j) + P(a_k) = 1/2$. Obsérvese que $P(\{a_3, a_4\} \cup \{a_1, a_4, a_6\})$ no es la suma de las probabilidades de los sucesos $\{a_3, a_4\}$ y $\{a_1, a_4, a_6\}$, puesto que no son incompatibles, por tener en común el elemento a_4 .

Ejemplo 2. Para el lector familiarizado con la noción de medida de conjuntos sobre la recta real (lo que no hará falta en lo sucesivo, pues nos limitaremos a conjuntos formados por la unión de segmentos disjuntos y entonces la medida es la suma de las longitudes de los mismos), podemos decir que, dado el segmento $(a, b), a < b$ y la σ -álgebra de los subconjuntos medibles del mismo, para cada uno de estos subconjuntos, por ejemplo α , se puede definir $P(\alpha) = \mu(\alpha)/(b - a)$, donde $\mu(\alpha)$ es la medida de α . Evidentemente, esta definición de probabilidad es admisible, por cumplirse los axiomas I, II, III. Vemos así la vinculación estrecha entre la noción de probabilidad y la de medida. Como $b - a = \mu((a, b))$, la probabilidad aparece como cociente de medidas. Para conjuntos de puntos del plano, del espacio o de \mathbb{R}^n puede establecerse la misma definición.

17

2. CONSECUENCIAS DE LOS AXIOMAS

A partir de los axiomas que definen la probabilidad y mediante la aplicación de simples relaciones de la teoría de conjuntos, se pueden deducir las siguientes consecuencias:

1. $P(\emptyset) = 0$. En efecto, aplicando el axioma II a los conjuntos \emptyset y E , resulta $P(\emptyset \cup E) = P(\emptyset) + P(E)$. Pero $\emptyset \cup E = E$ y $P(E) = 1$ (axioma III). Por tanto, queda $1 = P(\emptyset) + 1$, de donde $P(\emptyset) = 0$.

2. Llamando $\bar{A} = E - A =$ suceso *opuesto* al A , es

$$P(\bar{A}) = 1 - P(A). \quad [2.1]$$

En efecto, aplicando el axioma II a $A \cup \bar{A} = E$, se tiene $P(A) + P(\bar{A}) = P(E)$, y por el axioma III, resulta [2.1].

3. Si $A_1 \subset A_2$, entonces $P(A_1) \leq P(A_2)$.

En efecto, es $A_2 = A_1 \cup (A_2 \cap \bar{A}_1)$, y por los axiomas II y I, resulta $P(A_2) = P(A_1) + P(A_2 \cap \bar{A}_1) \geq P(A_1)$.

4. Cualquiera que sea A es $P(A) \leq 1$. Consecuencia de ser $A \subset E$ y de la consecuencia anterior.

5. Si A y B son sucesos compatibles, o sea $A \cap B = C \neq \emptyset$, entonces

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad [2.2]$$

Demostración. Observemos las identidades

$$A = (A \cap B) \cup (A \cap \bar{B}), \quad B = (A \cap B) \cup (\bar{A} \cap B)$$

de las cuales, por el axioma II, se deduce

$$P(A) = P(A \cap B) + P(A \cap \bar{B}), \quad P(B) = P(A \cap B) + P(\bar{A} \cap B). \quad [2.3]$$

Por otra parte, la identidad

$$A \cup B = (A \cap B) \cup (A \cap \bar{B}) \cup (\bar{A} \cap B)$$

nos da

$$P(A \cup B) = P(A \cap B) + P(A \cap \bar{B}) + P(\bar{A} \cap B) \quad [2.4]$$

y de [2.3] y [2.4] resulta inmediatamente la relación [2.2] que se quería demostrar.

Esta relación se puede generalizar. Pongamos $A = A_1$, $B = A_2 \cup A_3$. Resulta

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2 \cup A_3) - P(A_1 \cap (A_2 \cup A_3)) = \\ &= P(A_1) + P(A_2) + P(A_3) - P(A_2 \cap A_3) - P((A_1 \cap A_2) \cup (A_1 \cap A_3)) = \\ &= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + \\ &+ P(A_1 \cap A_2 \cap A_3). \end{aligned}$$

En general, por inducción, resulta la fórmula

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_n) &= \sum_1 P(A_i) - \sum_{i,j} P(A_i \cap A_j) + \quad [2.5] \\ &+ \sum_{i,j,k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n) \end{aligned}$$

donde las sumas se extienden en cada caso a todas las combinaciones posibles entre los diferentes índices i, j, k, \dots .

3. LA DEFINICION CLASICA

Veamos cómo los axiomas I, II, III y sus consecuencias se relacionan con la definición clásica de Laplace dada en el capítulo 1. Supongamos que el espacio muestral se compone de un número finito de sucesos A_1, A_2, \dots, A_n , o sea, $E = A_1 \cup A_2 \cup \dots \cup A_n$, que sean incompatibles entre sí, o sea, $A_i \cap A_j = \emptyset$ para todo $i \neq j$. Entonces, según los axiomas II y III,

$$P(E) = P(A_1) + P(A_2) + \dots + P(A_n) = 1. \quad [2.6]$$

Si suponemos, además, que todos los sucesos A_i tienen la misma probabilidad (que es lo que quiere decir la frase "igualmente posibles" de la definición de Laplace), tenemos $P(A_1) = P(A_2) = \dots = P(A_n)$ y, por tanto, según [2.6] será $P(A_i) = 1/n$. Sea ahora un suceso A formado por la unión de m sucesos A_i , o sea, $A = A_{i_1} \cup A_{i_2} \cup \dots \cup A_{i_m}$. La probabilidad de A será

$$P(A) = m P(A_i) = m/n.$$

Se tiene así el resultado coincidente con la definición clásica.

4. PROBABILIDAD CONDICIONAL

Queremos ahora incluir en el esquema axiomático el principio de las probabilidades compuestas mencionadas en el capítulo 1. Allí nos referíamos a sucesos "independientes", aunque quedaba un poco impreciso el significado de estas palabras, advirtiendo que la fórmula [1.2] perdía su validez si los sucesos no eran independientes. Vamos a ver ahora cómo se pueden tratar estas ideas dentro de la teoría axiomática.

Sea (E, \mathcal{B}, P) un espacio de probabilidad compuesto del espacio muestral E , la σ -álgebra \mathcal{B} y la probabilidad P . Sea $A \in \mathcal{B}$ un suceso tal que $P(A) \neq 0$.

Definición 2.5. Se llama *probabilidad condicional* de un suceso $B \in \mathcal{B}$, dado el suceso $A \in \mathcal{B}$, y se representa por $P(B|A)$, al cociente

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \quad [2.7]$$

Hay que demostrar que esta definición es admisible, es decir, que la probabilidad condicional $P_A(B) = P(B|A)$, definida para todo $B \in \mathcal{B}$, una vez fijado $A \in \mathcal{B}$, cumple con los axiomas I, II y III de toda probabilidad. En otras palabras, hay que probar que (E, \mathcal{B}, P_A) es otro espacio de probabilidad.

El axioma I dice que debe ser $P(B|A) \geq 0$, lo cual es inmediato por ser $P(B \cap A) \geq 0$ y $P(A) > 0$. El axioma II se cumple por ser

$$\begin{aligned} P_A(\cup B_i) &= \frac{P(A \cap (\cup B_i))}{P(A)} = \frac{P(\cup (A \cap B_i))}{P(A)} = \\ &= \frac{\sum_i P(A \cap B_i)}{P(A)} = \sum_i P(B_i|A) = \sum_i P_A(B_i), \end{aligned}$$

para toda sucesión B_1, B_2, \dots de sucesos incompatibles.

El axioma III se cumple por ser

$$P_A(E) = \frac{P(A \cap E)}{P(A)} = \frac{P(A)}{P(A)} = 1.$$

Por tanto la definición 2.5 es admisible. A partir de ella se puede escribir

$$P(A \cap B) = P(A) P(B|A).$$

Análogamente, fijado el suceso B y supuesto $P(B) \neq 0$, se tiene la probabilidad condicional

$$P(A|B) = \frac{P(B \cap A)}{P(B)}. \quad [2.8]$$

Obsérvese que la ecuación [2.6] no se demuestra ni es ningún axioma: es una definición, y se complementa con la siguiente definición.

Definición 2.6. Se dice que dos sucesos son *independientes* si se verifica que

$$P(B|A) = P(B). \quad [2.9]$$

Obsérvese que en este caso, según [2.7], es

$$P(A \cap B) = P(A) \cdot P(B) \quad (A \text{ y } B \text{ independientes}) \quad [2.10]$$

y por tanto, comparando con [2.9], es también

$$P(A|B) = P(A). \quad [2.11]$$

Esto nos dice que la independencia es una propiedad simétrica, que puede definirse por cualquiera de las igualdades [2.9] ó [2.11].

Puesto que el hecho de que un elemento pertenezca a la intersección $A \cap B$ significa que a la vez pertenece a A "y" a B, la relación [2.10] equivale a la [1.2] (principio de las probabilidades compuestas). Nótese, sin embargo, que si bien las definiciones 2.5 y 2.6 tienen su origen en la idea intuitiva de sucesos independientes, al ser ahora establecidas como definiciones desaparece toda posible ambigüedad, y adquieren un sentido bien preciso.

La definición de independencia se generaliza a más de dos sucesos de la manera siguiente:

Definición 2.7. Se dice que varios sucesos A_1, A_2, \dots, A_n son *independientes* (o *completamente independientes*) si se verifica

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_k}) \quad [2.12]$$

para $k = 2, 3, \dots, n$, donde (i_1, i_2, \dots, i_k) es una combinación cualquiera de los n números $1, 2, \dots, n$.

21

Por ejemplo, para que 3 sucesos A, B, C sean independientes, se debe cumplir

$$P(A \cap B) = P(A) P(B),$$

$$P(A \cap C) = P(A) P(C),$$

$$P(B \cap C) = P(B) P(C),$$

$$P(A \cap B \cap C) = P(A) P(B) P(C).$$

Obsérvese que la última relación no puede deducirse de las tres primeras, es decir, las cuatro condiciones son necesarias.

Naturalmente, no hay que confundir la independencia con la incompatibilidad. Según las definiciones 2.4 y 2.6 se tiene: Dos sucesos A y B son *independientes* si se verifica

$$P(A \cap B) = P(A) P(B) \quad [2.13]$$

y son *incompatibles* si $A \cap B = \phi$.

5. EJEMPLOS

Vamos a resolver algunos problemas clásicos que resultan de la aplicación directa de la definición de probabilidad y de sus primeras propiedades. Son problemas del mismo tipo que los estudiados en el capítulo 1, pero ahora se pueden puntualizar más las etapas de la solución. Son problemas más bien de interés histórico, pero que tienen importancia por ser representativos de un amplio campo de problemas análogos.

Problema 2.1. *Se tiran dos dados al azar y se quiere hallar la probabilidad de que la suma de los puntos obtenidos sea igual a 10.*

Solución. El espacio muestral es el conjunto de los 36 pares ordenados (i, j) , donde i puede tomar los valores $1, 2, \dots, 6$ del primer dado y j los valores $1, 2, \dots, 6$ del segundo. Siempre que se trate de un problema con dados y no se indique lo contrario, se supone que el dado es correcto y por tanto la probabilidad de que salga cada cara es la misma. En consecuencia, la probabilidad de cada par (i, j) será también la misma, cualquiera que sea el par, y según el No. 3, esta probabilidad será $1/36$, puesto que hay 36 casos posibles y la suma de las probabilidades debe ser 1. El problema consiste en hallar la probabilidad del suceso $A = \{(i, j); i + j = 10\}$, o sea, la probabilidad del suceso formado por los pares $(4, 6), (5, 5), (6, 4)$. Tratándose de 3 pares, todos con la misma probabilidad, por el axioma II, será $P(A) = 3/36 = 1/12$.

22

Problema 2.2. (Problema de las coincidencias). *Se tienen dos urnas con n bolillas cada una, numeradas de 1 a n . Se va sacando simultáneamente 1 bolilla de cada urna, y se quiere hallar la probabilidad de que, al terminar la extracción de todas las bolillas, se haya extraído, por lo menos una vez, el mismo número de cada urna.*

Solución. El espacio muestral E está formado por todas las matrices de la forma

$$\begin{pmatrix} i_1 & i_2 & \dots & i_n \\ j_1 & j_2 & \dots & j_n \end{pmatrix} \quad [2.14]$$

donde i_1, i_2, \dots, i_n son los números entre 1 y n que salen de la primera urna, y j_1, j_2, \dots, j_n son los que salen de la segunda. El número total de elementos de E es $(n!)^2$ (casos posibles) y la probabilidad de cada uno será $(n!)^{-2}$. Sea A_i el conjunto de elementos de E , en los cuales el número i de la primera fila coincide con el número $j = i$ de la segunda, independientemente del lugar en que

ocurra la coincidencia. Por ejemplo, A_3 es el conjunto de las matrices de la forma

$$\begin{pmatrix} \dots & 3 & \dots \\ \dots & 3 & \dots \end{pmatrix}.$$

Se trata de calcular $P(A_1 \cup A_2 \cup \dots \cup A_n)$. Para ello se aplica la fórmula [2.5], que reduce el problema al cálculo de las probabilidades del segundo miembro. Vamos a buscar el valor de cada sumando:

a) Probabilidad $P(A_1)$. El número de elementos [2.14] en que coinciden los números i , se calcula de la siguiente manera. Fijado el lugar en que ocurre la coincidencia, los restantes $n-1$ números de la primera y de la segunda filas pueden ser cualesquiera, y por tanto se tienen $((n-1)!)^2$ casos. Como el lugar de la coincidencia puede ser también cualquiera, se tienen, en cada caso, n posibilidades más. Por tanto, A_1 está compuesto de $n((n-1)!)^2$ elementos [2.14], y en consecuencia

$$P(A_1) = \frac{n((n-1)!)^2}{(n!)^2} = \frac{1}{n}, \quad \sum_1 P(A_i) = \frac{n}{n} = 1.$$

23

b) Probabilidad $P(A_i \cap A_j)$. Fijados los lugares en que ocurren las coincidencias de los números i y j , los $n-2$ números restantes pueden ser cualesquiera; por tanto quedan $((n-2)!)^2$ casos. Como los lugares de las coincidencias pueden ser cualesquiera de los n , resultan, para cada caso, $n(n-1)$ posibilidades más; en total, $A_i \cap A_j$ está compuesto de $((n-2)!)^2 n(n-1)$ casos y por tanto

$$\begin{aligned} P(A_i \cap A_j) &= \frac{n(n-1)((n-2)!)^2}{(n!)^2} = \frac{1}{n(n-1)}, \quad \sum_{i,j} P(A_i \cap A_j) = \\ &= \binom{n}{2} \frac{1}{n(n-1)} = \frac{1}{2}. \end{aligned}$$

c) Análogamente,

$$P(A_1 \cap A_j \cap A_k) = \frac{((n-3)!)^2 n(n-1)(n-2)}{(n!)^2} = \frac{1}{n(n-1)(n-2)}$$

y por tanto

$$\sum_{i,j,k} P(A_i \cap A_j \cap A_k) = \binom{n}{3} \frac{1}{n(n-1)(n-2)} = \frac{1}{3!}.$$

Se tiene así como resultado final (procediendo sucesivamente) que la probabilidad de por lo menos una coincidencia es

$$P = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} \dots + (-1)^{n-1} \frac{1}{n!}. \quad [2.15]$$

Para $n \rightarrow \infty$, $P = 1 - e^{-1} = 0,6321\dots$ (Véase el Apéndice II). Es curioso que para valores relativamente pequeños de n se obtienen ya valores muy próximos a este valor límite. Así, se tiene

$$\begin{aligned} P(1) &= 1, & P(2) &= 0,5, & P(3) &= 0,666\dots, & P(4) &= 0,625, \\ P(5) &= 0,6333\dots, & P(6) &= 0,6319\dots, & P(7) &= 0,6321\dots \end{aligned}$$

y para valores mayores de 7, quedan invariables las cuatro primeras cifras decimales.

Este problema se encuentra por primera vez en el libro de P. R. Montmort, titulado *Essai d'Analyse sur les jeux de Hasard*, publicado en 1708, y presenta muchas variantes. Por ejemplo:

24

a) *Dos jugadores tienen una baraja de 40 cartas cada uno y van sacando simultáneamente las cartas una a una. ¿Cuál es la probabilidad de que, al terminar las cartas, los dos jugadores hayan sacado la misma carta por lo menos una vez?*

Evidentemente, la probabilidad es la misma de antes $P(40) = 0,6321\dots$. Ello quiere decir que repitiendo el experimento un gran número de veces, por término medio habrá coincidencia 6 veces de cada 10. En cuanto sea $n > 2$, es más probable de que haya coincidencia de que no la haya.

b) *Se tiene un cierto número de cartas y sus correspondientes sobres. Se mezclan unas y otros, y se va metiendo una carta cualquiera dentro de cada sobre. ¿Cuál es la probabilidad de que una carta por lo menos se haya metido en su correspondiente sobre? El resultado está dado siempre por la fórmula [2.15] y, como vimos, da siempre prácticamente el mismo valor 0,63... en cuanto sea $n \geq 5$.*

Problema 2.3. (Problema de los nacimientos). *En una reunión de r personas, ¿cuál es la probabilidad de que, por lo menos dos de ellas, cumplan años el mismo día?*

Solución. Prescindimos de la posibilidad de que alguien haya nacido el 29 de febrero y por tanto suponemos que el año tiene 365 días. El espacio muestral se compone de los posibles conjuntos

de r fechas y tiene por tanto 365^r elementos, todos con la misma probabilidad 365^{-r} . En vez del suceso cuya probabilidad se busca, consideremos el opuesto: el de que ningún par de personas cumpla años el mismo día. El número de elementos de este suceso se calcula así: la primera persona tiene 365 posibilidades; la segunda, no habiendo nacido el mismo día que la primera, tiene 364 posibilidades; la tercera persona tiene 363 posibilidades y así sucesivamente, la última persona tiene $365 - (r - 1)$ posibilidades. En total, el suceso opuesto consta de $365 \cdot 364 \cdot 363 \dots (365 - (r - 1))$ elementos. Su probabilidad es igual a este número dividido por 365^r , y la probabilidad del suceso objeto del problema será (suponiendo $r > 1$)

$$P_r = 1 - \frac{365 \cdot 364 \dots (365 - r + 1)}{365^r}.$$

Este número no es fácil de calcular directamente. Los siguientes valores dan una idea de su comportamiento:

No. de personas	$r =$	5	10	20	23	30	40	60
Probabilidad	$P_r =$	0,027	0,117	0,411	0,507	0,706	0,89	0,99.

Se ha señalado el número $r = 23$, pues en tal caso la probabilidad es prácticamente $1/2$. Obsérvese que si las personas son 60 o más, la probabilidad es superior al 99%, es decir hay casi certeza de que por lo menos dos personas cumplan años el mismo día, resultado un tanto sorprendente a primera vista.

Problema 2.4. (Esquema de contagio de Polya). *Supongamos una urna con n bolillas negras y r rojas. Se saca una bolilla y se devuelve a la urna junto con otras c bolillas del mismo color. Se desea hallar la probabilidad de que en $N = k_1 + k_2$ extracciones salgan k_1 bolillas negras y k_2 bolillas rojas.*

Solución. Consideremos primero la probabilidad de que las bolillas negras salgan justo en las k_1 primeras extracciones. La probabilidad de bolilla negra en la primera extracción es $n/(n+r)$ y la probabilidad de que sea negra en la segunda extracción, condicionada a que haya sido negra la primera, es $(n+c)/(n+r+c)$. Por tanto, según [2.7] la probabilidad de dos negras sucesivas es

$$\frac{n}{n+r} \cdot \frac{n+c}{n+r+c}.$$

Si las dos primeras han salido negras, en el momento de hacer la tercera extracción la urna contiene $n + 2c$ bolillas negras y r

rojas, luego la probabilidad de que la tercera sea también negra es $(n + 2c)/(n + r + 2c)$, y por tanto la probabilidad de que las tres primeras sean negras es

$$\frac{n}{n+r} \cdot \frac{n+c}{n+r+c} \cdot \frac{n+2c}{n+r+2c}$$

Prosiguiendo este razonamiento se llega a la extracción k_1 -ésima, con la probabilidad

$$\frac{n}{n+r} \cdot \frac{n+c}{n+r+c} \dots \frac{n+(k_1-1)c}{n+r+(k_1-1)c}$$

de que todas las bolillas hayan salido negras. La probabilidad de que la siguiente sea roja es $r/(n+r+k_1c)$. Se añaden luego c bolillas rojas, con lo cual la probabilidad de una nueva extracción roja es $(r+c)/(n+r+(k_1+1)c)$. Procediendo sucesivamente resulta que la probabilidad de que las k_1 primeras extracciones sean negras y las k_2 siguientes sean rojas es

$$\frac{n(n+c) \dots (n+(k_1-1)c) r(r+c) \dots (r+(k_2-1)c)}{(n+r)(n+r+c) \dots (n+r+(N-1)c)}$$

26

donde $N = k_1 + k_2$.

Si se da otro orden cualquiera en que salgan k_1 bolillas negras y k_2 rojas, la probabilidad es siempre la misma (como se ve de inmediato por el mismo cálculo anterior). Por tanto, la probabilidad buscada de que salgan exactamente k_1 bolillas negras y k_2 rojas, independientemente del orden, será igual a la probabilidad anterior multiplicada por el número de combinaciones de k_1 (o k_2) objetos tomados entre N de ellos. En definitiva, el resultado es

$$P = \binom{N}{k_1} \frac{n(n+c) \dots (n+(k_1-1)c) r(r+c) \dots (r+(k_2-1)c)}{(n+r)(n+r+c) \dots (n+r+(N-1)c)}$$

Problema 2.5. *Se tiene un conjunto formado por m objetos de una misma clase M y por r objetos de una misma clase R. Se eligen al azar s objetos. Se busca la probabilidad de que entre ellos haya x de la clase M y $s-x$ de la clase R.*

Solución. Es un problema típico de combinatoria. El espacio E es el conjunto de las combinaciones de s objetos entre $m+r = n$ objetos dados. Su número es $\binom{n}{s}$ y cada uno de estos sucesos elementales tiene la misma probabilidad $\left(\frac{n}{s}\right)^{-1}$. Se trata de hallar la probabilidad del subconjunto A, cuyos elementos consten de x

objetos de la clase M y $s-x$ de la clase R . Las posibles combinaciones de x objetos de la clase M son $\binom{m}{x}$ y las posibles combinaciones de $s-x$ objetos de la clase R son $\binom{r}{s-x}$ por tanto, el número de elementos de A es $\binom{m}{x}\binom{r}{s-x}$, y la probabilidad buscada

$$P = \frac{\binom{m}{x}\binom{r}{s-x}}{\binom{n}{s}}.$$

Por ejemplo, si de una clase de 30 alumnos, compuesta de 20 varones y 10 mujeres, se eligen 5 alumnos al azar, ¿cuál es la probabilidad de que entre ellos haya exactamente 2 mujeres?

Aplicando la fórmula última donde $m = 20$, $r = 10$, $s = 5$ y $x = 3$ resulta: $P = 2850/7917 = 0,36\dots$

Problema 2.6. *Se desea hallar:* a) *la probabilidad de sacar por lo menos una vez el 6 al arrojar un dado 4 veces;* b) *la probabilidad de sacar dos 6 por lo menos una vez al lanzar dos dados 24 veces.*

Solución. a) La probabilidad de no sacar el 6 en una jugada es $5/6$ y la de no sacarlo vez alguna en 4 jugadas será $(5/6)^4$, puesto que cada jugada es independiente de las demás. Por tanto la probabilidad buscada es $P_1 = 1 - (5/6)^4 = 671/1296 = 0,517\dots$

b) La probabilidad de no sacar el 6 dos veces en una jugada es $35/36$, y la de no sacarlo vez alguna en 24 jugadas será $(35/36)^{24}$. Luego la probabilidad de sacar el 6 dos veces por lo menos una vez será

$$P_2 = 1 - (35/36)^{24} = 0,49\dots$$

El problema tiene interés histórico, y se considera como uno de los que dieron origen al cálculo de probabilidades. Según figura en las *Oeuvres* de Fermat, página 296 (Carta de B. Pascal a P. de Fermat, del 29 de julio de 1654), el problema fue propuesto a Pascal por un tal Caballero de Meré, quien no se explicaba que, en el caso a), la probabilidad fuera mayor que $1/2$ y en el caso b) fuera menor que $1/2$, siendo así que 4 es a 6 (razón del número de tiradas al número de posibilidades en el primer caso) como 24 es a 36 (razón del número de tiradas al número de posibilidades en el segundo caso). Locurioso es que el Caballero de Meré llegaba a esta conclusión, para él paradójica, por experiencia de jugador, a pesar de ser los resultados muy próximos a

1/2. Esto prueba tanto la precisión a que se puede llegar sin cálculo, con un fino espíritu de observación, como la extrema coincidencia de la probabilidad calculada *a priori* con la frecuencia obtenida mediante repetidos ensayos.

Problema 2.7. Si se lanzan 3 dados al azar, ¿cuál es la probabilidad de que la suma de los números obtenidos sea, respectivamente, 9, 10, 11 ó 12?

Solución. El espacio muestral está formado por las ternas (i, j, k) con $1 \leq i, j, k \leq 6$. Como todas ellas tienen igual probabilidad, resulta que ésta es $(1/6)^3 = 1/216$. Consideremos el suceso A_9 formado por las ternas tales que $i + j + k = 9$. Las ternas que lo constituyen son

(1, 2, 6), (2, 6, 1), (6, 1, 2), (2, 1, 6), (1, 6, 2), (6, 2, 1),
 (1, 3, 5), (3, 5, 1), (1, 5, 3), (5, 1, 3), (5, 3, 1), (3, 1, 5),
 (1, 4, 4), (4, 1, 4), (4, 4, 1), (2, 2, 5), (2, 5, 2), (5, 2, 2),
 (2, 3, 4), (3, 4, 2), (4, 2, 3), (2, 4, 3), (4, 3, 2), (3, 2, 4), (3, 3, 3).

Tenemos, en total, 25 ternas. Por tanto

$$P(A_9) = \frac{25}{216} = 0,1157\dots$$

De modo análogo, formando directamente las ternas, cuya suma es 10, 11 ó 12 resulta

$$P(A_{10}) = P(A_{11}) = \frac{27}{216} = 0,125\dots, \quad P(A_{12}) = P(A_9) = \\ = \frac{25}{216} = 0,1157.$$

También este problema tiene interés histórico. Figura en *Le Opere di Galileo Galilei*, vol. XIV, Firenze 1855, bajo el título *Considerazione sopra il giuoco dei dadi*, y fue propuesto a Galileo (1564-1642) por un jugador que se extrañaba de que la probabilidad fuera distinta, siendo así que los cuatro números 9, 10, 11 y 12 admiten el mismo número de descomposiciones en tres sumandos, a saber:

$$9 = 1 + 2 + 6 = 1 + 3 + 5 = 1 + 4 + 4 = 2 + 2 + 5 = 2 + 3 + 4 = 3 + 3 + 3, \\
 10 = 1 + 3 + 6 = 1 + 4 + 5 = 2 + 2 + 6 = 2 + 3 + 5 = 2 + 4 + 4 = 3 + 3 + 4, \\
 11 = 1 + 4 + 6 = 1 + 5 + 5 = 2 + 3 + 6 = 2 + 4 + 5 = 3 + 3 + 5 = 3 + 4 + 4, \\
 12 = 1 + 5 + 6 = 2 + 4 + 6 = 2 + 5 + 5 = 3 + 3 + 6 = 3 + 4 + 5 = 4 + 4 + 4.$$

La explicación está, como hemos visto, en que las descomposiciones no tienen la misma probabilidad, puesto que por permutación de los sumandos pueden provenir de ternas diferentes.

Problema 2.8. *Una lotería tiene N números y un solo premio. Un jugador compra n billetes de un solo sorteo y otro jugador compra un solo billete durante n sorteos consecutivos, de manera que los dos jugadores apuestan la misma cantidad. ¿Cuál tiene mayor probabilidad de sacar el premio?*

Solución. El espacio muestral está formado por los N números de la lotería. Al decir que se trata de una lotería, se sobrentiende que todos los números tienen la misma probabilidad de ser premiados, y por tanto esta probabilidad es $1/N$. El primer jugador, al comprar n números, tendrá una probabilidad de ganar el premio igual a $P_1 = n/N$. En el caso del segundo jugador se razona de la siguiente manera. La probabilidad de no sacar el premio en el primer sorteo es $(N - 1)/N$; la probabilidad de no sacar el premio ni en el primero ni en el segundo sorteo será $((N - 1)/N)^2$, y la probabilidad de no sacar ninguna vez el premio en n jugadas será $((N - 1)/N)^n$. Por tanto, la probabilidad de sacar el premio por lo menos una vez será

$$P_2 = 1 - (1 - \frac{1}{N})^n.$$

Obsérvese ahora que para todo par de enteros $N, m > 2$, es

$$(1 - \frac{1}{N})^n = (1 - \frac{1}{N})^{n-1} - (1 - \frac{1}{N})^{n-1} \cdot \frac{1}{N} > (1 - \frac{1}{N})^{n-1} - \frac{1}{N}.$$

Escribiendo esta desigualdad para $m = n, n - 1, \dots, 3, 2$ y sumando miembro a miembro, resulta

$$(1 - \frac{1}{N})^n > 1 - \frac{n}{N}$$

de lo cual se deduce $P_2 < P_1$, es decir, el primer jugador tiene una probabilidad mayor de sacar el premio. El resultado es natural. En efecto, el primer jugador sólo puede sacar el premio una sola vez, en tanto que el segundo, exponiendo el mismo dinero, tiene la posibilidad de ganarlo varias veces. Es natural, por tanto, que la probabilidad de sacarlo "por lo menos una vez" sea menor para este último que para el primero.

Problema 2.9. *Un tirador tiene la probabilidad p de dar en el blanco. Se le ofrecen dos alternativas: a) Hacer un solo disparo; b) hacer 3 disparos con la condición de dar por lo menos 2 veces en el blanco. ¿Cuál alternativa es más favorable al tirador?*

Solución. La probabilidad de dar 2 veces en el blanco en 3 disparos es $3p^2(1-p)$, puesto que $p^2(1-p)$ es la probabilidad de dar en el blanco 2 veces especificadas y errar la tercera. La probabilidad de dar las 3 veces en el blanco es p^3 . Por tanto, la probabilidad de dar por lo menos 2 veces en el blanco en 3 disparos es $p^3 + 3p^2(1-p)$. Para que la segunda alternativa sea más ventajosa que la primera debe ser, por tanto, $p^3 + 3p^2(1-p) > p$, o sea, $2p^2 - 3p + 1 < 0$, de donde $1/2 < p < 1$. Por tanto, si $p > 1/2$, es ventajosa la segunda alternativa, pero si $p < 1/2$ es preferible la primera. Si $p = 1/2$ las dos alternativas son equivalentes

Problema 2.10. *Se lanzan simultáneamente al azar 5 dados ¿Cuál es la probabilidad de que salgan los números 1, 2, 3, 4, 5?*

Solución. Si se especificara cuál de los dados debe dar el número 1, cuál el número 2, etcétera, la probabilidad sería el producto de las probabilidades respectivas, o sea, $(1/6)^5 = 1/7776$. Pero como el dado que sale 1 puede ser cualquiera (5 posibilidades), el que sale 2 puede ser cualquiera de los 4 restantes (4 posibilidades), el que sale 3, cualquiera de los 3 restantes (3 posibilidades), etcétera, resulta que hay $5 \cdot 4 \cdot 3 \cdot 2 = 120$ posibilidades más, y por tanto la probabilidad buscada es $120/7776 = 5/324 = 0,015$.

Otros Problemas. Damos a continuación otros ejemplos simples de problemas, de los que se indica únicamente la solución.

1. *Se lanzan 2 dados, hállese la probabilidad de que la diferencia entre el mayor y el menor de los números respectivos sea ≥ 3 .*

Solución. Por cómputo directo resulta $P = 12/36 = 1/3$.

2. *Se lanzan 2 dados, hállese la probabilidad de que la suma de los números obtenidos sea 7 ó 10.*

Solución. En el problema 2.1 se vio que $P(10) = 1/12$. Por cálculo directo resulta análogamente $P(7) = 6/36 = 1/6$. Por tanto, la probabilidad pedida es $P = 1/12 + 1/6 = 1/4$.

3. *Se lanzan 2 dados, hállese la probabilidad de que la suma de los números que salgan sea igual a 7 y, al mismo tiempo, la diferencia entre el mayor y el menor sea igual a 1.*

Solución. Si A es el suceso "sacar la suma 7" y B el suceso "diferencia igual a 1" es $P(A) = 1/6$ (problema anterior) y por cálculo directo resulta $P(B|A) = 1/3$. Por tanto, $P(A \cap B) =$

$= P(A) P(B|A) = (1/6)(1/3) = 1/18$. Obsérvese que el hecho de que $P(B|A) \neq P(B)$ prueba que los sucesos A y B no son independientes.

4. Se lanza un dado rojo y otro blanco, hállese la probabilidad de que el número del rojo sea mayor que el del blanco.

Solución. Contando los sucesos elementales que cumplen con el enunciado y el total de sucesos elementales resulta $P = 15/36$.

5. Una urna contiene 4 bolillas blancas y 8 rojas y se sacan dos bolillas sucesivamente, sin reposición de la primera sacada. Hállese: a) la probabilidad de que las dos bolillas sean blancas; b) la probabilidad de que una sea blanca y la otra roja.

Solución. a) $P_a = 1/11$; b) $P_b = 16/33$.

6. Se lanzan al azar n monedas, hállese la probabilidad de que por lo menos $n - 1$ de ellas salgan cara o salgan cruz.

Solución. La probabilidad de que n salgan cara es $(1/2)^n$ y la de que $n - 1$ salgan cara es $n(1/2)^n$. Luego la probabilidad de por lo menos $n - 1$ salgan cara es la suma $(n + 1)/2^n$. Análogamente, la probabilidad de por lo menos salgan $n - 1$ cruz es también $(n + 1)/2^n$. La probabilidad buscada es, portanto, la suma de las dos, o sea, $(n + 1)/2^{n-1}$. Esto supone $n > 2$. Para $n = 2$ la probabilidad es evidentemente igual a 1.

7. Tres amigos han acordado, si las circunstancias se lo permiten, acudir a una determinada cita. La probabilidad de cada uno de poder ir es p . ¿Cuál es la probabilidad de que acudan dos amigos y falte el tercero?

Solución. $P = 3p^2(1 - p)$.

8. La urna A contiene x bolillas blancas e y bolillas negras. La urna B contiene x' bolillas blancas e y' bolillas negras. Se saca una bolilla de cada urna, ¿cuál es la probabilidad de que ambas bolillas sean del mismo color?

Solución. $p = (xx' + yy')/(x + y)(x' + y')$.

9. En un tribunal de tres personas A, B y C, las probabilidades respectivas de dar un fallo justo son P_A , P_B y P_C . Se desea hallar la probabilidad de que el fallo del tribunal sea justo, sabiendo que el mismo se toma por mayoría de votos.

Solución. $p = p_A p_B + p_A p_C + p_B p_C - 2 p_A p_B p_C.$

10. Una determinada travesía puede hacerse con aviones bimotores o con cuádrimotores. Los bimotores pueden volar con un solo motor y los cuádrimotores sólo con dos. La probabilidad de que un motor falle durante la travesía es p . ¿Cuáles aviones son los más seguros?

Solución. La probabilidad de que un bimotor no termine la travesía es p^2 y la de que no termine la travesía un cuádrimotor, $p^4 + 4p^3(1-p)$. Por tanto, suponiendo $p < 1$, los bimotores son más seguros si $p > 1/3$, y son más seguros los cuádrimotores en caso contrario.

11. Un avión trimotor puede volar con sólo el motor central o bien con los dos motores laterales. En una determinada travesía, la probabilidad de fallar el motor central es p_0 y la de fallar cada uno de los motores laterales es p_1 . ¿Cuál es la probabilidad de que el avión no pueda terminar la travesía?

Solución. $p = p_0 p_1 (2 - p_1).$

3

VARIABLES ALEATORIAS, FUNCIONES DE PROBABILIDAD

1. VARIABLES ALEATORIAS

Los elementos del espaciomuestral E , que hemos llamado sucesos elementales, son elementos abstractos y, en consecuencia, también lo son los sucesos o elementos de la σ -álgebra \mathbf{B} definida en E . La probabilidad P , por otra parte, es una función cuyo dominio es el conjunto de los sucesos y cuyo codominio es el intervalo $[0, 1]$ de los números reales. Para poder aplicar el cálculo matemático, es conveniente que el dominio de la función P pertenezca también al conjunto de los números reales R , a fin de que P sea una función de números reales en números reales.

Spongamos, por ejemplo, el experimento aleatorio de lanzar dos monedas y ver si éstas salen cara C o cruz F . El espacio muestral E es el conjunto de los 4 pares:

$$(C, C), (C, F), (F, C), (F, F). \quad [3.1]$$

Si sólo interesa el número de caras, de manera que los dos pares intermedios puedan considerarse equivalentes, podemos introducir la función $X: E \rightarrow R$ definida por $X =$ número de caras, o sea,

$$X(C, C) = 2, \quad X(C, F) = X(F, C) = 1, \quad X(F, F) = 0.$$

Tenemos así un nuevo conjunto $\{2, 1, 0\}$, ahora formado por números reales, a cada uno de los cuales corresponde una probabilidad. Así

$$P(2) = P(C, C) = 1/4, \quad P(1) = P(C, F) + P(F, C) = 1/2, \quad P(0) = \\ = P(F, F) = 1/4.$$

Vemos, pues, que la función X hace corresponder a cada elemento de E un número real y que, además, el conjunto de elementos de E , cuya imagen es uno de estos números reales, es un elemento de \mathbf{B} , o sea, es un suceso, y tiene, por tanto, una determinada probabilidad.

Las funciones $X: E \rightarrow R$ que cumplen estas condiciones se llaman *variables aleatorias*. La palabra *variable* indica que la función puede tomar diversos valores (en el ejemplo anterior 2, 1, 0) y la palabra *aleatoria* indica que estos valores provienen de un experimento aleatorio (en el ejemplo anterior lanzar dos monedas) y, por tanto, a cada uno de sus valores corresponde una determinada probabilidad. En realidad sería más apropiado llamarlas *funciones aleatorias*, pero el uso ha sancionado la primera denominación.

En este capítulo 3 vamos a limitarnos a variables aleatorias que pueden tomar únicamente un número finito de valores. Se llaman variables aleatorias finitas. Su definición precisa es la siguiente:

Definición 3.1. Dado un espacio de probabilidad (E, \mathbf{B}, P) , se llama *variable aleatoria finita* a toda función $X: E \rightarrow R$ que pueda tomar únicamente un número finito de valores $X = x_1, x_2, \dots, x_n$, con la condición de que, para cualquiera de ellos, por ejemplo x_i , el conjunto $X^{-1}(x_i)$ sea un elemento de \mathbf{B} .

Por cierto que si E es finito y aceptamos el convenio de la página 16, según el cual \mathbf{B} es el conjunto de todos los subconjuntos de E , la última condición puede suprimirse, y variable aleatoria finita es cualquier función $X: E \rightarrow R$.

Puesto que el conjunto $X^{-1}(x_i)$ (elementos de E cuya imagen por X es el número x_i) es un elemento de \mathbf{B} , tendrá una cierta probabilidad, que se representa por

$$P(X^{-1}(x_i)) = P(X = x_i) = f(x_i) \quad [3.2]$$

y se establece la siguiente definición.

Definición 3.2. La función f definida por [3.2] se llama la *función de probabilidad* de la variable aleatoria X .

Se dice, abreviadamente, que $f(x_i)$ es la probabilidad de que la variable aleatoria X tome el valor x_i .

Sean x_1, x_2, \dots, x_n los valores de X y pongamos $A_i = X^{-1}(x_i) = \{a \in E; X(a) = x_i\}$. Según [3.2] y los axiomas de la definición de probabilidad, tendremos

$$\sum_{i=1}^n f(x_i) = \sum_{i=1}^n P(A_i) = P(A_1 \cup A_2 \cup \dots \cup A_n) = P(E) = 1.$$

Por otra parte, dado que los valores de f son probabilidades, es siempre $f \geq 0$. Las dos relaciones

$$\sum_{i=1}^n f(x_i) = 1, \quad f(x_i) \geq 0 \quad [3.3]$$

se cumplen siempre para cualquier función de probabilidad.

Supongamos que los valores de X están ordenados $x_1 < x_2 < \dots < x_n$. Muchas veces interesa la probabilidad de que X tome un valor *igual* o *menor* que x_1 . Su valor será

$$F(x_1) = P(X \leq x_1) = \sum_{h=1}^1 f(x_h) \quad [3.4]$$

lo cual da lugar a la siguiente definición.

Definición 3.2. La función F definida por [3.4] se llama la *función de distribución* de la variable aleatoria finita X .

Obsérvese que $F(x_n) = 1$.

Ejemplo 1. *Se lanzan 3 monedas. Análizese la variable aleatoria $X = \text{número de caras}$.*

35

Solución. El espacio muestral consta de 8 elementos:

(C, C, C), (C, C, F), (C, F, C), (F, C, C)
(C, F, F), (F, C, F), (F, F, C), (F, F, F)

cada uno de los cuales tiene la misma probabilidad $1/8$. Los números posibles de caras son $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$. Los valores de la función de probabilidad y de la función de distribución están dados en la siguiente tabla:

X	0	1	2	3
f	1/8	3/8	3/8	1/8
F	1/8	1/2	7/8	1

Obsérvese, como detalle que será importante en el próximo capítulo, que los valores de f están dados por la fórmula

$$f(x) = \binom{3}{x} (1/2)^3, \quad x = 0, 1, 2, 3.$$

Ejemplo 2. Una urna contiene 12 bolillas numeradas de 1 a 12. Se saca una bolilla y se quiere analizar la variable aleatoria $X =$ número de divisores del número obtenido.

Solución. Se tiene:

No. sacado	1	2	3	4	5	6	7	8	9	10	11	12
X	1	2	2	3	2	4	2	4	3	4	2	6

De aquí se deduce

$$X^{-1}(1) = \{1\}, \quad X^{-1}(2) = \{2, 3, 5, 7, 11\}, \quad X^{-1}(3) = \{4, 9\},$$

$$X^{-1}(4) = \{6, 8, 10\}, \quad X^{-1}(6) = \{12\}$$

con lo cual se puede escribir la siguiente tabla:

X	1	2	3	4	6
f	1/12	5/12	1/6	1/4	1/12
F	1/12	1/2	2/3	11/12	1

36

Esto nos dice, por ejemplo, que la probabilidad de que un número elegido al azar entre 1 y 12 tenga 3 divisores es $f(3) = 1/6$, y la probabilidad de que tenga *menos* de 3 divisores es $F(2) = 1/2$.

Ejemplo 3. Se lanzan 2 dados y sean i, j los números resultantes. Estúdiense las variables aleatorias $X = i + j$, $Y =$ m. c. d. (i, j) = máximo común divisor de i, j .

Solución. La probabilidad de cada par ordenado es $1/36$. Contando directamente en cada caso el número de pares que corresponden a cada valor de X o de Y, resultan las siguientes tablas:

X	2	3	4	5	6	7	8	9	10	11	12
f	1/36	1/18	1/12	1/9	5/36	1/6	5/36	1/9	1/12	1/18	1/36
F	1/36	1/12	1/6	5/18	15/36	21/36	13/18	15/18	11/12	35/36	1

Y	1	2	3	4	5	6
f	23/36	7/36	1/12	1/36	1/36	1/36
F	23/36	15/18	1/12	17/18	35/36	1

Estas tablas resuelven problemas del siguiente tipo: a) Lanzando dos dados al azar, la probabilidad de que la suma de los puntos sea 7 es $f(7) = 1/6$, y la de que la suma de los puntos sea ≤ 10 , es $F(10) = 11/12$; b) Dados al azar dos números entre 1 y 6, la probabilidad de que su m. c. d. sea menor que 5 vale $F(4) = 17/18$.

Ejemplo 4. Ejercicio. *Se lanzan dos dados y sean i, j los números obtenidos. Estúdiese la variable aleatoria $X =$ número de divisores de la suma $i + j$, y compárense los resultados con los del ejemplo 2.*

2. ESPERANZA MATEMATICA

Sea una variable aleatoria finita X que puede tomar los valores x_1, x_2, \dots, x_n con las probabilidades $f(x_1), f(x_2), \dots, f(x_n)$.

Definición 3.3. Se llama *esperanza matemática* o *valor medio* de la variable aleatoria X , a la expresión

$$E(X) = \sum_{i=1}^n x_i f(x_i). \quad [3.5]$$

Si $f =$ constante, puesto que $\sum f(x_i) = 1$, debe ser $f(x_i) = 1/n$, y la esperanza matemática resulta $E(X) = (1/n) \sum x_i$, o sea, es igual a la media aritmética de los valores de X .

37

Más generalmente, se establece la siguiente definición.

Definición 3.4. Se llama *esperanza matemática* o *valor medio* de una función $G(X)$ de la variable aleatoria X , a la expresión

$$E(G(X)) = \sum_{i=1}^n G(x_i) f(x_i). \quad [3.6]$$

Ejemplo 1. La esperanza matemática de la variable aleatoria $X =$ número que resulta al lanzar un dado, será

$$E(X) = 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) = 3,5.$$

Ejemplo 2. La esperanza matemática de la suma de los números resultantes al lanzar dos dados a la vez, según el ejemplo 3 de la sección 1, será

$$E(X) = (1/36)(2 + 3.2 + 4.3 + 5.4 + 6.5 + 7.6 + 8.5 + 9.4 + 10.3 + 11.2 + 12) = 7.$$

Ejemplo 3. La esperanza matemática del número de divisores de un número elegido al azar entre 1 y 12, según el ejemplo 2 de la sección 1, será

$$E(X) = (1/12)(1 + 2.5 + 3.2 + 4.3 + 6.1) = 2.91\dots$$

3. MOMENTOS DE UNA VARIABLE ALEATORIA

La esperanza matemática de una variable aleatoria es un dato muy importante de la misma, pero no dice nada acerca de la dispersión de sus valores alrededor de esta esperanza o valor medio. Por ejemplo, si X representa el capital de una persona escogida al azar de un grupo de dos, tanto si la primera posee \$99.000 y la segunda \$1.000, como si cada una tiene \$50.000, el valor medio o esperanza matemática tiene el mismo valor de \$50.000. Igualmente, del hecho de saber que la talla media de un conjunto de personas es 1,67 m no se puede deducir si todas tienen una altura aproximada a este valor medio o si hay algunas muy altas y otras muy bajas. Otro ejemplo típico es el caso en que X representa el caudal de las aguas de un río durante las 52 semanas del año. Del conocimiento de su valor medio no se puede deducir si se trata de un río de caudal aproximadamente constante durante todo el año, o bien de un río muy caudaloso en invierno y casi seco en verano.

Por este motivo se han introducido más datos que, en cierta manera, permiten medir esta dispersión de los valores de X alrededor de su valor medio $E(X)$. Estos datos suelen ser la esperanza matemática de ciertas funciones $G(X)$. Por ejemplo, se puede tomar $G(X) = |X - E(X)|$ = valor absoluto de la diferencia $X - E(X)$. Evidentemente esta esperanza nos dará una idea de la mayor o menor concentración de los valores de X en torno de $E(X)$. Sin embargo, como los valores absolutos son de incómodo manejo matemáticamente, se han preferido otras funciones. Las más importantes son las potencias de X , que dan lugar a los llamados momentos, que pasamos a definir.

Definición 3.5. Se llama *momento de orden k* de la variable aleatoria finita X a la esperanza matemática de X^k , o sea,

$$\alpha_k = E(X^k) = \sum_{i=1}^n x_i^k f(x_i). \quad [3.7]$$

En particular $\alpha_1 = E(X)$. Los *momentos centrados* se definen por la ecuación

$$\mu_k = E((X - \alpha_1)^k) = \sum_{i=1}^n (x_i - \alpha_1)^k f(x_i). \quad [3.8]$$

Es particularmente importante el momento centrado de segundo orden, que da lugar a la siguiente definición.

Definición 3.6. Se llama *varianza* o *variancia* de una variable aleatoria X al momento centrado de segundo orden. Se representa por σ^2 o por $\sigma^2(X)$, o sea

$$\text{Var}(X) = \sigma^2 = \sigma^2(X) = E((X - \alpha_1)^2) = \sum_{i=1}^n (x_i - \alpha_1)^2 f(x_i). \quad [3.9]$$

Definición 3.7. El número no negativo σ , raíz cuadrada de la varianza, se llama *desviación típica* o *desviación "standard"* de la variable aleatoria X .

Se comprende que σ^2 mida en cierta manera la separación de los valores de X de su valor medio α_1 . En efecto, como σ^2 es una suma de términos positivos, si σ^2 es pequeña, todos ellos deben ser pequeños y por tanto, o bien $x_i - \alpha_1$ es pequeño (para todo x_i) o bien es pequeña la probabilidad $f(x_i)$. Al contrario, si σ^2 es grande, ello significa que hay valores de X muy separados de su valor medio. En resumen, la idea que se debe tener de σ^2 es que si tiene un valor pequeño se trata de una variable aleatoria X cuyos valores difieren poco de su valor medio y, si tiene un valor grande, significa que hay valores de X alejados de su valor medio.

39

Las siguientes relaciones en las cuales a, b son constantes son importantes y su demostración, en cada caso, es una consecuencia inmediata de las definiciones anteriores, por lo que nos limitamos a enunciarlas:

$$E(aX + b) = aE(X) + b \quad [3.10]$$

$$E(X - \alpha_1) = 0 \quad [3.11]$$

$$\sigma^2(X) = E(X^2) - (E(X))^2 \quad [3.12]$$

$$\sigma^2(aX + b) = a^2\sigma^2(X). \quad [3.13]$$

4. LA DESIGUALDAD DE TCHEBYCHEFF

Sea $H(X)$ una función de la variable aleatoria X que no tome valores negativos, o sea, $H(x_i) \geq 0$ para todos los valores x_i de X . Sea $K > 0$ una constante dada, y sean x_i ($i = 1, 2, \dots, n$) los valores de X , y x_α aquellos valores para los cuales sea $H(x_\alpha) \geq K$. Tenemos

$$E(H(X)) = \sum_{i=1}^n H(x_i) f(x_i) \geq \sum_{\alpha} H(x_\alpha) f(x_\alpha) \geq K \sum_{\alpha} f(x_\alpha)$$

Pero $\sum f(x_\alpha)$ indica la probabilidad de que $H(X)$ sea igual o mayor que K . Por tanto se tiene

$$P(H(X) \geq K) \leq \frac{E(H(X))}{K}. \quad [3.14]$$

En particular, tomando $H(X) = (X - \alpha_1)^2$ y $K = k^2 \sigma^2$, resulta la llamada *desigualdad de TCHEBYCHEFF*, a saber,

$$P(|X - \alpha_1| \geq k\sigma) \leq \frac{1}{k^2}. \quad [3.15]$$

Poniendo $k\sigma = k_1$, se puede escribir también

$$P(|X - \alpha_1| \geq k_1) \leq \frac{\sigma^2}{k_1^2}. \quad [3.16]$$

La desigualdad de Tchebycheff es muy importante, y aunque la hemos demostrado para variables aleatorias finitas, ella es válida con análoga demostración para las variables aleatorias discretas y continuas, que se definirán en el capítulo 6.

5. SUMA DE VARIABLES ALEATORIAS

Sean X e Y dos variables aleatorias y $\{x_i, f(x_i)\} (i=1, 2, \dots, n)$, $\{y_j, g(y_j)\} (j=1, 2, \dots, m)$ los valores respectivos de X e Y , junto con sus correspondientes probabilidades.

Definición 3.8. Se llama variable aleatoria *suma* $X + Y$ a la variable aleatoria que toma los valores $x_i + y_j (i=1, 2, \dots, n; j=1, 2, \dots, m)$, con las probabilidades $P(x_i, y_j) = P(X = x_i, Y = y_j) =$ probabilidad de que X tome el valor x_i e Y tome el valor y_j .

Obsérvense las relaciones

$$\sum_{j=1}^m P(x_i, y_j) = f(x_i), \quad \sum_{i=1}^n P(x_i, y_j) = g(y_j) \quad [3.17]$$

puesto que la suma, respecto de j , de la probabilidad de que ocurra (x_i, y_j) es igual a la probabilidad de que ocurra x_i , lo que da la primera igualdad. La segunda es análoga.

Se tiene el siguiente teorema.

Teorema 3.1. *Cualesquiera que sean las variables aleatorias X e Y se verifica*

$$E(X + Y) = E(X) + E(Y) \quad [3.18]$$

Demostración. Basta aplicar la definición y las igualdades [3.17]. En efecto

$$\begin{aligned}
E(X + Y) &= \sum_{i,j} (x_i + y_j)P(x_i, y_j) = \sum_i x_i \sum_j P(x_i, y_j) + \sum_j y_j \sum_i P(x_i, y_j) = \\
&= \sum_i x_i f(x_i) + \sum_j y_j g(y_j) = E(X) + E(Y),
\end{aligned}$$

donde las sumas respecto de i van de 1 a n , y las sumas respecto de j van de 1 a m .

Con la misma demostración, teniéndose en cuenta [3.10], resulta de manera general que si X_1, X_2, \dots, X_n son variables aleatorias cualesquiera y a_1, a_2, \dots, a_n son constantes, entonces

$$E(a_1 X_1 + a_2 X_2 + \dots + a_n X_n) = a_1 E(X_1) + a_2 E(X_2) + \dots + a_n E(X_n). \quad [3.19]$$

En la demostración se ha supuesto que X e Y son variables aleatorias finitas. Sustituyendo las sumas por series o por integrales, la misma demostración prueba que el teorema es válido también para las variables aleatorias discretas y continuas definidas en el capítulo 6. La misma observación es válida para los teoremas 3.2, 3.3 y 3.4.

El teorema anterior permite resolver algunos problemas de manera más fácil que por cálculo directo. Por ejemplo:

Problema 3.1. *Dos urnas contienen, cada una, bolillas numeradas de 1 a 10. Se saca una bolilla de cada urna y se suman los números obtenidos, ¿cuál es el valor medio de la suma?*

Solución. Si X e Y son las variables aleatorias que expresan el número sacado de cada urna, es $E(X) = E(Y) = (1/10)(1 + 2 + \dots + 10) = 11/2$ y por tanto la solución es $E(X + Y) = 11$.

Problema 3.2. *Una urna contiene 10 bolillas numeradas de 1 a 10. Si se sacan a la vez dos bolillas ¿cuál es el valor medio de la suma?*

Solución. Aquí las variables aleatorias X e Y , que expresan respectivamente el número de la primera y segunda bolilla, no son independientes, pues no pueden tomar el mismo valor. Sin embargo, puesto que el teorema 3.1 vale igualmente, resulta que la solución es la misma de antes, o sea, $E(X + Y) = 11$.

Antes de seguir necesitamos la siguiente definición, que está de acuerdo con [2.10].

Dos variables aleatorias X e Y cuyas funciones de probabilidad respectivas son f y g , se dice que son *independientes*, si, y sólo si, la probabilidad del par $(X = x_i, Y = y_j)$ es igual al producto de la probabilidad $f(x_i)$ por la probabilidad $g(y_j)$. Es decir, si se cumple

$$P(X = x_i, Y = y_j) = f(x_i)g(y_j)$$

para todo par de valores x_i, y_j de las variables.

Teorema 3.2. Si X e Y son dos variables aleatorias independientes, entonces

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y). \quad [3.20]$$

Demostración. Siendo X e Y independientes, se cumple la igualdad última y por tanto, poniendo $E(X) = \alpha_1$ y $E(Y) = \beta_1$, se tiene

$$\begin{aligned} \sigma^2(X + Y) &= \sum_{i,j} (x_i + y_j - \alpha_1 - \beta_1)^2 f(x_i)g(y_j) = \\ &= \sum_{i,j} (x_i - \alpha_1)^2 f(x_i)g(y_j) + \sum_{i,j} (y_j - \beta_1)^2 f(x_i)g(y_j) + \\ &\quad + 2 \sum_i (x_i - \alpha_1) f(x_i) \sum_j (y_j - \beta_1) g(y_j) \end{aligned}$$

y como $\sum_j g(y_j) = 1, \sum_i f(x_i) = 1, E(X - \alpha_1) = 0, E(Y - \beta_1) = 0$, resulta [3.20].

En general, de [3.20] y [3.13] resulta, para variables aleatorias *independientes*:

42

$$\sigma^2(a_1 X_1 + a_2 X_2 + \dots + a_n X_n) = a_1^2 \sigma^2(X_1) + a_2^2 \sigma^2(X_2) + \dots + a_n^2 \sigma^2(X_n) \quad [3.21]$$

En particular, si $a_1 = 1, a_2 = -1$,

$$\sigma^2(X - Y) = \sigma^2(X) + \sigma^2(Y). \quad [3.22]$$

6. PRODUCTO DE VARIABLES ALEATORIAS

Definición 3.9. Se llama *variable aleatoria producto* $X \cdot Y$ de las variables aleatorias $X(x_i, f(x_i)), Y(y_j, g(y_j))$, a la variable aleatoria que toma los valores $x_i y_j$ con probabilidades $P(x_i, y_j)$.

Teorema 3.3. Si X e Y son variables aleatorias independientes, se tiene

$$E(X \cdot Y) = E(X) \cdot E(Y). \quad [3.23]$$

Demostración. Siendo X e Y independientes, es $P(x_i, y_j) = f(x_i)g(y_j)$, y por tanto

$$E(X \cdot Y) = \sum_{i,j} x_i y_j f(x_i)g(y_j) = \sum_i x_i f(x_i) \sum_j y_j g(y_j) = E(X) \cdot E(Y).$$

Si las variables aleatorias X e Y no son independientes, conviene adoptar la siguiente definición.

Definición 3.10. Se llama *covarianza* o *covariancia* de dos variables aleatorias X e Y a la expresión

$$\text{cov}(X, Y) = E((X - \alpha_1)(Y - \beta_1)) = \sum_{i,j} (x_i - \alpha_1)(y_j - \beta_1)P(x_i, y_j),$$

donde, como siempre, $E(X) = \alpha_1$, $E(Y) = \beta_1$.

Otra forma de la covarianza, que se obtiene desarrollando $E((X - \alpha_1)(Y - \beta_1)) = E(XY - \alpha_1 Y - \beta_1 X + \alpha_1 \beta_1)$, es

$$\text{cov}(X, Y) = E(X \cdot Y) - E(X)E(Y). \quad [3.24]$$

De aquí se deduce que, en el caso de variables aleatorias dependientes, la fórmula [3.23] debe ser sustituida por

$$E(X \cdot Y) = E(X) \cdot E(Y) + \text{cov}(X, Y). \quad [3.25]$$

En particular, si X e Y son independientes, se verifica

$$\text{cov}(X, Y) = 0. \quad [3.26]$$

Teorema 3.4. *Para dos variables aleatorias X e Y independientes o no vale*

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y) + 2 \text{cov}(X, Y). [3.27]$$

Demostración. Se tiene

$$\begin{aligned} \sigma^2(X + Y) &= \sum_{i,j} (x_i - \alpha_1 + y_j - \beta_1)^2 P(x_i, y_j) = \\ &= \sum_i (x_i - \alpha_1)^2 \sum_j P(x_i, y_j) + \sum_j (y_j - \beta_1)^2 \sum_i P(x_i, y_j) + \\ &+ 2 \sum_{i,j} (x_i - \alpha_1)(y_j - \beta_1) P(x_i, y_j) \end{aligned}$$

y teniendo en cuenta [3.17] y la definición de covarianza, resulta el enunciado.

Problema 3.3. *Dos urnas contienen, cada una, 5 bolillas numeradas de 1 a 5. Si se saca una bolilla de cada urna y se multiplican los números obtenidos, ¿cuál es el valor medio de este producto?*

Solución. Como las variables aleatorias X e Y que representan los números sacados de cada urna, son independientes, y $E(X) = E(Y) = (1/5)(1 + 2 + \dots + 5) = 3$, resulta $E(X \cdot Y) = 9$.

Problema 3.4. Una urna contiene 5 bolillas numeradas de 1 a 5. Se sacan dos bolillas al azar y se multiplican los números obtenidos. ¿Cuál es el valor medio del producto?

Solución. Ahora las variables X e Y son dependientes y por tanto no puede aplicarse la fórmula [3.23]. Hay que hacer el cálculo directo. Los casos posibles son $(1, 2), (1, 3), \dots, (4, 5)$, y de cada uno de ellos la probabilidad es $1/10$. Por tanto, $E(X \cdot Y) = 8,5$, resultado distinto del anterior.

7. CORRELACION

Puesto que $\text{cov}(X, Y) = 0$ en el caso en que X e Y son independientes, puede decirse que la $\text{cov}(X, Y)$ mide, en cierta manera, el grado de dependencia entre X e Y . Sin embargo, $\text{cov}(X, Y)$ tiene el inconveniente de depender de las unidades de medida, es decir, si en vez de X e Y se toman las variables aX, bY (a, b constantes) es $\text{cov}(aX, bY) = ab \text{cov}(X, Y)$, como resulta inmediatamente de la definición. Para evitar este inconveniente se ha recurrido al siguiente concepto:

Definición 3.11. Se llama *coeficiente de correlación* entre dos variables aleatorias, cuyas desviaciones típicas $\sigma(X), \sigma(Y)$ no sean nulas, al cociente

$$\rho = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}. \quad [3.28]$$

Este cociente no cambia si se multiplican las variables aleatorias por constantes positivas. Se va a demostrar que $-1 \leq \rho \leq 1$. Para ello, siendo λ, μ dos constantes, consideremos

$$E(\lambda(X - \alpha_1) + \mu(Y - \beta_1))^2 = \lambda^2\sigma^2(X) + \mu^2\sigma^2(Y) + 2\lambda\mu \text{cov}(X, Y). \quad [3.29]$$

Por ser la esperanza matemática un cuadrado, esta expresión es siempre ≥ 0 , y por tanto

$$\sigma^2(X) \cdot \sigma^2(Y) - (\text{cov}(X, Y))^2 \geq 0 \quad [3.30]$$

de donde:

a) $\rho^2 \leq 1$, y por tanto $-1 \leq \rho \leq 1$.

b) Si $\rho = \pm 1$, existen valores $\lambda = \lambda_0, \mu = \mu_0$ (no ambos nulos), para los cuales [3.29] se anula, o sea $E(\lambda_0(X - \alpha_1) + \mu_0(Y - \beta_1))^2 = 0$, lo que exige que sea

$$\lambda_0(X - \alpha_1) + \mu_0(Y - \beta_1) = 0. \quad [3.31]$$

Esta igualdad significa que los únicos pares de valores x_1, y_1 de las variables X, Y que tienen probabilidad distinta de cero de ocurrir, son los que verifican $\lambda_0(x_1 - \alpha_1) + \mu_0(y_1 - \beta_1) = 0$, es decir, los pares (x_1, y_1) son coordenadas de puntos pertenecientes a la recta [3.31]. Esto implica que entre X e Y hay una correspondencia funcional lineal, o sea, se tiene el siguiente teorema.

Teorema 3.5. *El coeficiente de correlación es un número real ρ comprendido entre -1 y +1, tal que $\rho = \pm 1$ implica que entre X e Y existe una dependencia funcional lineal.*

Si X e Y son independientes, hemos visto que $\rho = 0$. Sin embargo, esta condición necesaria no es suficiente para la independencia de las variables X e Y. Consideremos el siguiente ejemplo. Sea U la variable aleatoria que representa el número de un dado lanzado al azar, y sea V la variable análoga para un segundo dado, independiente del primero. Consideremos las variables aleatorias $X = U + V$ e $Y = U - V$. Tendremos $E(X \cdot Y) = E(U^2 - V^2) = E(U^2) - E(V^2) = 0$. Por otra parte, también es $E(Y) = E(U) - E(V) = 0$. Por consiguiente, según [3.24] es $\text{cov}(X, Y) = 0$, de donde $\rho = 0$. Sin embargo, X e Y no son independientes, pues ellas toman valores pares o impares a la vez.

8. FUNCION GENERATRIZ DE MOMENTOS

Definición 3.12. Se llama *función generatriz de momentos*, o simplemente, *función generatriz* de la variable aleatoria X, a la esperanza matemática de la función e^{xt} , o sea, a la función de t definida por

$$\Psi(t) = E(e^{xt}) = \sum_{i=1}^n e^{x_i t} f(x_i). \quad [3.32]$$

El nombre proviene de que, conocida la función $\Psi(t)$, sus derivadas sucesivas en el punto $t = 0$, son los momentos de X. En efecto, se tiene

$$\Psi(0) = \sum_{i=1}^n f(x_i) = 1$$

$$\Psi'(0) = \sum_{i=1}^n x_i f(x_i) = E(X) = \alpha_1$$

$$\Psi''(0) = \sum_{i=1}^n x_i^2 f(x_i) = E(X^2) = \alpha_2$$

y, en general,

$$\Psi^{(r)}(0) = \sum_{i=1}^n x_i^r f(x_i) = E(X^r) = \alpha_r.$$

Teorema 3.6. *Si $\Psi_x(t)$, $\Psi_y(t)$ son las funciones generatrices de las variables aleatorias independientes X e Y, y $\Psi_{x+y}(t)$ es la función generatriz de la variable aleatoria X + Y, se verifica*

$$\Psi_{x+y}(t) = \Psi_x(t) \cdot \Psi_y(t). \quad [3.33]$$

Demostración. En efecto, si X e Y son independientes, también lo serán e^{Xt} y e^{Yt} , y por tanto, según [3.22]

$$\Psi_{X+Y}(t) = E(e^{(X+Y)t}) = E(e^{Xt}e^{Yt}) = E(e^{Xt})E(e^{Yt}) = \Psi_X(t)\Psi_Y(t). \quad [3.34]$$

Repitiendo el razonamiento, resulta que si X_1, X_2, \dots, X_n son variables aleatorias independientes, se tiene

$$\Psi_{X_1+X_2+\dots+X_n}(t) = \Psi_{X_1}(t) \dots \Psi_{X_n}(t). \quad [3.35]$$

Por la definición [3.32], la función de probabilidad determina la función generatriz de momentos. Ahora se plantea el problema inverso: Dada la función generatriz $\Psi(t)$, ¿queda unívocamente determinada la función de probabilidad f ? Este es un teorema matemático difícil. Se trata de buscar una fórmula de inversión que permita despejar las $f(x_i)$ en [3.32]. Se puede demostrar que bajo ciertas condiciones muy amplias, que se cumplen siempre en los casos usuales, la inversión es posible, es decir, la función generatriz determina unívocamente la función de probabilidad. Nosotros admitiremos este resultado sin demostración.

Aunque más adelante ya veremos otras aplicaciones del último teorema, vamos ahora a aplicarlo a la solución de un problema clásico de De Moivre (*The doctrine of chances*, 1718).

46

Problema 3.5. *Una urna contiene $n + 1$ bolillas numeradas $1, 2, \dots, n + 1$. Se saca r veces una bolilla y se devuelve a la urna después de anotar su número. Se desea saber la probabilidad de que la suma de los r números asíobtenidos sea igual a un valor dado s .*

Solución. La probabilidad es distinta de cero sólo si $r \leq s \leq r(n + 1)$; supongamos que esta condición se cumple. Consideremos las r variables aleatorias X_1, X_2, \dots, X_r correspondientes a los números de cada extracción. Cada una de estas variables puede tomar, independientemente una de otra, los valores $1, 2, \dots, n + 1$, con probabilidad $1/(n + 1)$ para cada uno. Por tanto, para cualquier X_i , es

$$\Psi_{X_i}(t) = \frac{1}{n + 1} (e^t + e^{2t} + \dots + e^{(n+1)t})$$

y por consiguiente, según [3.35],

$$\Psi_{X_1+X_2+\dots+X_r}(t) = \left(\frac{1}{n + 1}\right)^r (e^t + e^{2t} + \dots + e^{(n+1)t})^r. \quad [3.36]$$

Por otra parte, según la definición de función generatriz, es

$$\Psi_{X_1+X_2+\dots+X_r}(t) = \sum_{h=r}^{r(n+1)} e^{ht} P_h \quad [3.37]$$

donde P_h es la probabilidad de que la suma de las r bolillas sea h . Por consiguiente, el coeficiente de e^{st} en la expresión [3.36] es la probabilidad buscada. Elevando la potencia r -ésima del segundo miembro de [3.36] y juntando términos semejantes se obtiene el resultado

$$P_s = \frac{1}{(n+1)^r} (C_{s-1, r-1} - C_{r, 1} C_{s-n-2, r-1} + C_{r, 2} C_{s-2n-3, r-1} - \dots). \quad [3.38]$$

donde $C_{i,j}$ indican los números combinatorios (véase el Apéndice I) y la suma alternada debe proseguirse mientras ninguno de los índices resulte negativo ni el segundo mayor que el primero.

Ejemplo 1. Se desea hallar la probabilidad de sacar la suma 9 arrojando 3 dados. Hay que hacer $s = 9$, $r = 3$, $n = 5$ en la fórmula general y resulta $P_9 = (1/6^3)(C_{8,2} - C_{3,1}C_{2,2}) = 25/216$ de acuerdo con el problema 2.7.

Ejemplo 2. Se desea saber cuál es la probabilidad de sacar la suma 15 arrojando 6 dados. Hay que tomar $s = 15$, $r = 6$, $n = 5$, y resulta $P_{15} = (1/6^6)(C_{14,5} - C_{6,1}C_{8,5}) = 833/23328$.

9. FUNCIÓN CARACTERÍSTICA

47

La función generatriz puede definirse igualmente para variables aleatorias con una infinidad numerable de valores. Basta, para ello, sustituir la suma de [3.32] por una serie desde $i = 1$ hasta $i = \infty$. En este caso, sin embargo, la función generatriz sólo tendrá sentido si la serie resulta convergente. Si en lugar de e^{xt} , se toma e^{ixt} , donde i es la unidad imaginaria, la convergencia es siempre más segura. Por otra parte, en el caso de variables aleatorias continuas, que veremos más adelante, la sustitución de e^{xt} por e^{ixt} presenta todavía otras ventajas (poder aplicar, por ejemplo, la fórmula de inversión de Fourier). De aquí que en los tratados de probabilidades desde un punto de vista superior, en vez de la función generatriz de momentos, se suele utilizar la llamada *función característica*, definida por

$$\Psi(t) = \Psi(it) = E(e^{ixt}).$$

Como nosotros no vamos a utilizar esta función característica, nos contentamos con la definición.

10. REGRESION

Sean dos variables aleatorias X e Y y sea $f_{i,j} = f(x_i, y_j)$ la probabilidad del par $X = x_i$ e $Y = y_j$. De manera análoga a [2.8], se tiene ahora

$$f(x_i, y_j) = h(x_i)f(y_j|x_i), \quad h(x_i) = \sum_j f_{ij} \quad [3.39]$$

donde $f(y_j|x_i)$ representa la probabilidad de que sea $Y = y_j$, condicionada a $X = x_i$, y $h(x_i)$ es la probabilidad de $X = x_i$. Para cada valor de X , por ejemplo $X = x_i$, la esperanza de Y es un cierto valor

$$y'_i = \sum_j y_j f(y_j|x_i).$$

Se tiene así una función y' cuyo dominio es el conjunto de valores de X , que se llama la *regresión* de Y sobre X .

Se trata de aproximar esta función mediante una expresión lineal $y = a + bx$. Para ello se aplica el método de los mínimos cuadrados, que consiste en determinar las constantes a y b de manera que la esperanza

$$E = \sum_i \left(\sum_j y_j f(y_j|x_i) - a - bx_i \right)^2 h(x_i)$$

sea mínima. Para determinar los valores de a y b que hacen mínima E , deben ser nulas las derivadas parciales de esta expresión respecto de a y b y se tienen las ecuaciones (teniendo en cuenta [3.39] y que $\sum_{i,j} f_{ij} = 1$)

$$E(Y) - a - bE(X) = 0, \quad E(X \cdot Y) - aE(X) - bE(X^2) = 0$$

donde, aplicando [3.12] y [3.24]

$$a = E(Y) - \frac{\text{cov}(X, Y)}{\sigma^2(X)} E(X), \quad b = \frac{\text{cov}(X, Y)}{\sigma^2(X)}.$$

Estos son los coeficientes de la llamada *recta de regresión* de Y sobre X . El coeficiente b se llama *coeficiente de regresión*.

DISTRIBUCION BINOMIAL, LEY DE LOS GRANDES NUMEROS

1. DISTRIBUCION BINOMIAL

En 1713 se publicó un libro famoso y fundamental en la historia del cálculo de probabilidades, titulado *Ars Conjectandi*, cuyo autor, Jacobo Bernoulli, había fallecido ocho años antes. En este libro Bernoulli introduce el siguiente modelo probabilístico, llamado de las "pruebas repetidas" o "pruebas de Bernoulli".

Supongamos una urna con n_1 bolillas blancas y n_2 bolillas negras. La probabilidad de sacar una bolilla blanca es $p = n_1 / (n_1 + n_2)$ y la de sacar una bolilla negra es $q = n_2 / (n_1 + n_2)$. Se hacen n pruebas, devolviendo cada vez la bolilla a la urna para que todas las pruebas estén en las mismas condiciones. Se desea hallar la probabilidad de que en el curso de tales pruebas salgan r bolillas blancas y $n - r$ bolillas negras (independientemente del orden en que salgan).

49

La solución es fácil. Si se pidiera la probabilidad del mismo problema, pero dando el orden en que van saliendo las bolillas, como en cada extracción la probabilidad de blanca es p y la de negra es q y se trata de sucesos independientes, según la relación [2.12] aplicada sucesivamente, resulta que la probabilidad buscada sería $p^r q^{n-r}$. Si el orden no interesa, hay que sumar esta probabilidad para todas las ordenaciones posibles, que son $\binom{n}{r}$, puesto que las r bolillas blancas pueden distribuirse entre las n extracciones. Resulta así que la probabilidad buscada es

$$P_r = \binom{n}{r} p^r q^{n-r}. \quad [4.1]$$

Esta función de probabilidad que, una vez dados n y p , tiene un valor determinado para cada r , se llama la *función binomial*, por ser sus valores iguales a los términos del desarrollo de $(p+q)^n$ por la fórmula del binomio de Newton. Puesto que $p+q=1$, esto prueba, además, que $\sum_r P_r = 1$ (la suma extendida a $r = 0, 1, \dots, n$), como debe ser según la primera condición [3.3].

El modelo o esquema de Bernoulli tiene muchas aplicaciones. Para enunciarlo de manera general, dentro del sistema axiomático del capítulo 2, podemos establecer que todo esquema de Bernoulli consta de los siguientes elementos: a) Un experimento aleatorio del cual pueden resultar un *éxito* A con probabilidad p o un *fracaso* B con probabilidad $q = 1 - p$; b) un número n de pruebas que, anotando cada vez si tiene lugar A o B, da lugar al espacio muestral E formado por los 2^n elementos siguientes

AAA...A, AA...AB, AA...BA, ..., BB...B

cada uno de los cuales consta de n letras, elegidas entre A y B, de todas las maneras posibles, teniendo en cuenta el orden; c) el número r de éxitos A en cada prueba, lo que da origen a la variable aleatoria X, cuyos valores posibles son $0, 1, 2, \dots, n$ con las probabilidades respectivas P_0, P_1, \dots, P_n dadas por [4.1].

Todo proceso en las condiciones anteriores se dice que sigue la *distribución binomial*, o la *ley binomial*, con función de probabilidad [4.1]. Es costumbre adoptar la notación

$$b(r;n, p) = \binom{n}{r} p^r q^{n-r}, \quad r = 0, 1, \dots, n \quad [4.2]$$

y también es muchas veces útil la *función de distribución binomial*, a saber,

$$F(x) = \sum_{r \leq x} \binom{n}{r} p^r q^{n-r} \quad [4.3]$$

que da la probabilidad de que el número de éxitos A sea $\leq x$.

Para ver cómo varía la función $b(r;n, p)$, dados p y n , formemos el cociente

$$\frac{b(r;n, p)}{b(r-1;n, p)} = \frac{(n-r+1)p}{rq} = 1 + \frac{(n+1)p-r}{rq}$$

Distingamos dos casos: a) Existe un entero $r = r_0$ tal que $(n+1)p = r_0$. En este caso, para este valor r_0 es $b(r_0;n, p) = b(r_0-1;n, p)$ y para valores $r > r_0$ es $b(r;n, p) < b(r-1;n, p)$ (o sea b es decreciente), en tanto que para valores $r < r_0$ es $b(r;n, p) > b(r-1;n, p)$ (o sea, b es también decreciente hacia la izquierda). En resumen, $b(r;n, p)$ tiene dos valores iguales, correspondientes a $r_0 = (n+1)p$ y a r_0-1 , para los cuales toma el valor máximo. b) No existe un entero r_0 que sea igual a $(n+1)p$. En este caso, sea r_0 el único entero comprendido entre $(n+1)p-1$ y $(n+1)p$. Este será el único valor de r para el cual $b(r_0-1;n, p) < b(r_0;n, p) > b(r_0+1;n, p)$,

en tanto que para $r < r_0$, la función $b(r; n, p)$ crece y para $r > r_0$ decrece. Es decir, $b(r; n, p)$ tiene en este caso un solo valor r_0 en el que toma su valor máximo. Como veremos más adelante, este valor máximo es aproximadamente igual a $1/(2\pi npq)^{1/2}$. En las figuras 1 y 2 se dan ejemplos de ambos casos.

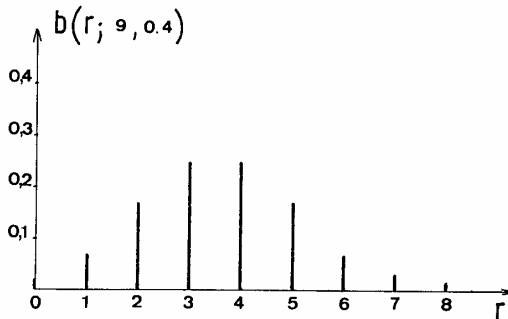


Fig. 1.

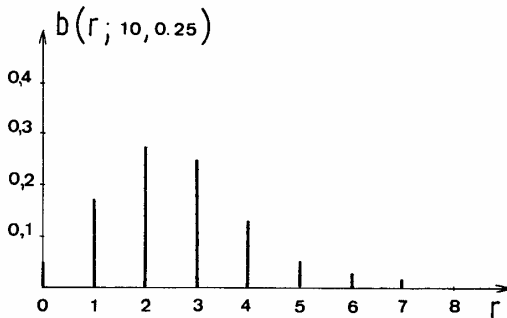


Fig. 2.

El cálculo de P no es fácil para valores un poco grandes de n y r . Se pueden aplicar tablas de los números combinatorios o tablas de factoriales (véase el Apéndice III) o la fórmula de Stirling (Apéndice I). Sin embargo, en general, es más práctico sustituir la ley binomial por otra ley aproximada, como la de Poisson o la normal, que veremos en el próximo capítulo.

Problema 4.1. *Se lanza una moneda 20 veces. Se busca: a) la probabilidad de sacar 14 veces cara; b) el número más probable de caras y la probabilidad de que salga este número.*

Solución. Los cálculos pueden hacerse directamente para darse cuenta de lo engorrosos que son, pero aquí han sido hechos mediante la tabla de factoriales.

$$a) P_{14} = \binom{20}{14} \left(\frac{1}{2}\right)^{20} = \frac{20!}{6!14!2^{20}} = \frac{38760}{1048576} = 0,036 \dots$$

b) La probabilidad máxima corresponde a $r = np = 10$, y es

$$P_{10} = \binom{20}{10} \left(\frac{1}{2}\right)^{20} = \frac{184756}{1048576} = 0,176 \dots$$

2. ESPERANZA MATEMÁTICA Y VARIANZA DE UNA VARIABLE BINOMIAL

En un esquema de Bernoulli podemos siempre considerar la variable aleatoria X_B , llamada *variable de Bernoulli*, que solamente puede tomar dos valores: el $x_1 = 1$, con probabilidad p , si el experimento aleatorio resulta un éxito, y el valor $x_2 = 0$, con probabilidad q , si el mismo resulta un fracaso. Entonces la variable aleatoria binomial X es la suma de n variables de Bernoulli.

La esperanza matemática de X_B es $E(X_B) = 1 \cdot p + 0 \cdot q = p$. Por tanto, según el teorema 3.1, es

$$E(X) = np. \quad [4.4]$$

Por otra parte, es $\sigma^2(X_B) = E((X_B - p)^2) = E(X_B^2) - p^2 = p - p^2 = pq$, y según el teorema 3.2, tratándose de variables aleatorias independientes (puesto que cada experimento es independiente de los anteriores), resulta

$$\sigma^2(X) = npq. \quad [4.5]$$

Estas fórmulas [4.4] y [4.5] se pueden también obtener directamente. En efecto, pruébese, como ejercicio, que

$$\sum_{r=0}^{\infty} r b(r; n, p) = np, \quad \sum_{r=0}^{\infty} (r - np)^2 b(r; n, p) = npq. \quad [4.6]$$

La función generatriz de la variable binomial X , según la definición 3.12, es

$$\Psi(t) = \sum_{r=0}^{\infty} e^{rt} \binom{n}{r} p^r q^{n-r} = (e^t p + q)^n. \quad [4.7]$$

De aquí se deduce,

$$E(X) = \Psi'(0) = np = \alpha_1$$

$$E(X^2) = \Psi''(0) = n(n-1)p^2 + np = \alpha_2$$

y por tanto, nuevamente

$$\sigma^2(X) = E(X^2) - (E(X))^2 = npq.$$

3. TEOREMA DE BERNOULLI

Apliquemos la desigualdad de Tchebycheff [3.15] a una variable binomial. Siendo $E(X) = \alpha_1 = np$, $\sigma^2 = npq$, resulta

$$P(|X - np| \geq k\sqrt{npq}) \leq \frac{1}{k^2} \quad [4.8]$$

que puede escribirse

$$P\left(\left|\frac{X}{n} - p\right| \geq k\sqrt{\frac{pq}{n}}\right) \leq \frac{1}{k^2} \quad [4.9]$$

Dado un número positivo cualquiera ϵ , siempre se puede determinar k de manera que sea

$$k\sqrt{\frac{pq}{n}} > \epsilon, \quad k > \epsilon\sqrt{\frac{n}{pq}} \quad [4.10]$$

con lo cual [4.9] se puede escribir

$$P\left(\left|\frac{X}{n} - p\right| > \epsilon\right) < \frac{pq}{\epsilon^2 n}. \quad [4.11]$$

Recordemos que X denota el número de éxitos en una sucesión de pruebas de Bernoulli, de manera que X/n es igual a la frecuencia relativa con que aparece el éxito en n pruebas. Por consiguiente, observando que el segundo miembro en [4.11] tiende a 0 para $n \rightarrow \infty$, queda demostrado el siguiente teorema fundamental.

Teorema de Bernoulli. *En una sucesión de pruebas de Bernoulli, dado un número positivo ϵ arbitrario, la probabilidad de que la frecuencia relativa del éxito en n pruebas difiera de la probabilidad p en una cantidad mayor que ϵ , tiende a cero cuando $n \rightarrow \infty$.*

Conviene adoptar la siguiente definición.

Definición 4.1. Una variable aleatoria Z_n , dependiente de un entero n , se dice que converge en probabilidad al límite p , si $P(|Z_n - p| > h)$ tiende a cero, para $n \rightarrow \infty$, cualquiera que sea $h > 0$.

Con esta definición, el teorema de Bernoulli se puede enunciar brevemente:

En toda sucesión de pruebas de Bernoulli, la frecuencia relativa X/n converge en probabilidad a p .

Este teorema es uno de los más importantes de la teoría de las probabilidades. Por su intermedio quedan vinculados los conceptos de probabilidad experimental o frecuencia y el de probabilidad teórica. Por ejemplo, en el lanzamiento de una moneda, la probabilidad de salir cara es $p = 1/2$. Si se hace el experimento n veces y se divide el número de veces que ha salido cara por n , el teorema de Bernoulli dice que este cociente converge (en probabilidad) a $1/2$. Decir "en probabilidad" significa que si bien nunca puede haber seguridad absoluta, ello ha de ocurrir con una probabilidad tan cerca de 1 como se quiera.

En vez de [4.11], es útil a veces la relación equivalente

$$P\left(\left|\frac{X}{n} - p\right| < \epsilon\right) \geq 1 - \frac{pq}{\epsilon^2 n}. \quad [4.12]$$

54

A veces no se conoce p para poder aplicar [4.11] ó [4.12]. En tal caso, observemos que $(p - q)^2 \geq 0$, de donde $p^2 + q^2 \geq 2pq$, y por tanto $p^2 + q^2 + 2pq \geq 4pq$, o sea $(p + q)^2 \geq 4pq$, y como $p + q = 1$, resulta que siempre se verifica la desigualdad

$$pq \leq \frac{1}{4} \quad [4.13]$$

de manera que, tanto en [4.11] como en [4.12], en el segundo miembro se puede sustituir pq por $1/4$.

Veamos los siguientes ejemplos.

Problema 4.2. *Se lanza una moneda 1.000 veces. Se desea una acotación de la probabilidad de que salga cara un número de veces comprendido entre 450 y 550.*

Solución. Se quiere que sea $0,45 \leq X/n \leq 0,55$, o sea $|X/1.000 - 0,5| \leq 0,05$; en [4.12] es $\epsilon = 0,05$, $n = 1.000$, $p = q = 1/2$, con lo cual se tiene

$$P\left(\left|\frac{X}{1.000} - 0,5\right| \leq 0,05\right) \geq 1 - \frac{1}{4 \cdot (0,05)^2 \cdot 1.000} = 0,9,$$

o sea que hay una probabilidad igual o superior al 90% de que el número de caras esté comprendido entre 450 y 550.

Problema 4.3. *¿Cuántas veces hay que lanzar un dado para que, con probabilidad $\geq 0,9$, la frecuencia relativa con que salga el 1 difiera de $1/6$ en no más de $0,01$?*

Solución. Se quiere que sea $|X/n - p| \leq 0,01$. Aplicando [4.12] en el caso $p = 1/6$, $q = 5/6$, $\epsilon = 0,01$, resulta que debe ser

$$1 - \frac{5 \cdot 10^4}{36n(0,01)^2} \geq 0,9$$

de donde, $n \geq 13.888$.

Problema 4.4. *Se quiere saber la frecuencia de fumadores en una cierta población (o sea, el cociente de dividir el número de fumadores por el total de la población). Para ello se eligen n personas al azar y se halla la frecuencia de las que fuman. Se desea saber qué valor debe tener n para que esta frecuencia no difiera de la real en más de $0,005$, con una probabilidad $\geq 0,95$.*

Solución. En este caso no se conoce p , de manera que tomaremos la acotación [4.13]. Siendo $\epsilon = 0,005 = 5 \cdot 10^{-3}$, se deduce de [4.12] que debe ser

$$1 - \frac{10^6}{4 \cdot 25 \cdot n} \geq 0,95,$$

de donde $n \geq 200.000$.

Esta acotación es excesiva. Más adelante, después de estudiar la distribución normal (problema 6.3) se verá que puede reducirse mucho. Las desigualdades [4.11] ó [4.12] no deben, casi nunca, aplicarse en este tipo de problemas, pues las acotaciones resultantes son siempre, en general, mucho peores que las que pueden obtenerse por otros métodos. Sin embargo, desde el punto de vista teórico estas desigualdades son fundamentales.

4. LEYES DE LOS GRANDES NUMEROS

El teorema de Bernoulli pertenece a un tipo general de teoremas conocidos con el nombre de "leyes de los grandes números". Ellos difieren entre sí en el grado de generalidad, pero son siempre teoremas límites que relacionan frecuencias con probabilidades o con valores medios. Vamos a dar otro de ellos.

Consideremos un dado. Si X es la variable aleatoria que indica los puntos de sus caras, es $E(X) = 3,5$. Supongamos que se lanza el dado n veces y se halla la "media experimental" $\bar{x}_n/n =$ suma total de los puntos obtenidos dividida por n . ¿Cuál es la probabilidad de que esta media experimental difiera de la teórica

$E(X) = 3, 5$ en menos de un número dado $\epsilon > 0$? Estamos ante una situación análoga a la del teorema de Bernoulli, sólo que en vez de la probabilidad se trata del valor medio.

Planteemos el problema en general. Sea $X_1, X_2, \dots, X_n, \dots$ una sucesión de variables aleatorias, dos a dos independientes, todas con la misma distribución de probabilidad de la variable aleatoria X . Sea $E(X_i) = E(X) = \alpha_1$, $\sigma^2(X_i) = \sigma^2(X) = \sigma^2$. Pongamos $Z_n = X_1 + X_2 + \dots + X_n$. Tendremos

$$E(Z_n) = n\alpha_1, \quad E\left(\frac{Z_n}{n}\right) = \alpha_1 \quad [4.14]$$

y por haber supuesto que las X_i son independientes dos a dos, también

$$\sigma^2(Z_n) = n\sigma^2, \quad \sigma^2\left(\frac{Z_n}{n}\right) = \frac{\sigma^2}{n}. \quad [4.15]$$

Aplicando la desigualdad de Tchebycheff a la variable aleatoria Z_n/n , tenemos

$$P\left(\left|\frac{Z_n}{n} - \alpha_1\right| \geq k\frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}$$

o bien, poniendo $k\sigma/\sqrt{n} = \epsilon$,

$$P\left(\left|\frac{Z_n}{n} - \alpha_1\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}, \quad \text{o bien} \quad P\left(\left|\frac{Z_n}{n} - \alpha_1\right| \leq \epsilon\right) \geq 1 - \frac{\sigma^2}{n\epsilon^2}. \quad [4.16]$$

Estas desigualdades permiten enunciar el siguiente teorema (que es otra "ley de los grandes números"):

Teorema de Bernoulli Generalizado. *Dada una sucesión X_1, X_2, \dots de variables aleatorias, dos a dos independientes, con una misma distribución de probabilidad y con media α_1 y varianza σ^2 finitas, se verifica (para todo $\epsilon > 0$)*

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{Z_n}{n} - \alpha_1\right| \geq \epsilon\right) = 0, \quad [4.17]$$

donde $Z_n = X_1 + X_2 + \dots + X_n$.

En otras palabras: *el límite, en probabilidad, de la media experimental Z_n/n , para $n \rightarrow \infty$, es igual a la media teórica α_1 .*

En este teorema la condición de que la varianza sea finita es excesiva, pero si se prescinde de ella, la demostración es más compleja (véase, por ejemplo, Feller⁽¹⁾).

Todavía el teorema puede generalizarse, suponiendo que las X_i tienen diferentes medias $E(X_i) = \alpha_i$ y diferentes varianzas σ_i^2 .

Entonces la suma Z_n tendrá cierta media $E(Z_n) = m_n$ y cierta varian-za s_n^2 . Bajo ciertas condiciones muy amplias, se verifica enton-ces que, para todo $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{Z_n - m_n}{n} \right| > \epsilon \right) = 0. \quad [4.18]$$

Cuando esta igualdad se cumple, se dice que la sucesión X_i sa-tisface la *ley débil* de los grandes números.

Si, todavía de manera más precisa, para cada $\epsilon > 0$, $\delta > 0$ existe un entero N tal que, con probabilidad $> 1 - \delta$, se cumplan, para todo $r > 0$, las desigualdades

$$\left| \frac{Z_n - s_n}{n} \right| < \epsilon, \quad n = N, N + 1, \dots, N + r, \quad [4.19]$$

se dice que la sucesión X_i satisface la *ley fuerte* de los grandes números.

La búsqueda de las condiciones mínimas para que una sucesión X_1, X_2, \dots de variables aleatorias cumpla la ley débil o la ley fuerte de los grandes números da lugar a cuestiones difíciles. Muchos teoremas al respecto, junto con ejercicios y comentarios, pueden verse en los libros de Feller,⁽¹⁾ Reny⁽²⁾ y Ríos⁽³⁾ citados en la bibliografía.

5

DISTRIBUCION DE POISSON

1. FUNCION DE PROBABILIDAD DE POISSON

Ya dijimos que la expresión $b(r;n,p)$ [4.2] de la función de probabilidad binomial es difícil de calcular directamente para valores un poco grandes de r y n . Se puede hacer mediante tablas apropiadas, pero muchas veces es preferible sustituir la expresión $b(r;n,p)$ por otra de mejor manejo para el cálculo y suficientemente aproximada en las aplicaciones.

Un primer caso en que esto es posible es para valores pequeños de p , tales que el producto np sea relativamente pequeño aun para valores bastante grandes de n . Planteemos el problema de buscar el límite de la función de probabilidad binomial para el caso en que p tiende a cero, al mismo tiempo que n tiende a infinito, de manera que el valor medio [4.4] se mantenga igual a una constante positiva λ , o sea,

$$np = \lambda. \quad [5.1]$$

Desde luego que se trata de un caso límite teórico, pues la probabilidad p en cualquier experimento tiene un valor fijo y, por tanto, el producto np crece con n . Sin embargo el límite nos dará un valor aproximado de $b(r;n,p)$ para los valores de n y p tales que su producto no difiera mucho de λ .

Se tiene

$$\begin{aligned} b(r;n,p) &= \binom{n}{r} p^r q^{n-r} = \binom{n}{r} \left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^{n-r} = \\ &= \frac{n(n-1) \dots (n-r+1)}{r! (1 - \lambda/n)^r} \left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^n = \\ &= \frac{\lambda^r (1 - 1/n)(1 - 2/n) \dots (1 - (r-1)/n)}{r! (1 - \lambda/n)^r} \left(1 - \frac{\lambda}{n}\right)^n. \end{aligned}$$

Para $n \rightarrow \infty$ (manteniendo fijo r), el numerador del segundo quebrado tiende a 1, puesto que es el producto de un número finito de factores, cada uno de los cuales tiende a 1. El denominador también tiende a 1. El último factor tiende a $e^{-\lambda}$ (véase el Apéndice II). Por tanto queda

$$\lim_{n \rightarrow \infty} b(r; n, p) = \frac{\lambda^r}{r!} e^{-\lambda} \quad [5.2]$$

donde se sobreentiende que n y p están ligados por la relación [5.1].

Se tiene así, siempre que $\lambda > 0$, una nueva función de probabilidad para una variable aleatoria X que tome los valores $0, 1, 2, \dots$. Esta función de probabilidad

$$P_r = \frac{\lambda^r}{r!} e^{-\lambda}, \quad r = 0, 1, 2, \dots \quad [5.3]$$

se llama la *función de probabilidad de Poisson*. Una variable aleatoria que pueda tomar los valores $r = 0, 1, 2, \dots$, con la probabilidad P_r , se llama una *variable de Poisson*. De acuerdo con la definición 3.2, la *función de distribución de Poisson* será

$$F(r) = \sum_{i=1}^r P_i. \quad [5.4]$$

Teniendo en cuenta cómo ha sido obtenida, resulta que la función P_r nos da un valor aproximado de $b(r; n, p)$ para valores pequeños de p . Prácticamente se considera que la aproximación es aceptable si $p < 0,1$ y $np < 5$. Por esto se llama también la función o la *ley de las pequeñas probabilidades*. La figura 3 representa la función $b(r; 10, 0,1)$ y la figura 4 la función $e^{-1}/r!$ pudiéndose apreciar la analogía.

Obsérvese que P_r sirve para variables aleatorias que pueden tomar una infinidad numerable de valores ($r = 0, 1, \dots$). Para estas variables, las definiciones de esperanza matemática, varianza y momentos son las mismas que las estudiadas en el capítulo 3 pero, en vez de sumas finitas, hay que tomar series.

Por ejemplo, se pueden comprobar, como ejercicio, las siguientes fórmulas:

$$\sum_{r=0}^{\infty} P_r = \sum_{r=0}^{\infty} e^{-\lambda} \frac{\lambda^r}{r!} = 1,$$

$$E(X) = \sum_{r=0}^{\infty} r \frac{\lambda^r}{r!} e^{-\lambda} = \lambda,$$

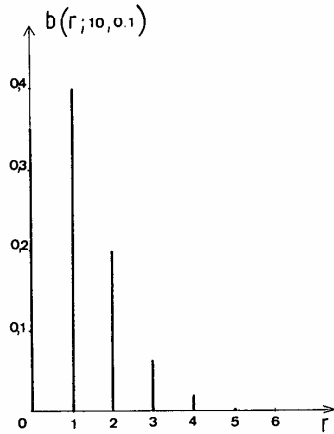


Fig. 3.

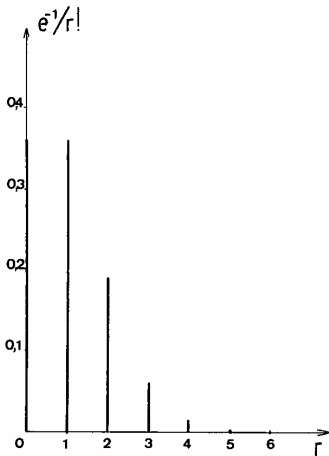


Fig. 4.

$$\sigma^2(X) = \sum_{r=0}^{\infty} (r-\lambda)^2 \frac{\lambda^r}{r!} e^{-\lambda} = \lambda.$$

Sin embargo, es más fácil calcular la media y la varianza de una variable de Poisson de la siguiente manera. Puesto que la media de una variable binomial es np , que ahora es la constante λ , y la varianza de la variable binomial es $npq = np(1-p) = \lambda(1-\lambda/n) \rightarrow \lambda$, resulta que la esperanza y la varianza de la variable aleatoria de Poisson de parámetro λ son

$$E(X) = \lambda, \quad \sigma^2(X) = \lambda. \quad [5.5]$$

La función generatriz de momentos de la variable de Poisson es

$$\Psi(t) = \sum_{r=0}^{\infty} \frac{(e^t \lambda)^r}{r!} e^{-\lambda} = e^{\lambda(e^t-1)} \quad [5.6]$$

y de aquí se deduce, nuevamente,

$$E(X) = \Psi'(0) = \lambda, \quad \sigma^2(X) = \Psi''(0) - \lambda^2 = \lambda.$$

62

Aplicación. Por el teorema 3.6, la expresión anterior de $\Psi(t)$ nos dice que si X_1, X_2, \dots, X_n son variables aleatorias de Poisson, con parámetros $\lambda_1, \lambda_2, \dots, \lambda_n$, la variable aleatoria suma $X = X_1 + X_2 + \dots + X_n$ es también una variable de Poisson con parámetro $\lambda_1 + \lambda_2 + \dots + \lambda_n$.

Problema 5.1. Se tiran dos dados 50 veces. Se desea hallar la probabilidad de que salgan dos 6 exactamente 5 veces.

Solución. La probabilidad de que salgan dos 6 es $1/36$. Por tanto, según el esquema binomial, la probabilidad pedida es

$$P_5 = b(5; 50, 1/36) = \binom{50}{5} \left(\frac{1}{36}\right)^5 \left(\frac{35}{36}\right)^{45}. \quad [5.7]$$

Este valor no es fácil de calcular directamente. Se pueden utilizar tablas apropiadas o aplicar la fórmula de Stirling (Apéndice II), pero es más cómodo aplicar la fórmula de Poisson para $np = 1,4 = \lambda$ (teniendo en cuenta que $p = 0,028 < 0,1$ y que $np < 5$). Utilizando la tabla de la función de probabilidad de Poisson (Apéndice III), resulta $P_5 = 0,011$. Si la fórmula [5.7] se hubiera calculado directamente, el resultado exacto sería $0,0099$. Vemos, pues, que la aproximación es buena.

Problema 5.2. Una fábrica produce ciertas piezas y se sabe que la probabilidad de que una pieza sea defectuosa es $p = 0,02$.

Se desea hallar la probabilidad de que, en un lote de 100 piezas, no haya piezas defectuosas, y también la probabilidad de que haya, a lo sumo, 3 piezas defectuosas.

Solución. Aplicando la función de Poisson para $\lambda = np = 2$, $r = 0$, resulta

$$P_0 = e^{-2} = 0,135$$

y la probabilidad de que haya a lo sumo 3 defectuosas será:

$$F(3) = P(r \leq 3) = P_0 + P_1 + P_2 + P_3 = \frac{19}{3} e^{-2} = 0,857\dots$$

Problema 5.3. *Un artillero dispara a un blanco y sabe que la probabilidad de acertar es $p = 1/100$. ¿Cuántos disparos tendrá que hacer para tener una probabilidad mayor que 90% de dar en el blanco por lo menos una vez?*

Solución. Aplicando directamente la función binomial, resulta que la probabilidad de dar en el blanco por lo menos una vez es la complementaria de no dar en el blanco vez alguna, o sea $1 - (0,99)^n$. Si se quiere que esta probabilidad cumpla la condición $1 - (0,99)^n > 0,9$, resulta $n > (2 - \log 99)^{-1} = 228$.

Como p es pequeña, en vez del cálculo directo se puede aplicar la aproximación de Poisson. La probabilidad de no dar vez alguna en el blanco es $e^{-\lambda}$, y la probabilidad de dar por lo menos una vez será $1 - e^{-\lambda}$. El problema impone que sea $1 - e^{-\lambda} > 0,9$, o sea, $0,1 > e^{-0,01 n}$, de donde $n > 100/\log e = 230$. Vemos, pues, que la aproximación de Poisson es muy buena.

2. DISTRIBUCION UNIFORME DE PUNTOS SOBRE UNA RECTA O DE SUCESOS EN EL TIEMPO

Un proceso muy corriente en que aparece la función de Poisson es el siguiente.

Supongamos un segmento A de longitud a y otro segmento B de longitud b contenido en A. Supongamos, además, que la probabilidad de que un punto x , dado al azar en A, pertenezca a B, sea igual a b/a , independientemente de la posición de B dentro de A. Si se dan al azar n puntos en A, la probabilidad de que r de ellos se hallen en B será $b(r; n, p)$. El cociente n/a es el número medio de puntos, dados al azar, por unidad de longitud. Sea $n/a = \lambda$, y supongamos que a crece hasta cubrir toda la recta, al mismo tiempo que n crece también de manera que el número medio λ de puntos por unidad de longitud permanezca constante. Será

$$np = n \frac{b}{a} = \frac{n}{a} b = \lambda b.$$

Por tanto se cumple la condición [5.1], con el parámetro λb en vez de λ , y para $n \rightarrow \infty$ resulta que la probabilidad de que r de los puntos dados al azar pertenezcan a B será

$$P_r = \frac{(\lambda b)^r}{r!} e^{-\lambda b}, \quad r = 0, 1, 2, \dots \quad [5.8]$$

Para $r=1$ y b pequeño, sea $b = \Delta x$; utilizando el desarrollo en serie de $e^{-\lambda \Delta x}$ (véase el Apéndice II) y despreciando los términos del orden de $(\Delta x)^2$, resulta que la *probabilidad de que un intervalo de longitud Δx contenga uno de los puntos distribuidos al azar* es

$$P = \lambda \cdot \Delta x. \quad [5.9]$$

El hecho de haber supuesto $p = b/a$ independiente de la posición del segmento B dentro del A significa que al extenderse A a toda la recta, la probabilidad de que uno de los puntos dados al azar caiga sobre B depende únicamente de la longitud de B, pero no de la posición de B sobre la recta. Esta suposición, en vista de la fórmula [5.8] a que se llega, se enuncia diciendo que los puntos están distribuidos sobre la recta según un *proceso de Poisson*. Precizando, sentamos la siguiente definición

Definición 5.1. Se dice que infinitos puntos están distribuidos sobre la recta según un *proceso de Poisson*, cuando la probabilidad de que un punto pertenezca a un segmento B (de longitud b), condicionada a que pertenezca al segmento A (de longitud a), tal que $B \subset A$, es b/a , cualquiera que sea la posición de A sobre la recta y la posición de B dentro de A.

Según esta definición, el resultado [5.8] se enuncia de acuerdo con el siguiente teorema

Teorema 5.1. *Supuestos dados sobre la recta infinitos puntos al azar según un proceso de Poisson, con un promedio de λ puntos por unidad de longitud, la probabilidad de que r de ellos pertenezcan a un segmento dado de longitud b está dada por [5.8].*

Consideremos ahora un punto cualquiera M de la recta y busquemos la probabilidad de que el r -ésimo punto, a partir de M, de un proceso de Poisson se encuentre a una distancia comprendida entre x y $x + \Delta x$ de M. Según [5.8], la probabilidad de que en el segmento de longitud x haya $r - 1$ puntos es $e^{-\lambda x} (\lambda x)^{r-1} / (r-1)!$, y la probabilidad de que en el segmento Δx haya un punto, es $\lambda \cdot \Delta x$ (según [5.9]). La probabilidad buscada es igual al producto de estas dos, o sea, $f_r(x) \Delta x$, donde

$$f_r(x) = e^{-\lambda x} \frac{\lambda^r x^{r-1}}{(r-1)!} \quad [5.10]$$

De aquí se deduce, por ejemplo, que la *distancia media* entre un punto cualquiera y el r -ésimo punto del proceso que le sigue es

$$E(x) = \int_0^{\infty} x f_r(x) dx = \frac{\lambda^r}{(r-1)!} \int_0^{\infty} e^{-\lambda x} x^r dx \quad [5.11]$$

Poniendo $\lambda x = t$, $\lambda dx = dt$, y recordando la integral que define la función $\Gamma(x)$ (véase el Apéndice II), resulta

$$E(x) = \frac{r}{\lambda} \quad [5.12]$$

Naturalmente que en todo lo anterior, la recta puede representar el tiempo t y cada punto de la misma ser un instante determinado. Entonces el esquema anterior se presta a muchas interpretaciones, como vamos a ver mediante algunos ejemplos.

Problema 5.4. *Se sabe que durante ciertas horas las llamadas telefónicas a una central están distribuidas al azar según un proceso de Poisson, con un promedio de 3 llamadas por minuto. Se desea hallar la probabilidad: a) de que transcurran 4 minutos sin llamadas; b) de que en 1 minuto haya exactamente 3 llamadas; c) de que en 2 minutos haya 8 llamadas.*

Solución. Se aplica siempre [5.8]. Resulta: a) $\lambda = 3$, $b = 4$, $r = 0$, por tanto, $P = e^{-12} < 10^{-4}$; b) $\lambda = 3$, $b = 1$, $r = 3$; por tanto, $P = (27/6)e^{-3} = 0,224$; c) $\lambda = 3$, $b = 2$, $r = 8$; por tanto, $P = (6^8/8!)e^{-6} = 0,103$.

Problema 5.5. *Por un punto de una carretera pasa un promedio de 5 automóviles por minuto. Suponiendo que el paso de los mismos sigue un proceso de Poisson, ¿cuál es la probabilidad de que en un minuto no pase automóvil alguno?*

Solución. Aplicando [5.8], resulta $P = e^{-5} = 0,0067$.

Problema 5.6. *Supongamos (lo cual se adapta bastante a la realidad) que los momentos en que tienen lugar nacimientos se reparten al azar durante todas las horas del día. Supongamos también que en una ciudad determinada tenga lugar por término medio un nacimiento cada 65 minutos. ¿Cuál es la probabilidad de que durante 2 horas no tenga lugar ningún nacimiento?*

Solución. Basta aplicar [5.8] para $r = 0$, $\lambda = 1/65$, $b = 120$. Resulta $p = e^{-1,83} = 0,160\dots$

Problema 5.7. *Se sabe que el número de partículas emitidas por una substancia radiactiva, con el tiempo, obedece a una ley de Poisson. Supongamos que la emisión ocurre a razón de 30 partículas por minuto. Se desea saber la probabilidad de que durante 7,5 segundos sean emitidas exactamente 3 partículas.*

Solución. Se aplica [5.8] para $\lambda = 0,5$ (número de partículas emitidas por segundo), $b = 7,5$, y resulta $P = (3,75)^3 e^{-3,75} (1/6)$. La tabla de la función de Poisson da entonces $P = 0,20$.

Caso del plano o del espacio. Si en vez de una recta se considera el plano o el espacio, el razonamiento que condujo a la fórmula [5.8] sigue siendo válido.

Supongamos, por ejemplo, el caso del espacio. Sea B un dominio de volumen b , contenido en un dominio A de volumen a . Representétese la probabilidad de que un punto dado al azar en A pertenezca a B, por b/a (proceso de Poisson). Dados al azar n puntos en A, la probabilidad de que r de ellos estén en B, es $b(r;n, p)$, y suponiendo que A crece hasta llenar todo el espacio, al mismo tiempo que también n crece de manera que la densidad media $n/a = \lambda$ se mantenga constante ($\lambda =$ número medio de puntos por unidad de volumen), entonces, pasando al límite, resulta que la probabilidad de que B contenga exactamente r puntos está dada por la misma fórmula [5.8]. Por ejemplo, es típico el siguiente problema:

Problema 5.8. *Se sabe que un líquido contiene ciertas bacterias a razón de 4 bacterias por cm^3 . Se desea saber la probabilidad de que una muestra de 1 cm^3 no contenga bacteria alguna y también la probabilidad de que en $1/2 \text{ cm}^3$ haya por lo menos una bacteria.*

Solución. En el caso de la primera pregunta es $\lambda = 4$, $b = 1$, $r = 0$; por tanto $P = e^{-4} = 0,0183$. En el de la segunda, puesto que la probabilidad de que no contenga alguna bacteria es e^{-2} , la probabilidad de que contenga por lo menos una bacteria será $P = 1 - e^{-2} = 0,864$.

3. EL PROBLEMA DE LAS FILAS DE ESPERA

Un proceso aleatorio importante que tiene varias interpretaciones es el llamado de manera general "proceso de nacimiento y muerte" o "proceso de inmigración y emigración". Se trata de estudiar el comportamiento de un conjunto de elementos (partículas, personas, llamadas telefónicas, ...) en el cual, con el transcurso de tiempo, se incorporan nuevos elementos (según cierta probabi-

lidad) y desaparecen otros (también con cierta probabilidad). Para fijar las ideas, aunque las fórmulas son las mismas en cualquier interpretación, vamos a estudiar el problema de las "colas" o filas de espera.

Consideremos una sola ventanilla de atención de clientes, los cuales, a medida que llegan, se van colocando en la cola o fila de espera, hasta el momento de ser atendidos. Supongamos que la llegada de los clientes sigue un proceso de Poisson, de parámetro λ . Esto quiere decir que la probabilidad de que lleguen r clientes en un intervalo de tiempo b está dada por [5.8], y la probabilidad de que llegue un cliente en el intervalo Δt es $\lambda \Delta t$ [5.9], es decir, depende sólo del intervalo Δt , pero no del momento en que este intervalo se considera (o sea, no se tiene en cuenta el caso de que en ciertas horas la afluencia de clientes puede ser mayor que en otras). Igualmente, se hace la hipótesis de que la probabilidad de que en el intervalo Δt un cliente termine de ser atendido y abandone la fila es igual a $\kappa \Delta t$, siendo κ otra constante. Ello equivale a suponer que la salida de los clientes sigue también una ley de Poisson, de parámetro κ . Recordemos que λ significa el número medio de clientes que llegan por unidad de tiempo, y κ el número medio de clientes que son atendidos por unidad de tiempo.

Sea $P_n(t)$ la probabilidad de que en el instante t la cola tenga n clientes. Puesto que la llegada y la salida se consideran independientes, tendremos (salvo términos en $(\Delta t)^2$ que no tenemos en cuenta)

$$P_n(t) = P_{n-1}(t - \Delta t)\lambda \Delta t + P_{n+1}(t - \Delta t)\kappa \Delta t + P_n(t - \Delta t)(1 - \lambda \Delta t - \kappa \Delta t), \quad [5.13]$$

lo cual expresa que la probabilidad de que en el instante t haya n clientes es igual a la probabilidad de que en el instante $t - \Delta t$ haya $n - 1$ por la probabilidad de que llegue uno nuevo, más la probabilidad de que en el instante $t - \Delta t$ haya $n + 1$, por la probabilidad de que uno se retire, más la probabilidad de que en el instante $t - \Delta t$ haya n clientes, por la probabilidad de que en el intervalo considerado no salga ni llegue ninguno. La probabilidad de que en el intervalo $(t - \Delta t, t)$ se produzca más de un suceso (llegada o salida de clientes) es infinitésima de orden igual o superior a $(\Delta t)^2$ y por tanto no se ha tenido en cuenta en [5.13].

Para $n = 0$, [5.13] se reduce a

$$P_0(t) = P_1(t - \Delta t)\kappa \Delta t + P_0(t - \Delta t)(1 - \lambda \Delta t). \quad [5.14]$$

Pasando al primer miembro de [5.13] y [5.14] el término que no contiene Δt , dividiendo ambos miembros por Δt y pasando al límite para $\Delta t \rightarrow 0$, quedan las ecuaciones diferenciales

$$\frac{dP_n(t)}{dt} = \lambda P_{n-1}(t) + \kappa P_{n+1}(t) - (\lambda + \kappa)P_n(t) \quad [5.15]$$

$$\frac{dP_0(t)}{dt} = \kappa P_1(t) - \lambda P_0(t).$$

Estas son las ecuaciones fundamentales del proceso.

Supongamos t lo bastante grande para que desaparezca la influencia de los primeros momentos en que la cola se inicia. Es decir, supongamos un estado estacionario, con las P_n independientes de t . Entonces, en las ecuaciones [5.15], los primeros miembros son nulos y los segundos son independientes de t , quedando

$$\lambda P_{n-1} + \kappa P_{n+1} - (\lambda + \kappa)P_n = 0, \quad \kappa P_1 - \lambda P_0 = 0. \quad [5.16]$$

Escribiendo estas ecuaciones para $n=1, 2, \dots, n-1$ y sumando, queda $\kappa P_n = \lambda P_{n-1}$, de donde

$$P_n = \rho^n P_0, \quad \text{donde } \rho = \frac{\lambda}{\kappa}. \quad [5.17]$$

Puesto que debe ser $\sum_0^{\infty} P_n = 1$, resultan las condiciones $\rho < 1$ y $P_0 = 1 - \rho$, y se obtiene

$$P_n = (1 - \rho)\rho^n. \quad [5.18]$$

68

Esta es la probabilidad de que la cola tenga n clientes. En particular, la probabilidad de encontrar la ventanilla desocupada es $P_0 = 1 - \rho$.

El *número medio* de clientes será

$$E(n) = \sum_0^n n P_n = (1 - \rho)\rho \sum_0^{\infty} n \rho^{n-1} = \frac{\rho}{1 - \rho} \quad [5.19]$$

donde hemos utilizado que $\sum_0^{\infty} n \rho^{n-1} = (1 - \rho)^{-2}$, como resulta al derivar la suma $\sum_0^{\infty} \rho^n = (1 - \rho)^{-1}$.

En todas estas fórmulas se supone $\rho < 1$.

Sea τ el tiempo que transcurre desde que un cliente llega hasta que sale, después de haber sido atendido. Sea $f(\tau)\Delta\tau$ la probabilidad de que este tiempo esté comprendido entre τ y $\tau + \Delta\tau$. Si al llegar, la cola tiene n clientes, según la fórmula [5.10] la probabilidad de que en el tiempo transcurrido hasta el momento de salir (tiempo de atender a los n clientes de la cola, más el tiempo de atender al que llega) esté comprendido entre τ y $\tau + \Delta\tau$, es $f(\tau | n)\Delta\tau$, donde

$$f(\tau | n) = e^{-\kappa\tau} \kappa^{n+1} \frac{\tau^n}{n!}. \quad [5.20]$$

Obsérvese, efectivamente, que si la recta considerada en la fórmula [5.10] es el eje de los tiempos, la distancia x que allí figura es ahora el tiempo τ , dado κ . Por tanto

$$\begin{aligned} f(\tau) &= \sum_{n=0}^{\infty} f(\tau | n) P_n = \sum_{n=0}^{\infty} e^{-\kappa\tau} \kappa^{n+1} \tau^n \frac{\rho^n (1-\rho)}{n!} = \\ &= \kappa(1-\rho) e^{-\kappa\tau} \sum_{n=0}^{\infty} \frac{(\kappa\tau\rho)^n}{n!} = \kappa(1-\rho) e^{-(1-\rho)\kappa\tau}. \end{aligned} \quad [5.21]$$

O sea: la función de densidad de probabilidad del tiempo de espera (desde que el cliente llega hasta que sale) es la [5.21] para $\tau \geq 0$ y $f = 0$ para $\tau < 0$.

El tiempo medio de espera será

$$E(\tau) = \int_0^{\infty} \tau f(\tau) d\tau = \frac{1}{\kappa(1-\rho)} = \frac{1}{\kappa - \lambda}. \quad [5.22]$$

Si se desea el tiempo medio de espera "hasta llegar a la ventanilla", teniendo en cuenta que el tiempo medio de atención es $1/\kappa$ (puesto que κ es el número medio de clientes atendidos por unidad de tiempo), resulta

$$E(\tau_1) = E(\tau) - \frac{1}{\kappa} = \frac{\rho}{\kappa(1-\rho)}. \quad [5.23]$$

Este tiempo medio de espera "hasta llegar a la ventanilla" es útil, por ejemplo, para muchas personas que sólo se impacientan hasta el momento de llegar a la ventanilla, pero que luego no les importa el tiempo durante el cual son atendidos y no hacen nada para abreviarlo.

Ejemplo 1. Una oficina debe atender un promedio de 200 clientes que llegan al azar cada 4 horas, y se quiere que la longitud media de la cola no sea superior a 3 clientes. ¿Cuántos clientes por hora debe ser capaz de atender la oficina?

Solución. Es $\lambda = 50$, $E(n) = 3$. Por tanto [5.19] da $\kappa = 67$ clientes por hora.

Ejemplo 2. Si cada ventanilla de una oficina puede atender a 40 clientes por hora y se sabe que acuden en promedio 200 clientes por hora, ¿cuántas ventanillas deben habilitarse para que la longitud media de la cola no supere a los 3 clientes por ventanilla?

Solución. Llamando v al número de ventanillas, es $\lambda = 200$, $\kappa = 40v$, $E(n) = 3v$, y por tanto, se tiene la ecuación $3v^2 - 15v - 5 = 0$, donde $v = 6$.

Ejemplo 3. *Un doctor sabe que el tiempo medio de que dispone para atender a cada paciente es de 20 minutos. Suponiendo que los pacientes llegan al azar, ¿cuántos pacientes debe citar por hora para que el promedio de los que esperan no sea superior a dos?*

Solución. Dos pacientes por hora.

Ejemplo 4. *En una estación de servicio llega un auto cada 6 minutos y tarda 4 minutos en ser atendido. Se pide: a) la probabilidad de que el auto que llega no tenga que esperar y b) el tiempo medio de espera desde que llega hasta el momento en que es atendido.*

Solución. a) $P_0 = 1 - (2/3) = 1/3$; b) $E(\tau_1) = 12 - 4 = 8$ minutos.

Ejemplo 5. *El tiempo medio de atención de una ventanilla es de 10 minutos y llega un cliente cada 11 minutos. Hállese la longitud de la cola y en cuánto disminuye la misma si el tiempo de atención se reduce a 9 minutos.*

Solución. La longitud media de la cola en el primer caso es $E(n) = 10$, y en el segundo, $E(n) = 5$, o sea, se reduce a la mitad. Esto prueba la rapidez con que disminuye la cola al aumentar la eficiencia del servicio.

6

DISTRIBUCION NORMAL. VARIABLES ALEATORIAS CONTINUAS

1. APROXIMACION DE LA DISTRIBUCION BINOMIAL POR LA NORMAL

Hemos visto que suponiendo $np = \lambda = \text{constante}$, la distribución binomial tiende a la de Poisson. Consideremos ahora el caso en que p no sea demasiado pequeño y que tanto n como r sean grandes. En tal caso se puede demostrar el siguiente teorema

Teorema de De Moivre - Laplace. *Si $r \rightarrow \infty$ y $n \rightarrow \infty$ de manera que*

$$\left| \frac{r - np}{\sqrt{npq}} \right| \rightarrow x < \infty, \quad [6.1]$$

entonces se verifica que

$$\sqrt{npq} \binom{n}{r} p^r q^{n-r} \rightarrow \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad [6.2]$$

No se da la demostración de este teorema por exigir cálculos un poco largos y delicados. Ella se puede hallar en los libros de Feller⁽¹⁾ o Reny.⁽²⁾ En realidad, la condición [6.1] se puede reemplazar por otra menos exigente. En efecto, se puede demostrar que la relación asintótica [6.2] es válida con sólo suponer que n y r tienden a ∞ de manera tal que, siendo p constante, sea $\lim (r - np)^2/n^2 = 0$ (véase, Feller⁽¹⁾).

La fórmula [6.2] permite calcular $b(r; n, p)$ por la fórmula aproximada

$$b(r; n, p) = \binom{n}{r} p^r q^{n-r} \sim \frac{1}{\sqrt{2\pi npq}} e^{-x^2/2} \quad [6.3]$$

donde

$$x = \frac{r - np}{\sqrt{npq}} \quad [6.4]$$

Esta aproximación de $b(r; n, p)$ se llama la *aproximación por la función normal*, de la que nos ocuparemos a continuación. Los intervalos de valores n, r, p , para los cuales la aproximación normal

[6.3] es buena, y se pueden tomar ambos miembros como iguales, no son fáciles de establecer. En general, se suele considerar que la aproximación [6.3] es buena para $p > 0,1$ y $np > 5$.

Por ejemplo, es $b(20;25,0,6) = 0,029$ y, usando la aproximación dada por [6.3], resulta $b(20;25,0,6) = 0,022$. Igualmente es $b(15;100,0,2) = 0,048$ y la aproximación [6.3] da $b(15;100,0,2) = 0,046$.

Es clásico un experimento de Galton para comprobar la fórmula [6.3] para el caso $p = 1/2$. Consiste en una tabla como se ilustra en la figura 5, colocada verticalmente o inclinada, por cuya parte superior se introducen numerosas bolillas (perdigones), las cuales, al caer, se encuentran con obstáculos que las obligan a desviarse hacia la derecha o hacia la izquierda, con probabilidad $1/2$. En el caso de la figura, cada bolilla se encuentra con 6 obstáculos ($n = 6$). Las bolillas que se han desviado r veces hacia la izquierda y $6 - r$ hacia la derecha van a parar a la casilla r ($r = 0,1,2, \dots, 6$). La probabilidad de que una bolilla vaya a parar en la casilla r es

$$Pr = \binom{6}{r} \left(\frac{1}{2}\right)^6 \sim \frac{1}{\sqrt{3\pi}} e^{-(r-3)^2/3}. \quad [6.5]$$

72

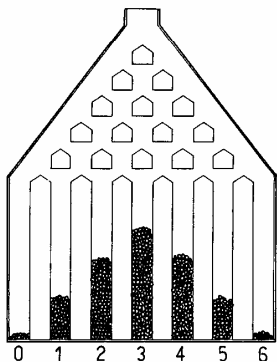


Fig. 5.

Haciendo el experimento con N bolillas, el número más probable de bolillas que van a parar a la casilla r es igual al valor medio NP_r . Efectivamente, la experiencia prueba que ellas van tomando la forma de campana expresada por la curva que indica el segundo miembro de [6.5], como muestra la figura 5. Esta curva de campana es la que vamos a estudiar en lo que sigue, pero antes conviene dar algunas definiciones sobre variables aleatorias continuas.

2. VARIABLES ALEATORIAS CONTINUAS

Hasta ahora hemos considerado variables aleatorias que pueden tomar un número finito de valores (variables aleatorias finitas) o bien, por extensión natural, variables aleatorias que pueden tomar una infinidad numerable de valores, como la variable de Poisson. Se dice que estas variables son *discretas*.

En ambos casos, representando sobre el eje x los valores x_1, x_2, \dots, x_n , de la variable aleatoria X y tomando sobre cada x_i el valor $f(x_i)$ de la probabilidad correspondiente, se tiene la gráfica de la función de probabilidad de X , que es un conjunto discreto de puntos cuyas coordenadas son $(x_i, f(x_i))$, $i = 1, 2, 3, \dots$. Así se hizo en las figuras 1, 2, 3 y 4.

El resultado [6.3], donde x puede tomar cualquier valor real, conduce a estudiar variables aleatorias el número de cuyos valores no es finito ni numerable. Para simplificar, vamos a suponer que son variables que pueden tomar todos los valores reales, desde $-\infty$ hasta $+\infty$, atribuyendo la probabilidad 0 al conjunto de los valores que no puede tomar. Se llaman variables aleatorias continuas. Recordemos que para el caso finito, la definición de variable aleatoria va ligada a la definición de función de probabilidad. En el caso general es mejor asociarla a la función de distribución. Suponiendo un espacio de probabilidad general (E, \mathbf{B}, P) , se establece la siguiente definición.

Definición 6.1. Se llama *variable aleatoria* definida en un espacio de probabilidad (E, \mathbf{B}, P) a toda función $X: E \rightarrow \mathbf{R}$ de E en el conjunto de los números reales, tal que, para todo $x \in \mathbf{R}$, el conjunto de elementos $a \in E$ para los cuales $X(a) \leq x$, pertenece a \mathbf{B} .

73

La última condición nos dice que el conjunto de elementos $a \in E$ para los cuales $X(a) \leq x$, tiene asignada una probabilidad, que representaremos abreviadamente por

$$F(x) = P(X \leq x), \quad [6.6]$$

y se llama la *función de distribución* de la variable aleatoria X . Se dice que $F(x)$ es igual a la probabilidad de que X tome un valor $\leq x$.

Observemos que esta definición de variable aleatoria es aplicable también a variables aleatorias finitas. En efecto, en tal caso, suponiendo la ordenación $x_1 < x_2 < \dots < x_n$ y llamando E_i al conjunto de elementos a de E tales que $X(a) \leq x_i$, resulta que el conjunto $X^{-1}(x_i)$ es $E_i \cap \bar{E}_{i-1}$, y por tanto, si E_i y E_{i-1} pertenecen a \mathbf{B} , también pertenece a \mathbf{B} el conjunto $X^{-1}(x_i)$.

De [6.6] se deducen inmediatamente las siguientes propiedades de toda función de distribución: 1. $F(x)$ es monótona no decreciente; 2. $F(-\infty) = 0$ (probabilidad del conjunto vacío) y $F(+\infty) = 1$ (probabilidad de E). Además

$$P(a < x \leq b) = F(b) - F(a). \quad [6.7]$$

En el caso de una variable aleatoria discreta, la función de distribución $F(x)$ es una función escalonada, expresada por sumas de la forma [3.4]. Vamos a considerar ahora el caso en que $F(x)$ es una función continua, en cuyo caso diremos que la correspondiente variable aleatoria es *continua*. $F(x)$ puede ser una función complicada, pero, para simplificar, vamos a considerar únicamente *funciones de distribución que puedan expresarse mediante una integral de la forma*

$$F(x) = \int_{-\infty}^x f(x)dx, \quad [6.8]$$

donde f es llamada *función de densidad o de densidad de probabilidad* de la variable aleatoria X . También por simplicidad, supondremos que f es continua o, por lo menos, "continua por pedazos", o sea, compuesta de un número finito o de una infinidad numerable de pedazos continuos.

De [6.7] y [6.8] se deduce

$$P(a < x \leq b) = \int_a^b f(x)dx \quad [6.9]$$

y las mencionadas propiedades de F nos dicen que toda función de densidad cumple las condiciones

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x)dx = 1. \quad [6.10]$$

Obsérvese que la condición [6.9] excluye el caso de las probabilidades discretas. En efecto, en este caso según [6.9], *la probabilidad de que X tome exactamente un valor dado x_0 es siempre nula*. Es el caso que ya observamos en el capítulo 1.1 donde la probabilidad nula no significa imposibilidad. Esto hace que sea lo mismo considerar la probabilidad $P(a \leq x \leq b)$ que la $P(a < x < b)$.

Veamos la interpretación geométrica de estos conceptos.

Sea f una función de densidad. Su gráfica será una curva como la de la figura 6 (que puede ser discontinua, como en el punto x_0). El área limitada entre la curva y el eje x es 1. La función de distribución F nos da, para cada x , el área comprendida entre el eje x y la curva, desde $-\infty$ hasta la ordenada correspondiente al punto x (área rayada en la figura). Esta área $F(x)$ representa la probabilidad de que X tome un valor $\leq x$.

Las definiciones de esperanza matemática, varianza y momentos de una variable aleatoria continua son análogas a las dadas para variables aleatorias discretas, sólo que se debe sustituir las sumas por integrales, es decir:

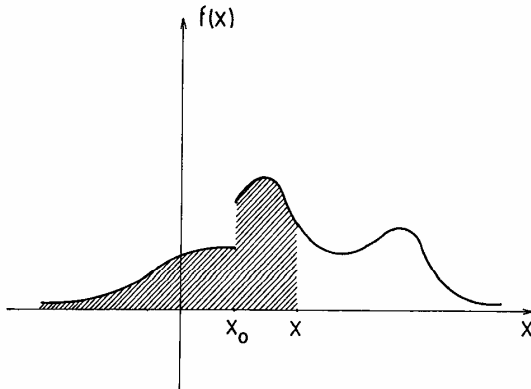


Fig. 6.

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx, \quad \sigma^2(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx, \quad [6.11]$$

y análogamente se definen los momentos de orden superior.

La función generatriz de momentos se define también como la esperanza de la función e^{xt} , o sea:

$$\Psi(t) = \int_{-\infty}^{\infty} e^{xt} f(x)dx. \quad [6.12]$$

3. FUNCION DE DENSIDAD DE PROBABILIDAD Y FUNCION DE DISTRIBUCION NORMAL

Consideremos una función de densidad de la forma

$$f(x) = a e^{-b(x-a)^2} \quad \text{con } a, b > 0. \quad [6.13]$$

La primera condición [6.10] se cumple, y la segunda conduce a la relación $a\sqrt{\pi/b} = 1$ (véase el Apéndice II). La esperanza y la varianza de la variable aleatoria X con la función de densidad de probabilidad $f(x)$ valen (Apéndice II)

$$E(X) = c = a, \quad \sigma^2(X) = \frac{1}{2b} \quad [6.14]$$

de manera que poniendo de manifiesto estos valores, [6.13] se escribe

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2} \quad [6.15]$$

Definición 6.2. Toda función de la forma [6.15] se llama una *función de densidad de probabilidad normal*. Las variables aleatorias con una tal función de densidad de probabilidad se llaman *normales* y la correspondiente función de distribución se llama una función de distribución *normal*.

Se dice que la función de densidad de probabilidad o la variable aleatoria o la función de distribución de la definición anterior son del tipo $N(\alpha, \sigma^2)$.

La función generatriz de momentos de la función [6.15] es (véase el Apéndice II)

$$\Psi(t) = e^{\alpha t + \frac{1}{2} \sigma^2 t^2} \quad [6.16]$$

Una aplicación importante es la siguiente. Si X e Y son variables aleatorias normales independientes de medias α_x, α_y , y varianzas σ_x^2, σ_y^2 , por el teorema 3.6, la función generatriz de momentos de la variable $X + Y$ será

$$\Psi_{X+Y}(t) = \Psi_X(t) \Psi_Y(t) = e^{(\alpha_x + \alpha_y)t + \frac{1}{2}(\sigma_x^2 + \sigma_y^2)t^2}. \quad [6.17]$$

76

Comparando con [6.16] resulta

Teorema 6.1. *La variable aleatoria suma de dos variables aleatorias normales independientes de tipo $N(\alpha_x, \sigma_x^2), N(\alpha_y, \sigma_y^2)$ es otra variable aleatoria normal de tipo $N(\alpha_x + \alpha_y, \sigma_x^2 + \sigma_y^2)$.*

Evidentemente, el teorema se generaliza de inmediato a la suma de n variables aleatorias normales independientes. Más generalmente, dadas n variables aleatorias normales independientes X_i de tipos $N(\alpha_i, \sigma_i^2)$, la combinación lineal $\sum a_i X_i$ es otra variable aleatoria normal de tipo $N(\sum a_i \alpha_i, \sum a_i^2 \sigma_i^2)$.

Particularmente interesante es el caso $N(0, 1)$, que corresponde a la función de densidad de probabilidad

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad [6.18]$$

y a la función de distribución

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx.$$

Estas funciones $\varphi(x), \Phi(x)$ han sido tabuladas. Tablas más o menos extensas se encuentran en todos los libros de cálculo de

probabilidades. En el Apéndice III se da un ejemplo de estas tablas.

La gráfica de la función $\varphi(x)$ tiene la forma de campana de la figura 7. $\Phi(x)$ representa el área entre la curva y el eje x desde $-\infty$ hasta el punto x . Obsérvese que, por simetría, el área rayada desde $-\infty$ hasta $-x$ es igual al área rayada desde x hasta ∞ , y como el área total es 1, resulta

$$\Phi(-x) = 1 - \Phi(x). \quad [6.20]$$

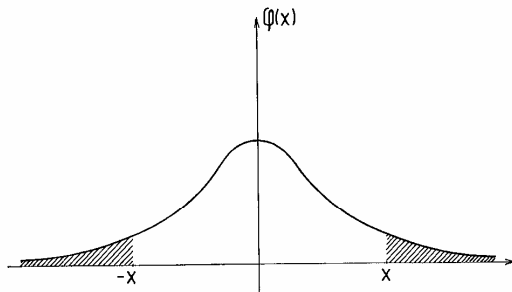


Fig. 7.

Esta relación hace que sea suficiente tabular $\Phi(x)$ para valores positivos de x . Un detalle que conviene tener en cuenta para no proseguir con cálculos innecesarios es que

$$\Phi(x) < 0,001 \quad \text{para } x < -3,1$$

$$\Phi(x) > 0,999 \quad \text{para } x > 3,1$$

de manera que, prácticamente, basta conocer los valores de $\Phi(x)$ en el intervalo $[0, 3]$.

Para calcular la función de densidad de probabilidad o la función de distribución de una variable normal $N(\alpha, \sigma^2)$ mediante las correspondientes de $N(0, 1)$ basta hacer el cambio de variable

$$X' = \frac{X - \alpha}{\sigma} \quad [6.21]$$

con lo cual la función de distribución $F(x)$ se transforma de la siguiente manera:

$$F(x) = \int_{-\infty}^x f(x) dx = \int_{-\infty}^{(x-\alpha)/\sigma} \varphi(x') dx' = \Phi\left(\frac{x-\alpha}{\sigma}\right). \quad [6.22]$$

Por tanto, recordando [6.7], se tiene el importante teorema siguiente.

Teorema 6.2. *Para una variable aleatoria normal de esperanza α y varianza σ^2 , la probabilidad de que esté contenida en el intervalo $[a, b]$ viene expresada por*

$$P(a < X \leq b) = \Phi\left(\frac{b-\alpha}{\sigma}\right) - \Phi\left(\frac{a-\alpha}{\sigma}\right). \quad [6.23]$$

Como la función Φ está tabulada, esta fórmula permite calcular $P(a < X \leq b)$ para toda variable aleatoria normal.

En particular, aplicando [6.20] resulta:

La probabilidad de que la variable aleatoria normal X de tipo $N(\alpha, \sigma^2)$ esté comprendida en el intervalo $[\alpha - \epsilon, \alpha + \epsilon]$ se expresa por

$$P(\alpha - \epsilon \leq X \leq \alpha + \epsilon) = 2\Phi\left(\frac{\epsilon}{\sigma}\right) - 1. \quad [6.24]$$

Poniendo $\epsilon = \alpha x$, esta igualdad puede también escribirse en la forma

$$P(\alpha - \alpha x \leq X \leq \alpha + \alpha x) = 2\Phi(x) - 1. \quad [6.25]$$

78

En general, son de interés las probabilidades que tienen los valores 0,90, 0,95, 0,99, ó 0,999. Como se deduce de las tablas, estos valores corresponden a $x = 1,65$, $x = 1,96$, $x = 2,58$, $x = 3,29$, de manera que, para toda variable normal $N(\alpha, \sigma^2)$, se cumple

$$\begin{aligned} P(\alpha - 1,65\sigma < X \leq \alpha + 1,65\sigma) &= 0,90 \\ P(\alpha - 1,96\sigma < X \leq \alpha + 1,96\sigma) &= 0,95 \\ P(\alpha - 2,58\sigma < X \leq \alpha + 2,58\sigma) &= 0,99 \\ P(\alpha - 3,29\sigma < X \leq \alpha + 3,29\sigma) &= 0,999. \end{aligned} \quad [6.26]$$

Recordemos que en todas estas fórmulas el signo \leq puede sustituirse por $<$, puesto que la probabilidad de que X tome un determinado valor es nula.

Volvamos ahora a la igualdad aproximada [6.3] que se deduce del teorema de De Moivre-Laplace. Siendo la esperanza de la función binomial igual a np y la varianza igual a npq , el último término de [6.3] es de la forma [6.15] y, por tanto, resulta que la función de probabilidad binomial, en el límite, es una función de densidad normal. Se dice que la función binomial es "asintóticamente normal". Por tanto, como ya observamos, tomando el límite como valor aproximado, resulta que, para valores grandes de n y r , y no demasiado

pequeños de p , se podrá calcular $b(r; n, p)$ mediante las tablas de la función $\varphi(x)$, después de hacer el cambio [6.21], que es el mismo [6.4]. Sin embargo, más importante que calcular el valor aproximado de $b(r; n, p)$, que para r y n grandes es siempre un valor pequeño, es el cálculo de la probabilidad de que la variable binomial X esté comprendida entre ciertos límites dados.

Consideremos, por ejemplo, el caso $b(r; 10, 0, 4)$ de la figura 8, y supongamos que se quiere hallar la probabilidad de que $3 \leq X \leq 6$. Será

$$P(3 \leq X \leq 6) = b(3; 10, 0, 4) + b(4; 10, 0, 4) + b(5; 10, 0, 4) + b(6; 10, 0, 4). \quad [6.27]$$

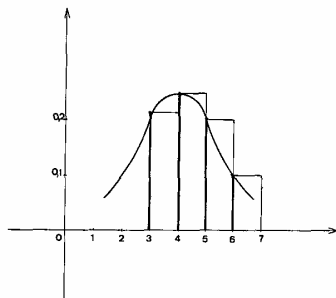


Fig. 8.

Esta suma es el área de los rectángulos de la figura 8, puesto que cada sumando es la altura de un rectángulo de base unidad. Considerando la aproximación [6.3], se tiene la curva

$$y = \frac{1}{\sqrt{2\pi} \sqrt{2,4}} e^{-(x-4)^2/4,8}, \quad [6.28]$$

representada en la figura 8. Si en vez del área de los rectángulos, tomamos el área limitada por la curva y las rectas $x = 3$, $x = 6$, se tendrá el valor aproximado de $P(3 \leq X \leq 6)$.

En este caso la aproximación no es muy buena, como se observa en la figura, pero lo hemos expuesto para comprender el paso de la suma de valores de $b(r; n, p)$ para distintos valores de r , al área limitada por la correspondiente curva normal, la cual se calcula fácilmente mediante una tabla de la función $\Phi(x)$. En el ejemplo anterior, el cálculo exacto del segundo miembro de [6.27], da $P(3 \leq X \leq 6) = 0,778$, y el cálculo mediante la función Φ , apli-

cando [6.25], donde $\alpha = np = 4$, $\sigma = \sqrt{npq} = 1,54$, resulta $\Phi(1,95) + \Phi(0,64) - 1 = 0,713$.

Para valores grandes de n y r la aproximación de la suma de rectángulos por el área limitada por la curva normal correspondiente es suficiente en todas las aplicaciones. Por ejemplo, la figura 9 representa los valores de $b(r;100,0,2)$ y los de la curva correspondiente

$$y = \frac{1}{4\sqrt{2\pi}} e^{-(x-20)^2/32},$$

donde se ve que la coincidencia es casi total.

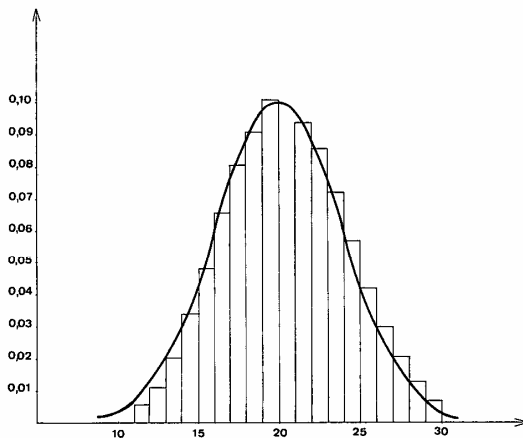


Fig. 9.

Es decir, del teorema de De Moivre-Laplace, que hemos enunciado como una aproximación puntual de la función de probabilidad binomial por la función de densidad normal, se deduce una aproximación entre la suma de las áreas de los rectángulos originados por la función de probabilidad binomial y el área de la función normal correspondiente. Puesto que el área de la función de densidad normal, comprendida entre dos ordenadas, es igual a la diferencia entre los valores de la función de distribución $\Phi(x)$, es posible enunciar:

Teorema de De Moivre-Laplace (Segunda parte). *Si X es la variable aleatoria que representa el número de éxitos en n pruebas de un proceso de Bernoulli de probabilidad p , se verifica*

$$P(a < X \leq b) \sim \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right). \quad [6.29]$$

Como antes, el símbolo \sim representa "aproximadamente igual". El verdadero significado es que, para $n, a, b \rightarrow \infty$ de manera tal que $(b - np)/\sqrt{npq}$ y $(a - np)/\sqrt{npq}$ tiendan a valores finitos, entonces el límite del primer miembro de [6.29] es igual al límite del segundo.

La obtención de acotaciones precisas del error que se comete al tomar la aproximación [6.29] es un problema sumamente difícil (véase, por ejemplo, Feller, ⁽¹⁾ pág. 140). Se suele admitir, de manera un poco ambigua, pues depende de la precisión que se desee en cada caso, que para valores no muy pequeños de p (por ejemplo, $p < 0, 1$) y relativamente grandes de n (por ejemplo, $n > 20$), la aproximación es muy buena, sobre todo si el intervalo $[a, b]$ no se aparta mucho del valor medio np . Según Cramer, ⁽⁴⁾ la aproximación es satisfactoria para $npq > 10$.

Hemos visto dos aproximaciones posibles de la distribución binomial: la de Poisson y la normal. Para valores grandes de np ambas tienden a coincidir. Por ejemplo, para $b(15; 100, 0, 2) = 0, 048$, la aproximación normal da 0,046 y la aproximación de Poisson da 0,051.

Nuevamente la ley de los grandes números. Consideremos [6.29] para $b = np + \epsilon$, $a = np - \epsilon$. Resulta

$$P(np - \epsilon < X \leq np + \epsilon) \sim 2\Phi\left(\frac{\epsilon}{\sqrt{npq}}\right) - 1 \quad [6.30]$$

y también, poniendo ϵn en vez de ϵ

$$P\left(\left|\frac{X}{n} - p\right| < \epsilon\right) \sim 2\Phi\left(\epsilon\sqrt{\frac{n}{pq}}\right) - 1. \quad [6.31]$$

Esta igualdad aproximada significa que, para $n \rightarrow \infty$, ambos miembros tienen el mismo límite. Pero, para $n \rightarrow \infty$, el segundo miembro tiende a 1 (pues $\Phi(\infty) = 1$), con lo cual queda demostrado nuevamente el teorema de Bernoulli (capítulo 4.3), a saber: *en una sucesión de pruebas de Bernoulli con probabilidad p , dado un número cualquiera $\epsilon > 0$, la probabilidad de que el número de éxitos dividido por n difiera de p en menos de ϵ , tiende a la unidad para $n \rightarrow \infty$.*

4. TEOREMA CENTRAL DEL LIMITE

El teorema de De Moivre-Laplace es un caso particular del siguiente teorema, cuya demostración no se va a dar (puede verse

en los libros de Feller, ⁽¹⁾ Reny ⁽²⁾ o Rños ⁽³⁾ citados en la Bibliografía), pero cuyo enunciado conviene tener presente en muchas cuestiones de probabilidad.

Teorema Central del Límite. *Si las variables aleatorias X_1, X_2, \dots son independientes y tienen todas la misma distribución, con esperanza matemática α y varianza σ^2 finitas y $\sigma^2 \neq 0$, entonces la nueva variable aleatoria*

$$Z_n = \frac{X_1 + X_2 + \dots + X_n - n\alpha}{\sigma\sqrt{n}}, \quad [6.32]$$

es asintóticamente normal, es decir, su función de distribución $F_n(z)$ cumple, para todo z , la relación

$$\lim_{n \rightarrow \infty} F_n(z) = \Phi(z) \quad [6.33]$$

Aplicación 1. Consideremos el caso particular en que las X_n son variables de Bernoulli, o sea, variables que toman únicamente los valores 0 y 1, de manera que $P(X_n = 1) = p$, $P(X_n = 0) = q = 1 - p$. En este caso es $\alpha = 1 \cdot p + 0 \cdot q = p$, $\sigma^2 = (1 - p)^2 \cdot p + p^2 \cdot q = pq$. La suma $X_1 + X_2 + \dots + X_n$ es la variable aleatoria que indica el número de éxitos en un proceso binomial de n pruebas. Poniendo $X = X_1 + X_2 + \dots + X_n$, el teorema central del límite nos dice que la nueva variable

$$Z_n = \frac{X - np}{\sqrt{npq}}$$

es asintóticamente normal. Esto es, precisamente, lo que afirma el teorema de De Moivre-Laplace.

Aplicación 2. Si las variables X_n son variables de Poisson de parámetro γ , y por tanto $E(X_n) = \gamma$, $\sigma^2(X_n) = \gamma$ (para todo n), la suma $X = X_1 + X_2 + \dots + X_n$ es otra variable de Poisson de parámetro $n\gamma = \lambda$ (véase la aplicación del capítulo 5.1). Entonces [6.32] y [6.33] permiten escribir

$$\lim_{n \rightarrow \infty} P\left(\frac{X - \lambda}{\sqrt{\lambda}} < z\right) = \Phi(z) \quad [6.34]$$

o sea

$$\lim_{n \rightarrow \infty} \sum_{x < \lambda + z\sqrt{\lambda}} \frac{\lambda^x}{x!} e^{-\lambda} = \Phi(z). \quad [6.35]$$

Esta relación nos dice que, para λ grande, la distribución de Poisson puede aproximarse mediante la distribución normal.

Por este motivo, las tablas de la función de Poisson no suelen pasar de $\lambda = 10$. Para valores $\lambda > 10$, cada sumando de [6.35] es despreciable, y la suma debe calcularse por tablas de la función $\Phi(z)$.

Ejemplo. La variable X tiene distribución de Poisson de parámetro $\lambda = 900$. Calcúlese $P(X < 950)$. Aplicando [6.35] tenemos $\lambda + z\sqrt{\lambda} = 900 + 30z = 950$, de donde $z = 1,666$, y la probabilidad pedida es $\Phi(1,666) = 0,95$.

Problema 6.1. Una ruleta tiene 35 números. Calcular la probabilidad de que en 1.000 jugadas el número 12 salga un número de veces comprendido entre 25 y 30.

Solución. Se aplica la fórmula [6.20]. Aunque $p = 1/35$ es pequeño, siendo $n = 1.000$ grande, se puede aplicar la aproximación [6.29], y se obtiene $P(25 \leq X \leq 30) = \Phi(0,585) - \Phi(-0,390) = \Phi(0,585) + \Phi(0,390) - 1 = 0,37$.

Problema 6.2. Se lanza un dado 100 veces. Se desea saber la probabilidad de que el número 6 salga más de 20 veces.

Solución. Aplicando [6.29] para $n = 100$, $p = 1/6$, resulta $P(X > 20) = 1 - \Phi(0,89) = 0,19$.

Problema 6.3. Tanto la desigualdad de Tchebycheff, en la forma [4.11] del teorema de Bernoulli, como la fórmula [6.31], permiten calcular el número n de pruebas necesarias para que la frecuencia experimental difiera de p en menos de un ϵ dado, con cierta probabilidad también dada. Sin embargo, como ya se observó al final del problema 4.4, la fórmula [4.11] da casi siempre acotaciones excesivas. Se recomienda aplicar siempre la fórmula [6.31].

Consideremos por ejemplo el mismo problema 4.4. Según [6.31], queremos que se cumpla

$$2\Phi\left(\epsilon\sqrt{\frac{n}{pq}}\right) - 1 > 0,95, \quad \text{o sea} \quad \Phi\left(\epsilon\sqrt{\frac{n}{pq}}\right) > 0,975.$$

Las tablas de la función Φ dan

$$\epsilon\sqrt{\frac{n}{pq}} > 1,96, \quad n > (1,96)^2 \frac{pq}{\epsilon^2}. \quad [6.36]$$

Según el enunciado, es $\epsilon = 5 \cdot 10^{-3}$. El producto pq no es conocido, pero según [4.13] es siempre $pq \leq 1/4$. Por tanto, [6.36]

se cumplirá con seguridad si $n > (1,96)^2 10^4 = 38.420$. Vemos que este número es muy inferior al 200.000 dado por la fórmula [4.11].

Problema 6.4. *Un centro turístico debe recibir 1.000 personas, las cuales deberán distribuirse entre dos hoteles iguales. Suponiendo que cada persona elige uno u otro de los hoteles con la misma probabilidad $1/2$, se desea saber la capacidad mínima de cada hotel (número de personas que se pueden albergar) para que, con probabilidad superior a 0,95, todas las personas encuentren alojamiento.*

Solución. Es $np = 500$, $npq = 250$. Hay que calcular x de modo que $P(500 - x \leq X \leq 500 + x) > 0,95$. Puesto que n es grande, podemos aplicar [6.29], resultando

$$P(500 - x \leq X \leq 500 + x) = \Phi\left(\frac{x}{\sqrt{250}}\right) - \Phi\left(\frac{-x}{\sqrt{250}}\right) = 2\Phi\left(\frac{x}{\sqrt{250}}\right) - 1.$$

Para que esta probabilidad sea $> 0,95$, usando las tablas de la función Φ , resulta que debe ser

$$\Phi\left(\frac{x}{15,81}\right) > 0,975, \quad \frac{x}{15,81} > 1,906, \quad x > 30.$$

84

Es decir, bastará que cada hotel tenga capacidad para 530 personas.

Problema 6.5. *En una votación de una población numerosa, un candidato obtiene el 60% de los votos. Se desea saber la probabilidad de que en una muestra de 100 votantes, tomados al azar, menos del 50% voten por el candidato.*

Solución. La ecuación [6.29] para $b = np - \epsilon$, $a = -\infty$, se puede escribir

$$P\left(\frac{X}{n} < p - \epsilon/n\right) = \Phi\left(\frac{-\epsilon}{\sqrt{npq}}\right) = 1 - \Phi\left(\frac{\epsilon}{\sqrt{npq}}\right).$$

En el caso actual es $p = 0,6$, $n = 100$, $\epsilon/n = 0,1$, $\epsilon = 10$, $\sqrt{npq} = \sqrt{24} = 4,9$, y las tablas de la función Φ dan $\Phi(2,04) = 0,979$. Por tanto, la probabilidad buscada es 0,021.

En la práctica, este problema se aplica a la inversa: sabiendo que una muestra tomada al azar en una población (supuesta binomial) ha dado como resultado el r por ciento de votos a favor de cierto candidato, averiguar la probabilidad de que dicho candidato obtenga el s por ciento o más en la elección verdadera.

Problema 6.6. *Un medicamento contra cierta enfermedad es eficaz en el 80% de los casos. Se ensaya una modificación del mismo, y se observa que en 100 enfermos tratados resultó eficaz en 90 de ellos. ¿Cuál es la probabilidad de que este resultado del tratamiento sea debido al azar?*

Solución. Es un caso análogo al anterior. Conviene ahora escribir [6.29] para $a = np + \epsilon$, $b = \infty$, y queda

$$P\left(\frac{X}{n} - p \geq \frac{\epsilon}{n}\right) = 1 - \Phi\left(\frac{\epsilon}{\sqrt{npq}}\right).$$

En el caso del problema, es $p = 0,8$, $n = 100$, $\epsilon/n = 0,1$, $\sqrt{npq} = 4$, $\Phi(2,5) = 0,994$. Por tanto, si $p = 0,8$ la probabilidad de haber obtenido $X/n \geq 0,9$ es 0,006. Como esta probabilidad es muy pequeña, hay que inferir que la modificación ha mejorado el medicamento.

Problema 6.7. *Se ha observado que de 100.000 nacimientos, 51.400 de los nacidos han sido varones y el resto mujeres. ¿Es consistente este hecho con la igualdad de las probabilidades de nacimiento de varón o mujer?*

Solución. Apliquemos [6.31] al caso $p = 0,5$, $X/n = 0,514$ y por tanto $\epsilon = 0,014$, $\sqrt{npq} = 632,4$. Resulta

$$P\left(\left|\frac{X}{n} - 0,5\right| > 0,014\right) = 2(1 - \Phi(8,85)).$$

Como $\Phi(8,85)$, que no se halla en las tablas, es del orden $1 - 10^{-9}$, resulta que la probabilidad de que el nacimiento de 51.400 o más varones sea debido al azar, si la probabilidad fuera 0,5, es insignificante. Por tanto, debe rechazarse la hipótesis de que sea $p = 0,5$.

Este tipo de problemas, se estudian con más detalle en el capítulo 8, al tratar de la "verificación de hipótesis".

5. OTRAS FUNCIONES DE DENSIDAD

Además de la función de densidad normal, que es la más importante, son también de interés las siguientes:

a) **Función de densidad uniforme.** Depende de dos parámetros a y b ($a < b$) y vale

$$f(x) = \frac{1}{b-a} \text{ para } x \in [a, b], \quad f(x) = 0 \text{ para } x \notin [a, b].$$

De aquí

$$E(X) = \frac{a+b}{2}, \quad \sigma^2(X) = \frac{(b-a)^2}{12}.$$

b) **Función de densidad exponencial.** Depende de un parámetro $\lambda > 0$ y vale

$$f(x) = \lambda e^{-\lambda x} \text{ para } x > 0, \quad f(x) = 0 \text{ para } x \leq 0.$$

Para esta función es

$$E(X) = \frac{1}{\lambda}, \quad \sigma^2(X) = \frac{1}{\lambda^2}.$$

c) **Función de densidad de Cauchy.** Depende de un parámetro a y vale

$$f(x) = \frac{1}{\pi} \frac{1}{1+(x-a)^2} \text{ para } -\infty < x < \infty.$$

La esperanza matemática y la varianza de esta función valen ∞ .

7

LA FORMULA DE BAYES

1. INFERENCIA ESTADISTICA

El cálculo de probabilidades es un capítulo de la matemática. Como tal, se construye a partir de unas definiciones y unos axiomas, únicas fuentes de que depende la validez de los resultados. Aunque los modelos y ejemplos del cálculo de probabilidades provienen de la experiencia (lanzamiento de monedas o dados, extracción de bolillas de urnas, etc.), de ella se toma sólo el esquema, para llegar a las ideas, pero la validez de los teoremas a que se llega no depende nunca de resultados experimentales. Más bien al revés, si el resultado de un teorema no coincide con el de la experiencia, la discrepancia se atribuye a que ésta ha sido mal realizada, o a que los elementos de la misma (moneda, dado, urna, ...) no estaban en buenas condiciones.

La estadística, en cambio, es una ciencia experimental. Trata de ordenar, estudiar y predecir el comportamiento de ciertas características de los elementos de un conjunto, llamado *población, colectivo o universo*, que existe de hecho en la naturaleza. Decimos que es una ciencia experimental porque, si bien utiliza a modo de herramienta el cálculo de probabilidades, su objeto es el estudio de conjuntos cuyo comportamiento no se puede cambiar, y el estadístico debe utilizar la parte de la matemática que más le convenga para llegar a resultados acordes con la realidad.

Dentro de la estadística, el capítulo más vinculado con la teoría de las probabilidades es el llamado de la inferencia estadística. Su objeto es en cierta manera el inverso del de los capítulos precedentes. Hasta ahora hemos estudiado funciones de probabilidad o de distribución, y de ellas hemos sacado consecuencias sobre las poblaciones que se adaptaban a ellas. La función de probabilidad era un dato. El problema de la inferencia estadística es el inverso: a partir ciertos datos de una población, averiguar la función de probabilidad que los rige. Los datos que se conocen de la población proceden de ciertas muestras de la misma, o sea, de ciertos subconjuntos elegidos en ella, y de las propiedades de las muestras hay que "inferir" las de toda la población.

A veces, a partir de las muestras, hay que estimar los valores de ciertos parámetros de la población (media, varianza), y otras veces hay que verificar hipótesis sobre los valores de estos parámetros. Ya veremos ejemplos en los capítulos que siguen.

En éste, como puente entre las probabilidades y la inferencia estadística, vamos a dar la llamada fórmula de Bayes.

2. FORMULA DE BAYES

Sean $A_1, \dots, A_2, \dots, A_n$ n sucesos que se excluyen mutuamente y cuya unión es todo el espacio muestral E , o sea, $A_1 \cup A_2 \cup \dots \cup A_n = E$. Sea B un suceso tal que $P(B) \neq 0$, y supongamos que se conocen tanto las probabilidades condicionales $P(B|A_i)$ como las probabilidades $P(A_i)$.

El problema de Bayes consiste en calcular, con los datos anteriores, las probabilidades $P(A_i|B)$. Según [2. 7] tenemos

$$P(B \cap A_i) = P(A_i) P(B|A_i) = P(B) P(A_i|B) \quad [7. 1]$$

y según el axioma II de la definición 2. 5 de probabilidad

$$\begin{aligned} P(B) &= P(B \cap E) = P(B \cap (\cup A_i)) = P(\cup (B \cap A_i)) = & [7. 2] \\ &= \sum_1^n P(B \cap A_i). \end{aligned}$$

De [7. 1] y [7. 2] se deduce

$$P(A_i|B) = \frac{P(A_i \cap B)}{\sum_1^n P(A_i \cap B)} = \frac{P(A_i) P(B|A_i)}{\sum_1^n P(A_i) P(B|A_i)}. \quad [7. 3]$$

Esta es la fórmula conocida con el nombre de *fórmula de Bayes* (publicada por J. Bayes en *Philosophical Transactions*, 1764), la cual expresa la llamada "probabilidad de las causas", pues resuelve el siguiente problema: Suponiendo que un suceso B puede producirse como consecuencia de cualquiera de los sucesos A_i y sabiendo que B se ha producido, averiguar la probabilidad de que haya sido debido a la causa A_i . Se suponen conocidas las probabilidades $P(A_i)$ y $P(B|A_i)$.

Problema 7.1. La urna A_1 contiene 10 bolillas blancas y 15 negras. La urna A_2 contiene 25 bolillas blancas y 5 negras. Se

ha extraído una bolilla de una de las urnas y ha resultado ser blanca. Se supone que, a priori, las dos urnas eran igualmente probables. ¿Cuál es la probabilidad de que la bolilla haya sido extraída de la urna A_1 ?

Solución. Es $P(A_1) = P(A_2) = 1/2$, $P(B|A_1) = 10/25 = 2/5$, $P(B|A_2) = 25/30 = 5/6$. Por tanto la fórmula de Bayes da $P(A_1|B) = 12/37$.

Problema 7.2. Una fábrica tiene tres máquinas A_1, A_2 , y A_3 para producir ciertas piezas, tales que: A_1 produce el 30% de las piezas con un porcentaje del 2% de piezas defectuosas; A_2 produce el 25% de las piezas con el 1% de defectuosas y A_3 produce el 45% de las piezas con el 3% de defectuosas. Se elige al azar una pieza que sale de la fábrica para la venta y resulta ser defectuosa, ¿cuál es la probabilidad de que provenga de la máquina A_1 , de la A_2 o de la A_3 ?

Solución. Es $P(A_1) = 0,3$, $P(A_2) = 0,25$, $P(A_3) = 0,45$, $P(B|A_1) = 0,02$, $P(B|A_2) = 0,01$, $P(B|A_3) = 0,03$. Con estos datos, las probabilidades $P(A_i|B)$ de que la pieza defectuosa (suceso B) provenga de la máquina A_i , resultan

$$P(A_1|B) = 0,273, \quad P(A_2|B) = 0,114, \quad P(A_3|B) = 0,614.$$

Problema 7.3. La fórmula de Bayes permite calcular la probabilidad "a posteriori", después de realizado un cierto experimento, de un suceso cuya probabilidad "a priori" se conoce. Supongamos, por ejemplo, una moneda desconocida a la que atribuimos una probabilidad p_1 de tener dos caras. Se lanza la moneda n veces consecutivas y en todas ellas sale cara. ¿Cuál es la probabilidad, después de estos lances, de que la moneda tenga dos caras?

Solución. Sea A_1 el suceso "tener dos caras", A_2 el suceso opuesto (tener una sola cara) y B el suceso "salir cara n veces consecutivas". Es $P(A_1) = p_1$, $P(B|A_1) = 1$, $P(B|A_2) = (1/2)^n$, $P(A_2) = 1 - p_1$. Por tanto, la fórmula de Bayes da para la probabilidad buscada

$$P(A_1|B) = \frac{p_1}{p_1 + (1 - p_1)(1/2)^n}.$$

Por ejemplo, si $p_1 = 10^{-7}$, resulta $P(A_1|B) \sim 10^{-4}$.

Observación. Este último problema se basa en la probabilidad a priori p_1 , que en cierto modo es un "grado de creencia" acerca de una determinada cuestión. El ejemplo pone de manifiesto la relación de la fórmula de Bayes con la inferencia estadística: a

partir de un valor estimado p_1 , mediante una muestra (en el ejemplo anterior el lanzamiento n veces de la moneda) se consigue un mejor valor de la probabilidad estimada.

Debido a la ambigüedad de la probabilidad *a priori*, el uso de la fórmula de Bayes en este tipo de problemas ha sido muy discutida. En general, se procura no usarla, si bien, modernamente, hay escuelas que procuran tomarla como base de toda la teoría de la inferencia estadística (véase el libro de Lindley, citado en la Bibliografía ⁽⁶⁾).

8

ESTIMACION POR PUNTO

1. MUESTRAS

Sea X una variable aleatoria, discreta o continua, cuya función de probabilidad o de densidad sea f , y cuyo dominio sea una cierta población (o espacio muestral) dada.

Una *muestra* de tamaño n , significa un conjunto x_1, x_2, \dots, x_n de valores de X . Para que una muestra sea útil hay que suponer que ha sido tomada "al azar" con la función de probabilidad o de densidad f . Esto significa que se cumplen las dos condiciones siguientes:

a) Cada uno de los valores x_i puede considerarse como un valor de una nueva variable aleatoria X_i que tiene la misma función de probabilidad o de densidad f que X . Por tanto

$$E(X_i) = E(X), \quad \sigma^2(X_i) = \sigma^2(X), \quad i = 1, 2, \dots, n. \quad [8.1]$$

b) Las variables aleatorias X_i son independientes.

Vamos a suponer siempre que estas dos condiciones se cumplen, lo que facilita la realización de inducciones a partir de la muestra. Supongamos, por ejemplo, que la población sea el conjunto de los habitantes de un país cuya edad está comprendida entre 20 y 21 años, y que X representa la talla de los mismos: Al hablar de una "muestra" de esta población se sobreentiende que se han tomado n habitantes independientes entre sí (no entre los de una misma ascendencia o entre los que practican un mismo deporte) y que, además, es lícito suponer que todos ellos tienen la misma probabilidad de tener una talla entre ciertos límites (o sea, todos están sujetos a la misma función de densidad de probabilidad).

El problema de elegir bien las muestras, de manera que se cumplan las mencionadas condiciones a) y b), no es un problema fácil. Hay diversas técnicas que se adaptan a cada caso particular. Para poblaciones finitas es útil el uso de las "tablas de números aleatorios", es decir, tablas de números que pueden considerarse como elegidos al azar. Entonces, para tomar una muestra de

tamaño n , se numeran todos los elementos del conjunto y se toman n números sucesivos de dichas tablas. Para poblaciones infinitas se pueden también utilizar las tablas de números aleatorios, sustituyendo la población infinita por otra población finita de análoga composición.

2. MEDIA Y VARIANZA DE UNA MUESTRA

Supongamos una variable aleatoria X , y sean x_1, x_2, \dots, x_n sus valores en una determinada muestra de tamaño n . La *media de la muestra* es

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) \quad [8.2]$$

y la *varianza de la muestra* es

$$s^2 = \frac{1}{n} \sum_1^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_1^n x_i^2 - \bar{x}^2 \quad [8.3]$$

Obsérvese la diferencia entre estos valores y la esperanza o media $E(X)$ y la varianza $\sigma(X)$. En estos últimos intervienen todos los posibles valores de X , en tanto que en [8.2] y de [8.3] intervienen únicamente n valores.

De manera análoga, teniendo en cuenta las definiciones del capítulo 3, se definen el *momento* de orden r de la muestra

$$a_r = \frac{1}{n} \sum_1^n x_i^r \quad [8.4]$$

y el *momento centrado de orden r*

$$m_r = \frac{1}{n} \sum_1^n (x_i - \bar{x})^r. \quad [8.5]$$

A veces interesa la llamada *asimetría* g_1 y el *exceso* g_2 de la muestra, que se definen por

$$g_1 = \frac{m_3}{s^3}, \quad g_2 = \frac{m_4}{s^4} - 3.$$

Considerando cada valor x_i de la muestra como valor de una variable aleatoria X_i , entonces \bar{x} será el valor de la variable aleatoria

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n). \quad [8.6]$$

Aplicando los teoremas del capítulo 3 referentes a la esperanza y a la varianza de una suma de variables aleatorias, y teniendo en cuenta [8.1] y [8.6] resulta

$$E(\bar{X}) = E(X) \quad [8.7]$$

y también, puesto que las X_i son independientes, $\sigma^2(\bar{X}) = (1/n^2)\sigma^2(X_1 + X_2 + \dots + X_n) = (1/n)\sigma^2(X)$, o sea,

$$\sigma^2(\bar{X}) = \frac{\sigma^2(X)}{n}. \quad [8.8]$$

La *varianza de la muestra* S^2 puede considerarse también como el valor de una nueva variable aleatoria S^2 , definida por

$$S^2 = \frac{1}{n} \sum_1^n (X_i - \bar{X})^2. \quad [8.9]$$

Para calcular la esperanza de S^2 , observemos que se puede escribir

$$\begin{aligned} S^2 &= \frac{1}{n} \sum_1^n [X_i - E(X) - (\bar{X} - E(X))]^2 = \\ &= \frac{1}{n} \sum_1^n [X_i - E(X)]^2 - [\bar{X} - E(X)]^2. \end{aligned}$$

Teniendo en cuenta que

$$E(X_i - E(X))^2 = E(X_i - E(X_i))^2 = \sigma^2(X_i) = \sigma^2(X)$$

y

$$E(\bar{X} - E(X))^2 = E(\bar{X} - E(\bar{X}))^2 = \sigma^2(\bar{X}) = \frac{\sigma^2(X)}{n}$$

resulta

$$E(S^2) = \frac{n-1}{n} \sigma^2(X). \quad [8.10]$$

Un poco más complicado, pero tampoco difícil, es calcular $\sigma^2(S^2)$. El resultado es

$$\sigma^2(S^2) = \frac{\mu_4(X)}{n} \left(\frac{n-1}{n}\right)^2 - \frac{(n-3)(n-1)}{n^3} \sigma^4(X). \quad [8.11]$$

Para el caso de una variable aleatoria normal, es $\mu_4 = 3\sigma^4$ y, por tanto, $\sigma^2(S^2) = (2(n-1)/n^2)\sigma^4$.

Observación. En el caso de una población finita, es interesante distinguir entre las *muestras con reemplazamiento*, en las que

Aplicando los teoremas del capítulo 3 referentes a la esperanza y a la varianza de una suma de variables aleatorias, y teniendo en cuenta [8.1] y [8.6] resulta

$$E(\bar{X}) = E(X) \quad [8.7]$$

y también, puesto que las X_i son independientes, $\sigma^2(\bar{X}) = (1/n^2)\sigma^2(X_1 + X_2 + \dots + X_n) = (1/n)\sigma^2(X)$, o sea,

$$\sigma^2(\bar{X}) = \frac{\sigma^2(X)}{n}. \quad [8.8]$$

La *varianza de la muestra* S^2 puede considerarse también como el valor de una nueva variable aleatoria S^2 , definida por

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2. \quad [8.9]$$

Para calcular la esperanza de S^2 , observemos que se puede escribir

$$\begin{aligned} S^2 &= \frac{1}{n} \sum_{i=1}^n [X_i - E(X) - (\bar{X} - E(X))]^2 = \\ &= \frac{1}{n} \sum_{i=1}^n [X_i - E(X)]^2 - [\bar{X} - E(X)]^2. \end{aligned}$$

Teniendo en cuenta que

$$E(X_i - E(X))^2 = E(X_i - E(X_i))^2 = \sigma^2(X_i) = \sigma^2(X)$$

y

$$E(\bar{X} - E(X))^2 = E(\bar{X} - E(\bar{X}))^2 = \sigma^2(\bar{X}) = \frac{\sigma^2(X)}{n}$$

resulta

$$E(S^2) = \frac{n-1}{n} \sigma^2(X). \quad [8.10]$$

Un poco más complicado, pero tampoco difícil, es calcular $\sigma^2(S^2)$. El resultado es

$$\sigma^2(S^2) = \frac{\mu_4(X)}{n} \left(\frac{n-1}{n} \right)^2 - \frac{(n-3)(n-1)}{n^3} \sigma^4(X). \quad [8.11]$$

Para el caso de una variable aleatoria normal, es $\mu_4 = 3\sigma^4$ y, por tanto, $\sigma^2(S^2) = (2(n-1)/n^2)\sigma^4$.

Observación. En el caso de una población finita, es interesante distinguir entre las *muestras con reemplazamiento*, en las que

para todo n (tamaño de la muestra) y todo valor posible de θ . Si [8.13] no se cumple, el estimador se llama *sesgado*, o que tiene *sesgo*.

Ejemplo 1. Supongamos que se quiere estimar el valor medio o esperanza de una variable aleatoria X , cualquiera que sea su densidad o función de probabilidad. El estimador

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n) \quad [8.14]$$

es un estimador insesgado, por cumplirse [8.7].

Recordando la desigualdad de Tchebycheff [3.15], para cualquier número $k > 0$ se tiene

$$P(|\bar{X} - E(X)| \geq k\sigma(\bar{X})) \leq \frac{1}{k^2} \cdot \quad [8.15]$$

Poniendo $k\sigma(\bar{X}) = \epsilon$ y aplicando [8.8] resulta

$$P(|\bar{X} - E(X)| \geq \epsilon) \leq \frac{\sigma^2(X)}{n\epsilon^2} \cdot \quad [8.16]$$

Esto nos dice que al tomar \bar{X} por $E(X)$, el error que se comete converge a cero en probabilidad (definición 4.1). Por tanto, el estimador \bar{X} de $E(X)$ es tanto mejor cuanto mayor es el tamaño de la muestra, lo cual es intuitivamente evidente (suponiendo, naturalmente, que $\sigma^2(X)$ tiene un valor finito).

Ejemplo 2. Para estimar la varianza $\sigma^2(X)$, para cualquier distribución de probabilidad, podemos tomar el estimador S^2 [8.9]. Entonces, la relación [8.10] nos dice que S^2 es un estimador sesgado. Por este motivo, suele tomarse como estimador de la varianza $\sigma^2(X)$ la función $nS^2/(n-1)$, pues entonces

$$E\left(\frac{n}{n-1} S^2\right) = \sigma^2(X). \quad [8.17]$$

Igual que en el ejemplo anterior, la desigualdad de Tchebycheff permite acotar la probabilidad del error que se comete al tomar $nS^2/(n-1)$ como valor de $\sigma^2(X)$. Teniendo en cuenta [8.11] y suponiendo que $\mu_4(X)$ y $\sigma^2(X)$ son finitos, resulta también que, en probabilidad $|nS^2/(n-1) - \sigma^2(X)| \rightarrow 0$, para $n \rightarrow \infty$.

Ejemplo 3. Hay estimadores que no son una función simple de los valores de la muestra y por tanto se prestan menos al cálculo, y deben ser estudiados en cada caso por métodos particulares. Consideremos, por ejemplo, la variable aleatoria X cuya función de densidad es

$$\begin{aligned}
 f(x; \theta) &= \frac{1}{\theta} \quad \text{para } |x| \leq \theta \\
 f(x; \theta) &= 0 \quad \text{para } |x| > \theta.
 \end{aligned}
 \tag{8.18}$$

Es el caso de una distribución uniforme en el intervalo $[-\theta, +\theta]$. Dada una muestra x_1, x_2, \dots, x_n tomamos como estimador de la amplitud θ la variable aleatoria X_n , cuyos valores son $x_n = \text{máximo de } |x_i|$. La función de densidad de probabilidad de la muestra es $(2\theta)^{-n}$ si todos los $|x_i|$ son $\leq \theta$, y cero si algún $|x_i|$ es $> \theta$. Con esto se calcula fácilmente que

$$E(X_n) = \frac{n}{n+1} \theta.$$

Se trata, por tanto, de un estimador sesgado, aunque el sesgo tiende a desaparecer en muestras grandes ($n \rightarrow \infty$).

4. ESTIMADORES CONSISTENTES

Definición 8.1. Un estimador $T(X_1, X_2, \dots, X_n)$ del parámetro θ se dice que es *consistente*, si T converge en probabilidad a θ .

96

De manera más precisa, se dice que T es consistente, si dados dos números positivos cualesquiera ϵ, η , arbitrariamente pequeños, existe un entero N tal que, para todo $n > N$, es $P(|T - \theta| < \epsilon) > 1 - \eta$.

Según esta definición y lo demostrado en los ejemplos del número anterior, los estimadores \bar{X} de $E(X)$ y $nS^2/(n-1)$ de $\sigma^2(X)$ son estimadores consistentes.

Estos dos son casos particulares del siguiente teorema

Teorema 8.1. Si para $n \rightarrow \infty$, se verifica

$$E(T) \rightarrow \theta, \quad \sigma^2(T) \rightarrow 0, \tag{8.19}$$

entonces T es un estimador consistente de θ .

Demostración. La desigualdad de Tchebycheff [3.15] se puede escribir

$$P(|T - E(T)| \leq k\sigma(T)) \geq 1 - \frac{1}{k^2} \tag{8.20}$$

Escribiendo $T - \theta = (T - E(T)) - (\theta - E(T))$, resulta que, siempre que se cumpla $|T - E(T)| \leq k\sigma(T)$, se cumplirá también

$$|T - \theta| \leq |T - E(T)| + |\theta - E(T)| \leq \kappa\sigma(T) + |\theta - E(T)|,$$

y por tanto, de [8.19] se deduce

$$P(|T - \theta| \leq \kappa\sigma(T) + |\theta - E(T)|) \geq 1 - \frac{1}{\kappa^2}. \quad [8.21]$$

Poniendo $1/\kappa^2 = \eta$ y puesto que, por hipótesis, $\kappa\sigma(T) + |\theta - E(T)| \rightarrow 0$, cualquiera que sea ϵ habrá un entero positivo N suficientemente grande tal que, para todo $n > N$, $\kappa\sigma(T) + |\theta - E(T)| < \epsilon$ y, en consecuencia,

$$P(|T - \theta| < \epsilon) \geq 1 - \eta$$

cualesquiera que sean ϵ, η , lo cual prueba el teorema.

5. ESTIMADORES SUFICIENTES

Supongamos una variable aleatoria cuya función de probabilidad sea la binomial. El parámetro θ es ahora la probabilidad p . Consideremos las variables de Bernoulli X_1, X_2, \dots, X_n , cada una de las cuales toma el valor 1 (éxito) con probabilidad p , y el valor 0 (fracaso) con probabilidad $q = 1 - p$. Puesto que $E(X_i) = p$, la variable aleatoria

$$T = \frac{1}{n}(X_1 + X_2 + \dots + X_n) \quad [8.22]$$

que da el número medio de éxitos en cada muestra, es un estimador insesgado de p . Supongamos que, intentando ser más precisos, considerásemos en la muestra, no sólo el número de éxitos (que es la suma $X_1 + X_2 + \dots + X_n$), sino también el *orden* de los mismos. Entonces, la probabilidad de una determinada sucesión ordenada con r éxitos es $p^r q^{n-r}$. Si no se tiene en cuenta el orden, la probabilidad de r éxitos es $\binom{n}{r} p^r q^{n-r}$. Para la estimación de p , lo mismo da conocer esta última probabilidad que la primera $p^r q^{n-r}$. Es decir, el hecho de conocer el orden en que se han sucedido los éxitos y fracasos no añade ninguna información suplementaria a la que proporciona el saber el número de éxitos. En otras palabras, de una muestra x_1, x_2, \dots, x_n lo único que interesa, para estimar p , es el número de éxitos. Por este motivo se dice que T es un estimador *suficiente*. La definición precisa es como sigue

Definición 8.2. Sean T, W dos estimadores del parámetro θ de una cierta función de probabilidad (o función de densidad). Si, dado T , la función de probabilidad condicional (o función de densidad condicional) de W no depende de θ , cualquiera que sea W , el estimador T se dice *suficiente*.

$$|T - \theta| \leq |T - E(T)| + |\theta - E(T)| \leq \kappa\sigma(T) + |\theta - E(T)|,$$

y por tanto, de [8.19] se deduce

$$P(|T - \theta| \leq \kappa\sigma(T) + |\theta - E(T)|) \geq 1 - \frac{1}{\kappa^2}. \quad [8.21]$$

Poniendo $1/\kappa^2 = \eta$ y puesto que, por hipótesis, $\kappa\sigma(T) + |\theta - E(T)| \rightarrow 0$, cualquiera que sea ϵ habrá un entero positivo N suficientemente grande tal que, para todo $n > N$, $\kappa\sigma(T) + |\theta - E(T)| < \epsilon$ y, en consecuencia,

$$P(|T - \theta| < \epsilon) \geq 1 - \eta$$

cualesquiera que sean ϵ, η , lo cual prueba el teorema.

5. ESTIMADORES SUFICIENTES

Supongamos una variable aleatoria cuya función de probabilidad sea la binomial. El parámetro θ es ahora la probabilidad p . Consideremos las variables de Bernoulli X_1, X_2, \dots, X_n , cada una de las cuales toma el valor 1 (éxito) con probabilidad p , y el valor 0 (fracaso) con probabilidad $q = 1 - p$. Puesto que $E(X_i) = p$, la variable aleatoria

$$T = \frac{1}{n}(X_1 + X_2 + \dots + X_n) \quad [8.22]$$

que da el número medio de éxitos en cada muestra, es un estimador insesgado de p . Supongamos que, intentando ser más precisos, considerásemos en la muestra, no sólo el número de éxitos (que es la suma $X_1 + X_2 + \dots + X_n$), sino también el *orden* de los mismos. Entonces, la probabilidad de una determinada sucesión ordenada con r éxitos es $p^r q^{n-r}$. Si no se tiene en cuenta el orden, la probabilidad de r éxitos es $\binom{n}{r} p^r q^{n-r}$. Para la estimación de p , lo mismo da conocer esta última probabilidad que la primera $p^r q^{n-r}$. Es decir, el hecho de conocer el orden en que se han sucedido los éxitos y fracasos no añade ninguna información suplementaria a la que proporciona el saber el número de éxitos. En otras palabras, de una muestra x_1, x_2, \dots, x_n lo único que interesa, para estimar p , es el número de éxitos. Por este motivo se dice que T es un estimador *suficiente*. La definición precisa es como sigue

Definición 8.2. Sean T, W dos estimadores del parámetro θ de una cierta función de probabilidad (o función de densidad). Si, dado T , la función de probabilidad condicional (o función de densidad condicional) de W no depende de θ , cualquiera que sea W , el estimador T se dice *suficiente*.

Ejemplo 1. Sea el caso ya considerado de una variable binomial de parámetro $\theta = p$. Suponiendo una muestra con r éxitos y teniendo en cuenta la ordenación es $f(x; \theta) = p^r q^{n-r}$. Respecto del estimador $T = (1/n)(X_1 + X_2 + \dots + X_n)$ es $t = r/n$. La descomposición [8.24] se cumple para $r(t; \theta) = \binom{n}{tn} p^{tn} (1-p)^{n-tn}$, $s = 1/\binom{n}{tn}$.

Ejemplo 2. Sea una variable normal de la que se quiere estimar la media, conocida la varianza σ^2 . Tomemos el estimador $T = (1/n)(X_1 + X_2 + \dots + X_n)$. La función de densidad de una muestra x_1, x_2, \dots, x_n es el producto de las funciones de densidad de cada x_i (puesto que son independientes), o sea

$$f(x_1, x_2, \dots, x_n; \theta) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} e^{-\sum_1^n (x_i - \theta)^2 / 2\sigma^2}$$

y puesto que, poniendo $t = \bar{x}$, es

$$\sum_1^n (x_i - \theta)^2 = \sum_1^n ((x_i - t) + (t - \theta))^2 = \sum_1^n (x_i - t)^2 + n(t - \theta)^2$$

resultan los dos factores

$$r(t; \theta) = e^{-n(t-\theta)^2/2\sigma^2}, \quad s(x_1, x_2, \dots, x_n) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} e^{-\sum_1^n (x_i - t)^2/2\sigma^2}.$$

En el último factor puede ponerse $t = (1/n)(x_1 + x_2 + \dots + x_n)$ para destacar que sólo depende de las x_i . Por tanto, $T = \bar{X}$ es un estimador suficiente del valor medio $E(X)$.

Ejemplo 3. Sea una variable de Poisson de parámetro λ . Tomemos $T = (1/n)(X_1 + X_2 + \dots + X_n)$ como estimador de $\theta = \lambda$. Resulta

$$f(x_1, x_2, \dots, x_n; \theta) = \frac{\theta^{x_1 + \dots + x_n}}{x_1! x_2! \dots x_n!} e^{-\theta n} = \theta^{nt} e^{-\theta n} \frac{1}{x_1! \dots x_n!}$$

y la descomposición [8.24] salta a la vista. Por tanto, T es un estimador suficiente de λ .

6. ESTIMADORES EFICIENTES

Un mismo parámetro θ puede estimarse por distintos estimadores. Por ejemplo, para estimar $E(X)$ se puede tomar cualquier estimador de la forma $T = c_1 X_1 + c_2 X_2 + \dots + c_n X_n$ (c_i constantes). Si se quiere que T sea insesgado bastará que se cumpla $c_1 + c_2 + \dots + c_n = 1$. Aparece entonces el problema de elegir, entre todos los posibles estimadores de un parámetro, el mejor de todos,

o sea, aquel cuyos valores posibles sean todos muy cercanos al valor de θ . Una manera de medir la "bondad" de un estimador T , o bien, de medir el "riesgo" de obtener valores demasiado apartados del verdadero valor de θ , consiste en considerar la varianza (suponiendo estimadores insesgados)

$$\sigma^2(T) = E(T - \theta)^2 \quad [8.25]$$

y pedir que ella sea lo menor posible.

Definición 8.3. Un estimador insesgado T de un parámetro θ se dice que es *eficiente* o de *mínima varianza* si, dado otro estimador insesgado cualquiera W de θ , se cumple siempre $\sigma^2(T) \leq \sigma^2(W)$.

Para un mismo θ puede haber más de un estimador eficiente. El problema de hallarlo no es, en general, fácil. Un criterio a veces útil es el que se deriva de la siguiente desigualdad, que enunciaremos sin demostración:

Desigualdad de Cramer-Rao. Si $f(x; \theta)$ es la función de probabilidad (o la función de densidad) de la variable aleatoria X , bajo ciertas condiciones muy amplias de la misma, la varianza de cualquier estimador insesgado T de θ cumple la desigualdad

$$\sigma^2(T) \geq \frac{1}{nE\left(\frac{\partial \log f(X; \theta)}{\partial \theta}\right)^2} \quad [8.26]$$

De esta manera, si se encuentra un estimador T cuya varianza resulte igual al segundo miembro de [8.26], es seguro que tal estimador es eficiente. Sin embargo, hay que tener en cuenta que muchas veces el mínimo expresado por [8.26] no es alcanzable por ningún estimador.

Ejemplo. Sea X una variable aleatoria normal; la función de densidad es la [6.15]. Por tanto

$$\log f(x; \alpha, \sigma^2) = -\frac{(x - \alpha)^2}{2\sigma^2} - \log(\alpha\sqrt{2\pi}) \quad [8.27]$$

Si se quiere estimar la media α , es $\partial \log f / \partial \alpha = (x - \alpha) / \sigma^2$, y $E((X - \alpha) / \sigma^2)^2 = (1 / \sigma^4) E(X - \alpha)^2 = 1 / \sigma^2$. Por tanto, la cota inferior dada por la desigualdad de Cramer-Rao es σ^2 / n . Como este valor es precisamente la varianza del estimador insesgado $T = \bar{X}$, resulta que \bar{X} es un estimador eficiente de la media.

Si se quiere estimar la varianza, de [8.27] se deduce

$$\frac{\partial \log f(x; \alpha, \sigma^2)}{\partial \sigma^2} = \frac{(x - \alpha)^2}{2\sigma^4} - \frac{1}{2\sigma^2}$$

y por tanto

$$E \left[\frac{\partial \log f(X; \alpha, \sigma^2)}{\partial \sigma^2} \right]^2 = \frac{\mu_4}{4\sigma^8} + \frac{1}{4\sigma^4} - \frac{1}{2\sigma^4} = \frac{1}{2\sigma^4},$$

ya que $\mu_4 = 3\sigma^4$. Por tanto [8.26] nos dice que todo estimador de la varianza de una variable normal tiene la varianza $\geq 2\sigma^4/n$.

Distingamos ahora dos casos. El estimador insesgado $(n/(n-1))S^2$ de σ^2 según [8.11], tiene la varianza $(2/(n-1))\sigma^4$, que es mayor que $2\sigma^4/n$ y por tanto la desigualdad de Cramer-Rao no permite decidir si es o no eficiente.

En cambio, si se conoce la media α , se puede tomar el estimador $S_0 = (1/n) \sum_{i=1}^n (X_i - \alpha)^2$, para el cual se tiene

$$E(S_0^2) = \sigma^2, \quad \sigma^2(S_0^2) = \frac{2}{n} \sigma^4. \quad [8.28]$$

Por tanto, en este caso, se trata de un estimador eficiente.

7. ESTIMADORES DE MAXIMA VEROSIMILITUD

Dentro de la estimación por punto, uno de los métodos más naturales y útiles es el llamado de *máxima verosimilitud*, que vamos a exponer. Supongamos una población sujeta a la función de probabilidad (o función de densidad) $f(x; \theta)$. Se extrae una muestra x_1, x_2, \dots, x_n . Puesto que la función de probabilidad (o de densidad) para cada x_i es $f(x_i; \theta)$, la función de probabilidad (o de densidad) de la muestra será el producto

$$P(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta). \quad [8.29]$$

El método de máxima verosimilitud consiste en tomar, como valor estimado de θ , el valor que hace máxima la función P. Es decir, se elige el valor del parámetro para el cual es máxima la probabilidad de haber sacado la muestra obtenida.

Puesto que si P es máxima también lo será su logaritmo; para convertir el producto en suma, se suele tomar la función $L = \log P$ y determinar θ por la ecuación

$$\frac{\partial L}{\partial \theta} = \sum_{i=1}^n \frac{\partial \log f(x_i; \theta)}{\partial \theta} = 0. \quad [8.30]$$

Si esta ecuación tiene alguna solución $\theta = \text{constante}$, independiente de las x_i , esta solución se deja de lado. Toda otra solución, que sea función de las x_i , constituye un estimador de máxima verosimilitud del parámetro θ .

Ejemplo 1. Distribución binomial. Hay que estimar p . Es $f(x; p) = p^x(1-p)^{1-x}$, y para una muestra x_1, x_2, \dots, x_n será

$$P(x_1, x_2, \dots, x_n; p) = p^{x_1+x_2+\dots+x_n}(1-p)^{n-(x_1+x_2+\dots+x_n)}.$$

Aplicando [8.30] resulta inmediatamente el estimador $\hat{p} = (1/n)(X_1 + X_2 + \dots + X_n)$.

Ejemplo 2. Distribución de Poisson. En este caso es $f(x; \lambda) = (\lambda^x/x!)e^{-\lambda}$, y por tanto

$$P(x_1, x_2, \dots, x_n; \lambda) = \frac{\lambda^{x_1+x_2+\dots+x_n}}{x_1!x_2!\dots x_n!} e^{-\lambda},$$

de donde resulta de inmediato que el estimador de máxima verosimilitud para λ es $\hat{\lambda} = (1/n)(X_1 + X_2 + \dots + X_n)$.

102

Ejemplo 3. Distribución normal. Hay dos parámetros posibles de estimar: la media α y la varianza σ^2 . Haciendo el producto de n funciones [6.15] y tomando logaritmos resulta

$$L = -\frac{1}{2\sigma^2} \sum_1^n (x_i - \alpha)^2 - \frac{n}{2} \log \sigma^2 - \log (\sqrt{2\pi})^n.$$

Si se supone conocida σ^2 , para α resulta el estimador usual $\bar{X} = (1/n)(X_1 + X_2 + \dots + X_n)$.

Si se supone conocida α , para σ^2 resulta

$$\frac{1}{2\sigma^2} \sum_1^n (x_i - \alpha)^2 - \frac{n}{2\sigma^2} = 0, \text{ de donde } \hat{\sigma}^2 = \frac{1}{n} \sum_1^n (X_i - \alpha)^2.$$

Si α y σ^2 son desconocidos, el sistema de ecuaciones $\partial L / \partial \alpha = 0$, $\partial L / \partial \sigma^2 = 0$ da

$$\hat{\alpha} = \frac{1}{n} \sum_1^n X_i = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_1^n (X_i - \bar{X})^2.$$

Ejemplo 4. Distribución uniforme. El caso de la función [8.18] da la función $P(x_1, x_2, \dots, x_n; \theta) = 1/(2\theta)^n$ siempre que para todo i sea $|x_i| \leq \theta$ y $P = 0$, si algún x_i es $> \theta$. En este caso no se puede aplicar directamente [8.30] (lo que daría $\theta = 0$, que no depende de

las x_i). Hay que seguir un razonamiento directo: *el máximo de P se obtiene para el mínimo valor de θ compatible con la muestra.* Pero este valor mínimo de θ es precisamente el máximo de los $|x_i|$. O sea, el método de máxima verosimilitud da nuevamente, para θ , el estimador del ejemplo 3 de la sección 3.

Observación 1. Se puede demostrar que, para n grande, la varianza de los estimadores de máxima verosimilitud tiene por límite la cota inferior dada por la desigualdad de Cramer-Rao. Es decir: *los estimadores de máxima verosimilitud son asintóticamente eficientes.*

Observación 2. La función $P(x_1, x_2, \dots, x_n; \theta)$ de [8.29] es la misma función $f(x_1, x_2, \dots, x_n; \theta)$ de [8.24]. Por tanto, si se tiene un estimador suficiente T , la ecuación [8.30] equivale a $\partial r(t, \theta) / \partial \theta = 0$. Esto nos dice que si existe un estimador suficiente cuyos valores sean t , todo estimador de máxima verosimilitud está dado por la ecuación $\partial r(t; \theta) / \partial \theta = 0$, o sea, es una función de T . En otras palabras: *si existe un estimador suficiente, todo estimador de máxima verosimilitud es una función del mismo.*

ESTIMACION POR INTERVALOS DE CONFIANZA. VERIFICACION DE HIPOTESIS

1. LA DISTRIBUCION χ^2 ("ji" cuadrado).

Antes de entrar en la estimación de parámetros por el método de intervalos de confianza son necesarias algunas distribuciones importantes que vamos a enunciar. Nos limitaremos a exponer sus principales propiedades sin entrar en su demostración, la que, sin ser difícil, exige el cálculo de ciertas integrales múltiples que se quieren evitar. Las respectivas demostraciones pueden verse, por ejemplo, en el libro de Ríos, ⁽³⁾ citado en la bibliografía.

Sean X_1, X_2, \dots, X_n variables aleatorias independientes normales $N(0, 1)$, es decir, de media 0 y varianza 1. Consideremos la nueva variable aleatoria

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2. \quad [9.1]$$

Se demuestra que la función de densidad de esta variable χ^2 es

$$h_n(x) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2}, \quad x > 0 \quad [9.2]$$

donde $\Gamma(n/2)$ es la función "gamma" que se define en el Apéndice II. La correspondiente función de *distribución* χ^2 es

$$P(\chi^2 \leq x) = \int_{-\infty}^x h_n(x) dx. \quad [9.3]$$

También nos interesaremos por

$$P(\chi^2 > \chi_0^2) = \int_{\chi_0^2}^{\infty} h_n(x) dx, \quad \chi_0^2 \geq 0. \quad [9.4]$$

Esta es una probabilidad, cuyo valor depende de χ_0^2 y de n . Representándola por p , escribiremos

$$P(\chi^2 > \chi_0^2) = p. \quad [9.5]$$

En general interesa, dados p y n , calcular χ_0^2 para que [9.5] se cumpla. Para ello se han construido las tablas de χ^2 (véase el Apéndice III), que dan el correspondiente χ_0^2 para distintos valores de p y de n . El número n se llama el número de *grados de libertad* de χ^2 .

En resumen, prescindiendo de su procedencia, lo que interesa es saber que las tablas de χ^2 resuelven el siguiente problema: "Dadas n variables aleatorias independientes normales $N(0, 1)$, sean X_1, X_2, \dots, X_n , y la probabilidad p , calcular el valor χ_0^2 tal que sea $P(\chi^2 > \chi_0^2) = p$, siendo χ^2 la suma [9.1]".

A partir de [9.2] y como ejercicio de cálculo integral se pueden calcular los valores

$$E(\chi^2) = n, \quad \sigma^2(\chi^2) = 2n. \quad [9.6]$$

Si las variables normales X_i son de tipo $N(\alpha, \sigma^2)$, basta reducirlas al tipo $N(0, 1)$, es decir, se considera la suma

$$\sum_1^n \left(\frac{X_i - \alpha}{\sigma} \right)^2 \quad [9.7]$$

que tendrá entonces una distribución χ^2 . En particular, escribiendo

$$S_0^2 = \frac{1}{n} \sum_1^n (X_i - \alpha)^2 \quad [9.8]$$

resulta que la *variable aleatoria* nS_0^2/σ^2 *tiene la distribución* χ^2 *con* n *grados de libertad*.

En cambio, si se considera la suma

$$S^2 = \frac{1}{n} \sum_1^n (X_i - \bar{X})^2 \quad [9.9]$$

donde $\bar{X} = (1/n)(X_1 + X_2 + \dots + X_n)$, los sumandos del segundo miembro ya no son independientes (puesto que $\sum_1^n (X_i - \bar{X}) = 0$) y el resultado anterior deja de ser válido. En este caso se demuestra que: *si las variables aleatorias independientes* χ_i *son normales de tipo* $N(\alpha, \sigma^2)$, *la nueva variable aleatoria* nS^2/σ^2 *tiene una distribución* χ^2 *con* $n-1$ *grados de libertad*.

Obsérvese que si $\chi_1^2, \chi_2^2, \dots, \chi_k^2$ son variables independientes distribuidas como χ^2 con n_1, n_2, \dots, n_k grados de libertad respectivamente, la suma $\chi_1^2 + \chi_2^2 + \dots + \chi_k^2$ será una variable distribuida como χ^2 con $n_1 + n_2 + \dots + n_k$ grados de libertad,

puesto que la suma de las χ_1^2 equivale a una suma de $n_1 + n_2 + \dots + n_k$ variables normales $N(0, 1)$. En particular, si de una población normal de varianza σ^2 , se toman k muestras independientes de tamaños n_1, n_2, \dots, n_k , y de cada una se forma la suma S_i^2 análoga a la [9.9], la cantidad

$$\frac{n_1 S_1^2 + n_2 S_2^2 + \dots + n_k S_k^2}{\sigma^2} \quad [9.10]$$

está distribuida como una χ^2 con $n_1 + n_2 + \dots + n_k - k$ grados de libertad.

Observación. Una propiedad importante de la distribución χ^2 es que ésta tiende a la distribución normal cuando el número de grados de libertad tiende a ∞ . Es decir, teniendo en cuenta los valores [9.6], resulta que la χ^2 tiende a una normal de tipo $N(n, 2n)$. Esto hace que las tablas de la distribución χ^2 estén, en general, calculadas hasta $n = 30$ solamente. Para más grados de libertad se utilizan las tablas de la función de distribución normal Φ .

2. LA DISTRIBUCION t DE STUDENT

Sea X una variable aleatoria cuya función de distribución es la χ^2 con n grados de libertad, y sea Y otra variable aleatoria, independiente de X , que sea normal $N(0, 1)$.

Interesa muchas veces la variable aleatoria

$$T = \frac{Y}{\sqrt{X/n}} \quad [9.11]$$

Se demuestra (véase, por ejemplo, Cramer⁽⁴⁾) que la función de densidad de T es

$$s_n = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \quad [9.12]$$

Esta función de densidad corresponde a la llamada *distribución t de Student*, debida al inglés W. S. Gosset, quien la publicó con el pseudónimo de Student en 1908.

El número n es el número de grados de libertad de t .

La curva $y = s_n(x)$ es simétrica respecto de $x = 0$. La probabilidad de que sea $|T| > t_0$ la suma de las dos áreas rayadas en la figura 10, y su valor, que dependa de t_0 y de n , es

$$P(|T| > t_0) = 2 \int_{t_0}^{\infty} s_n(x) dx = p. \quad [9.13]$$

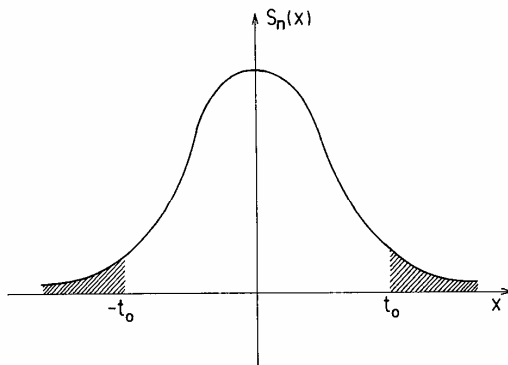


Fig. 10.

Análogamente al caso de χ^2 , se han tabulado los valores de t_0 en función de p y de n . Las tablas de la función t (véase el Apéndice III) resuelven el siguiente problema: "Dadas las variables aleatorias X e Y en las condiciones dichas, y una probabilidad p , hallar el $t_0 > 0$ para el cual es $P(|T| > t_0) = p$ ".

La esperanza y la varianza de T son (para $n > 1$ y $n > 2$, respectivamente)

$$E(T) = 0, \quad \sigma^2(T) = \frac{n}{n-2}. \quad [9.14]$$

Para $n \rightarrow \infty$, la distribución t converge a la distribución normal reducida $N(0, 1)$, de manera que para $n > 30$, en vez de las tablas de t , se utilizan las tablas de la función Φ , ya que para $n > 30$ los valores de la distribución t casi no se modifican.

Ejemplo 1. Sean X_1, X_2, \dots, X_n variables independientes normales $N(0, 1)$. Hemos visto que nS^2/σ^2 tiene una distribución χ^2 con $n-1$ grados de libertad. Por otra parte, $(\bar{X} - \alpha)/(\sigma/\sqrt{n})$ tiene distribución normal $N(0, 1)$ y se demuestra que es independiente de nS^2/σ^2 . Por tanto

$$T = \frac{\bar{X} - \alpha}{S/\sqrt{n-1}} \quad [9.15]$$

tiene una distribución de Student con $n-1$ grados de libertad. Obsérvese que en [9.15] no figura la varianza σ^2 .

Ejemplo 2. Sean $X_i (i = 1, 2, \dots, n_x)$ e $Y_j (j = 1, 2, \dots, n_y)$ muestras independientes de las variables normales X e Y , de medias

respectivas $E(X) = \alpha_x$, $E(Y) = \alpha_y$ y de la misma varianza $\sigma^2(X) = \sigma^2(Y) = \sigma^2$. Llamando S_x^2 y S_y^2 a las sumas [9.9], según [9.10], la variable aleatoria

$$\frac{n_x S_x^2 + n_y S_y^2}{\sigma^2} \quad [9.16]$$

tiene una distribución χ^2 con $n_x + n_y - 2$ grados de libertad. Por otra parte

$$\frac{(\bar{X} - \bar{Y}) - (\alpha_x - \alpha_y)}{\sigma(1/n_x + 1/n_y)^{1/2}} \quad [9.17]$$

tiene una distribución normal $N(0, 1)$, como resulta de aplicar el teorema 6.1 y teniendo en cuenta que $\sigma^2(\bar{X}) = \sigma^2/n_x$, $\sigma^2(\bar{Y}) = \sigma^2/n_y$. Se puede demostrar que [9.16] y [9.17] son independientes. Por tanto, la variable

$$T = \frac{(\bar{X} - \bar{Y}) - (\alpha_x - \alpha_y)}{[(n_x S_x^2 + n_y S_y^2)/(n_x + n_y - 2)]^{1/2} (1/n_x + 1/n_y)^{1/2}} \quad [9.18]$$

tiene una distribución de Student con $n_x + n_y - 2$ grados de libertad. Es importante observar que en esta fórmula no aparece la varianza σ^2 .

3. ESTIMACION POR INTERVALOS DE CONFIANZA

109

Volvamos ahora a la estimación de parámetros de una distribución de probabilidad. Consideremos una población descrita por la variable aleatoria X , cuya distribución depende de un parámetro desconocido θ . Si x_1, x_2, \dots, x_n es una muestra de la población, se desea encontrar dos funciones $a(x_1, x_2, \dots, x_n)$ y $b(x_1, x_2, \dots, x_n)$ tales que

$$P(a \leq \theta \leq b) = 1 - \eta \quad [9.19]$$

donde η es un valor dado, en general próximo a 0. El intervalo $[a, b]$ se llama un *intervalo de confianza* y $1 - \eta$ es el *coeficiente de confianza*.

Aunque existen varios criterios generales para hallar intervalos de confianza, nos vamos a limitar a los ejemplos más comunes de una población normal o "asintóticamente normal" (es decir, cuya distribución de probabilidad sea aproximadamente normal para muestras grandes).

Ejemplo 1. Distribución normal de varianza σ^2 conocida. En este caso, aplicando [6.25] a la variable $\bar{X} = (1/n)(X_1 + X_2 + \dots + \dots + X_n)$, tenemos

$$P(\alpha - x\sigma/\sqrt{n} < \bar{X} < \alpha + x\sigma/\sqrt{n}) = 1 - \eta \quad [9.20]$$

o sea

$$P(\bar{X} - x\sigma/\sqrt{n} < \alpha < \bar{X} + x\sigma/\sqrt{n}) = 1 - \eta \quad [9.21]$$

donde, por comparación con [6.24], es $\eta = 2(1 - \Phi(x))$. Dado η , las tablas de la función Φ permiten calcular x , y entonces tenemos, para la media α , el intervalo $(\bar{X} - x\sigma/\sqrt{n}, \bar{X} + x\sigma/\sqrt{n})$ cuyo coeficiente de confianza es $1 - \eta$.

Por ejemplo, dado $\sigma = 0,003$, $\eta = 0,05$, $\bar{x} = 0,124$, $n = 100$, se encuentra $x = 1,96$, y el intervalo en que se encuentra la media desconocida α , con un coeficiente de confianza $1 - \eta = 0,95$, es $(0,1234 \leq \alpha \leq 0,1246)$.

Ejemplo 2. Distribución normal de varianza desconocida.

a) Intervalo de confianza para la media. Si σ^2 no se conoce, no se puede aplicar [9.20], pero sí la variable aleatoria [9.15], que no depende de σ^2 y tiene una distribución de Student con $n - 1$ grados de libertad. Dado el coeficiente de confianza $1 - \eta$, las tablas de la distribución t nos dan el valor $t(n - 1; \eta)$ tal que

$$P(-t(n-1; \eta) \leq \frac{\bar{X} - \alpha}{S/\sqrt{n-1}} \leq t(n-1; \eta)) = 1 - \eta \quad [9.22]$$

110

que también se puede escribir

$$P(\bar{X} - t(n-1; \eta) \frac{S}{\sqrt{n-1}} \leq \alpha \leq \bar{X} + t(n-1; \eta) \frac{S}{\sqrt{n-1}}) = 1 - \eta. \quad [9.23]$$

Se tiene así un intervalo de confianza para α con el coeficiente $1 - \eta$. Por ejemplo, dados $\eta = 0,05$, $n = 17$, $\bar{x} = 4$, $s = 0,84$, las tablas dan $t(16; 0,05) = 2,12$ y resulta el intervalo de confianza $3,565 \leq \alpha \leq 4,445$, cuyo coeficiente de confianza es $0,95$.

b) Intervalo de confianza para la varianza. Sabemos que la variable aleatoria nS^2/σ^2 tiene distribución χ^2 con $n - 1$ grados de libertad. Dado $1 - \eta$, hay que hallar a, b tales que

$$P(a \leq \frac{nS^2}{\sigma^2} \leq b) = 1 - \eta. \quad [9.24]$$

Hay muchas maneras de elegir a, b que cumplan esta condición. Por comodidad de cálculo se suelen elegir de manera que sea

$$P(b < \frac{nS^2}{\sigma^2}) = \frac{\eta}{2}, \quad P(a \leq \frac{nS^2}{\sigma^2}) = 1 - \frac{\eta}{2}. \quad [9.25]$$

Los valores de a y b resultan de las tablas de χ^2 y entonces se cumple [9.24], que se puede escribir

$$P(\frac{nb}{a} \leq \sigma^2 \leq \frac{nS^2}{a}) = 1 - \eta. \quad [9.26]$$

Por ejemplo, dados $\eta = 0, 10$, $n = 20$, $s^2 = 1.000$, utilizando las tablas de χ^2 (para $n - 1$ grados de libertad), para que se cumplan las condiciones [9.25] resulta $a = 10, 117$ y $b = 30, 144$, y por tanto se tiene el intervalo $663 \leq \sigma^2 \leq 1.976$, con el coeficiente de confianza 0, 90.

4. VERIFICACION DE HIPOTESIS

Supongamos nuevamente una variable aleatoria X con función de probabilidad o de densidad $f(x; \theta)$ dependiente de un parámetro θ . Supongamos conocida la forma de la función f , pero no el parámetro θ . Se hace una hipótesis, H_0 , de que sea $\theta = \theta_0$. Se trata entonces, mediante una muestra, de verificar el grado de certeza de esta hipótesis. Por costumbre se dice que se trata de verificar la "hipótesis nula". A veces, junto con la H_0 se formulan otras hipótesis H_1 , por ejemplo, $\theta = \theta_1$ ó $\theta < \theta_0$ ó $\theta > \theta_0$. Se trata entonces de verificar también estas hipótesis, llamadas "hipótesis alternativas". Las líneas generales del método a seguir son las siguientes:

Toda muestra $x = (x_1, x_2, \dots, x_n)$, por ser un conjunto de n números reales, se puede representar por un punto x del espacio euclidiano de n dimensiones. Elaborar un test para verificar la hipótesis H_0 , significa buscar una región R de dicho espacio, tal que: a) Si la muestra $x = (x_1, x_2, \dots, x_n)$ corresponde a un punto de R , se rechaza la hipótesis H_0 ; b) Si la muestra x corresponde a un punto exterior a R , se acepta la hipótesis H_0 .

Si sólo hay dos alternativas H_0 ó H_1 , rechazar H_0 equivale a aceptar H_1 . Rechazar H_0 ($\theta = \theta_0$) cuando en verdad es $\theta = \theta_0$, se llama *error de tipo I*. La probabilidad del error de tipo I es

$$P(x \in R; \theta = \theta_0) = \epsilon, \quad [9.27]$$

que se llama nivel de significación.

Si sólo hay la alternativa H_1 ($\theta = \theta_1$), no rechazar H_0 cuando en realidad debiera hacerse, por ser $\theta = \theta_1$, se llama *error de tipo II*, y su probabilidad es $P(x \notin R; \theta = \theta_1) = \eta$. En este caso ($\theta = \theta_1$, como única alternativa) nos interesaremos también por la probabilidad

$$P(x \in R; \theta = \theta_1) = 1 - \eta \quad [9.28]$$

que se llama *potencia*.

Definición 9.1. Se llama *función de potencia* del test definido por la región R con respecto al valor θ del parámetro, a la probabilidad

$$P(x \in R; \theta). \quad [9.29]$$

Todo el problema consiste en hallar la región R , dados ϵ, η . Consideremos el siguiente problema ilustrativo.

Sea una población normal $N(\alpha, \sigma^2)$ con σ^2 conocida y α desconocida. Hacemos la hipótesis nula $\alpha = \alpha_0$. Queremos preparar un test de esta hipótesis. Sea $\bar{X} = (1/n)(X_1 + X_2 + \dots + X_n)$. Considerando una probabilidad $\epsilon = 0,05$ sabemos que

$$P(\alpha_0 - 1,96\sigma(\bar{X}) \leq \bar{X} \leq \alpha_0 + 1,96\sigma(\bar{X})) = 0,95 \quad [9.30]$$

que, poniendo $\sigma(\bar{X}) = \sigma/\sqrt{n}$ (σ es dato), se puede escribir

$$P(\bar{X} \leq \alpha_0 - 1,96\sigma/\sqrt{n}) + P(\bar{X} \geq \alpha_0 + 1,96\sigma/\sqrt{n}) = 0,05. \quad [9.31]$$

Por tanto, un test de la hipótesis $\alpha = \alpha_0$ al 5% puede ser: "Se elige una muestra y se calcula su media $\bar{x} = (1/n)(x_1 + x_2 + \dots + x_n)$. Si esta media cae fuera del intervalo $[\alpha_0 - 1,96\sigma/\sqrt{n}, \alpha_0 + 1,96\sigma/\sqrt{n}]$ se rechaza la hipótesis $\alpha = \alpha_0$ ".

La probabilidad de equivocarse, o sea, de rechazar la hipótesis siendo cierta, es 0,05.

Obsérvese que el test construido no es único. El mismo método permite construir otros muchos. Por ejemplo, se cumple también

$$P(\bar{X} \geq \alpha_0 + 1,65\sigma/\sqrt{n}) = 0,05$$

y por tanto, otro test podría ser: "Si la media de la muestra tiene un valor superior a $\alpha_0 + 1,65\sigma/\sqrt{n}$, se rechaza la hipótesis $\alpha = \alpha_0$ ". La probabilidad de error de tipo I es la misma 0,05 de antes.

La teoría de verificación de hipótesis da reglas o criterios para elegir el más conveniente, dentro de la naturaleza del problema, de los tests posibles.

5. COMPARACION DE DISTRIBUCIONES EXPERIMENTALES Y TEORICAS: TEST DEL χ^2

Sea X una variable aleatoria cuyo dominio es una cierta población. Hasta ahora hemos supuesto conocida la función de probabilidad o de densidad, salvo un parámetro de la misma. Supongamos ahora que desconocemos la distribución de probabilidad. En este caso se hace la hipótesis de que esta distribución es una determinada y se trata de verificar, mediante una muestra, si esta hipótesis es correcta o no. Para ello es útil el método del χ^2 , que pasamos a exponer.

Supongamos que la variable aleatoria X puede tomar los valores x_1, x_2, \dots, x_r y que, de acuerdo con la hipótesis hecha, las proba-

bilidades respectivas sean p_1, p_2, \dots, p_r , a las que llamaremos probabilidades teóricas. Tomemos una muestra de tamaño n y sean n_1, n_2, \dots, n_r los números de veces que aparecen, respectivamente, los valores x_1, x_2, \dots, x_r de manera que $n_1 + n_2 + \dots + n_r = n$. Formemos el estadístico

$$u = \sum_1^r \frac{(n_i - np_i)^2}{np_i} \quad [9.32]$$

que, en cierto modo, ya se ve que mide la discrepancia, en valor absoluto, entre los valores medios teóricos np_i y los obtenidos en la muestra. Un teorema muy importante de Pearson, cuya demostración no damos (véase, por ejemplo, Ríos ⁽³⁾), dice que *la distribución del estadístico u , para n grande, tiende a una distribución χ^2 con $r - 1$ grados de libertad*, cuando la hipótesis planteada es cierta.

Es decir, aunque no sea exacto (pues sólo lo es para $n \rightarrow \infty$), se acostumbra poner $u = \chi^2$ y entonces las tablas de χ^2 , dados r y ϵ , permiten calcular χ_0^2 tal que

$$P(\chi^2 > \chi_0^2) = \epsilon. \quad [9.33]$$

Entonces, fijado el nivel de significación ϵ , si el valor [9.32] es superior al χ_0^2 obtenido, rechazamos la hipótesis.

113

Problema 9.1. *Se quiere verificar la hipótesis de que el nacimiento de varones sigue la ley binomial con $p = 1/2$. Se tiene una muestra de 100.000 nacimientos, de los cuales 51.400 de los nacidos han sido varones y el resto mujeres. Este resultado ¿confirma o niega la hipótesis hecha, al nivel de significación del 1 por 1.000?*

Solución. Si la hipótesis es cierta, es $p_1 = p_2 = 1/2$, y [9.32] nos da

$$u = \frac{(51.400 - 50.000)^2}{50.000} + \frac{(48.600 - 50.000)^2}{50.000} = 78,4.$$

Puesto que $r = 2$, buscamos χ^2 con 1 grado de libertad. Las tablas nos dicen que $P(\chi^2 > 10,83) = 0,001$. Por tanto, utilizamos la región crítica $(10,83, \infty)$. Como el valor obtenido para el estadístico $u(78,4)$ pertenece a esta región crítica, la hipótesis debe rechazarse. La probabilidad del nacimiento de un varón es superior a la del nacimiento de una mujer.

El mismo problema fue tratado directamente en el problema 6.7, y se llegó al mismo resultado. El hecho es general en el sentido que: *cundo la variable aleatoria puede tomar sólo dos va-*

lores, $r = 2$, el método del χ^2 coincide con el método que consiste en utilizar la fórmula [6.31]. En efecto, en este caso, [9.32] se puede escribir

$$u = \frac{(n_1 - np)^2}{np} + \frac{(n_2 - nq)^2}{nq} = \frac{(n_1/n - p)^2}{pq/n}$$

y, por tanto, $P(u > u_0)$ se puede calcular por la fórmula [6.31].

Problema 9.2. Se lanza un dado 6.000 veces y resulta que las caras 1, 2, 3, 4, 5, 6 salen, respectivamente, los siguientes números de veces

800, 1.080, 960, 1.086, 1.304, 770.

¿Es el dado perfecto? Es decir, son estos resultados compatibles con la hipótesis $p_i = 1/6$ ($i = 1, 2, 3, 4, 5, 6$)? Se desea un nivel de significación igual a 0,001.

Solución. Aplicando el método del χ^2 al caso $p = 1/6$, resulta

$$u = \frac{1}{1.000}(200^2 + 80^2 + 40^2 + 86^2 + 304^2 + 230^2) = 200,69.$$

114

El número de grados de libertad es 5. Según las tablas de χ^2 , el valor 200 queda sin duda en la región crítica. Es decir, se concluye que, o bien el dado es defectuoso, o el experimento ha sido mal realizado.

Problema 9.3. De un conjunto de 17 plantas, 9 de ellas fueron sometidas a un tratamiento químico y el número de frutos fue, respectivamente,

17, 27, 18, 25, 27, 29, 27, 23, 17

mientras que las no sometidas al tratamiento, dieron los siguientes números de frutos

16, 16, 20, 16, 20, 17, 15, 21.

¿Puede considerarse que el tratamiento surtió efecto?

Solución. Suponiendo una distribución normal, podemos aplicar la distribución de la variable aleatoria [9.18] para verificar la hipótesis $\alpha_x = \alpha_y$. Tenemos $\bar{X} = 23,33$, $\bar{Y} = 17,62$, $(S_x^2 n_x + S_y^2 n_y) / (n_x + n_y - 2) = 14,79$, con lo cual resulta $t = 3,05$. El número de grados de libertad es 15. Por tanto, según las tablas de la función t de Student, el valor 3,05 queda fuera del intervalo de

las t , para el nivel de significación 0,01. Por tanto a este nivel de significación las dos muestras no pueden considerarse procedentes de una misma población, o sea, debe considerarse que el tratamiento surtió efecto.

Problema 9.4. *Un test de inteligencia entre 20 alumnos de una población X da un puntaje medio de 90 y una varianza de 100. El mismo test aplicado a 30 alumnos de la población Y da un puntaje medio de 110 y una varianza de 80. Se desea saber, al nivel de significación 0,01, si hay diferencias significativas entre los alumnos de las dos poblaciones.*

Solución. Suponiendo que el puntaje tiene una distribución normal, aplicamos el test de Student [9.18] para ver si la hipótesis $\alpha_x = \alpha_y$ de igual promedio es aceptable. Es $\bar{X} = 90$, $\bar{Y} = 80$, $n_x = 20$, $n_y = 30$, $s_x^2 = 100$, $s_y^2 = 80$. Por tanto, sustituyendo en [9.18] resulta $t = 3,62$. Como los grados de libertad son 48, las tablas de la función t , al nivel 0,01, dan $t_0 = 2,57$. Luego, como 3,62 queda fuera de este intervalo, debemos rechazar la hipótesis de igual promedio de inteligencia. El rechazo está fundado en que, si el promedio fuera el mismo, la probabilidad de resultados con $|t| > 2,57$ es 0,01 (uno por ciento).

115

En los problemas 9.1 y 9.2 se conocen las probabilidades teóricas p_i para poder escribir el estadístico u [9.32]. Hay veces, sin embargo, en que estas probabilidades teóricas no se conocen y hay que estimarlas mediante la misma muestra. En este caso, bajo condiciones muy amplias y, con ciertas precauciones sobre las que aquí no podemos entrar, también se puede aplicar el criterio del χ^2 , sustituyendo las probabilidades teóricas por las estimadas y prestando atención al número de grados de libertad, que es igual al número de parámetros independientes menos 1. El método resulta de mucha aplicación práctica, si bien hay que tener cuidado en casos extremos. Vamos a dar tres ejemplos típicos, que pueden servir de modelo para muchos casos análogos. El primer problema que sigue está tomado, con ligera variación en el enunciado, del clásico libro de Yule y Kendall titulado "Introduction to the Theory of Statistics":

Problema 9.5. *Un país está dividido en 8 regiones y se desea saber si la proporción de fumadores varía de una región a otra. Para ello se toma una muestra de cada una de estas regiones y los resultados se resumen en la siguiente tabla:*

Región	1	2	3	4	5	6	7	8	Total
Fumadores	56	87	142	71	88	72	100	142	758
No fumadores	17	20	58	20	31	23	25	48	242
Totales	73	107	200	91	119	95	125	190	1.000

¿Se puede inferir de estos resultados que la proporción de fumadores varía de una región a otra?

Solución. Si se hace la hipótesis de que la proporción no varía, la probabilidad experimental de que una persona sea fumador, calculada a partir de la muestra anterior, es $p = 0,758$, con lo cual resulta

$$\begin{aligned}\chi^2 &= (56 - 55)^2/55 + (87 - 81)^2/81 + (142 - 152)^2/152 + (71 - 69)^2/69 + \\ &+ (88 - 90)^2/90 + (72 - 72)^2/72 + (100 - 95)^2/95 + (142 - 144)^2/144 + \\ &+ (17 - 18)^2/18 + (20 - 26)^2/26 + (58 - 48)^2/48 + (20 - 22)^2/22 + \\ &+ (31 - 29)^2/29 + (23 - 23)^2/23 + (25 - 30)^2/30 + (48 - 46)^2/46 = \\ &= 6,28.\end{aligned}$$

116

El número de grados de libertad es 7, puesto que el número de parámetros independientes es 8. En efecto, sabiendo el tamaño de cada muestra, para conocer todo el experimento, basta conocer, por ejemplo, el número de fumadores de cada región. Tomando el nivel de significación 0,05 las tablas dan $\chi_0^2 = 14,06$. Puesto que 6,28 queda fuera de la región crítica, se deduce que el criterio del χ^2 no permite deducir que haya diferencia alguna en la proporción de fumadores de una región a otra.

Problema 9.6. *Hay que examinar 240 alumnos, los cuales se distribuyen entre 3 examinadores A, B y C. Los resultados obtenidos son los siguientes: el examinador A aprueba 60 alumnos, el B aprueba 70 y el C aprueba 68. Se desea verificar la hipótesis de que los 3 examinadores sean igualmente exigentes.*

Solución. Si los 3 examinadores son igualmente exigentes, la probabilidad de aprobar, estimada experimentalmente de la muestra, es $p = (60 + 70 + 68)/240 = 0,825$. Por tanto se tiene:

$$\begin{aligned}\chi^2 &= (60 - 66)^2/66 + (70 - 66)^2/66 + (68 - 66)^2/66 + (20 - 14)^2/14 + \\ &+ (10 - 14)^2/14 + (12 - 14)^2/14 = 4,8.\end{aligned}$$

El número de grados de libertad es 2 (puesto que hay 3 parámetros independientes, que son el número de alumnos aprobados por cada examinador). La tabla de χ^2 , al nivel 0,05 da $\chi_0^2 = 5,99$. Siendo este valor superior a 4,8, la hipótesis es aceptable, o sea, los 3 examinadores pueden considerarse igualmente exigentes.

Problema 9.7. *Se quiere verificar la eficacia de una vacuna. Se consideran 7.000 personas vacunadas y se observa que de ellas 60 han adquirido la enfermedad y no así las restantes 6.940. En cambio, de 12.000 personas no vacunadas se observa que 200 han adquirido la enfermedad y 11.800 no. ¿Qué puede deducirse de ello?*

Solución. Vamos a verificar la hipótesis de que la vacuna sea ineficaz. En tal caso la probabilidad de contraer la enfermedad, deducida de la muestra, es $p = (60 + 200)/19.000 = 0,0136$. De aquí

$$\chi^2 = (6.940 - 6.904)^2/6.904 + (11.800 - 11.836)^2/11.836 + (60 - 96)^2/96 + (200 - 164)^2/164 = 21,6.$$

El número de grados de libertad es 1, puesto que los parámetros independientes pueden ser los números de personas que adquieren la enfermedad. Por tanto, al nivel de significación 0,05 es $\chi_0^2 = 3,84$. Por consiguiente, la hipótesis debe rechazarse, o sea, la vacuna es eficaz o bien, por lo menos, tal es la conclusión a que llega el criterio del χ^2 .

Problema 9.8. *Se desea saber el porcentaje de televidentes que sintonizan un determinado programa de televisión, a partir de una muestra tomada al azar. Se admite un error de $\pm 0,05$ y un coeficiente de confianza de 0,95. Se pide el tamaño de la muestra a tomar, suponiendo muy grande el número de televidentes.*

Solución. En este tipo de problemas cabe suponer que la distribución de la muestra es la normal y que, por tanto, se puede aplicar [9.31], teniendo en cuenta que, para una variable de Bernoulli, es $\sigma = p(1-p)$; p es la probabilidad de que un televidente sintonice el programa en cuestión. Resulta $1,96\sigma/\sqrt{n} \leq 0,05$ y por tanto $n \geq (1,96/0,05)^2\sigma^2$. Si p no se conoce hay que usar la desigualdad $p(1-p) \leq 1/4$, y resulta que basta tomar una muestra de $n \geq (1,96/0,05)^2(1/4) = 384$ televidentes.

Si el error admisible fuera de $\pm 0,02$, el tamaño de la muestra debiera ser $n \geq (1,96/0,02)^2(1/4) = 2.401$ televidentes. Si por experiencia con otros programas análogos, se conociera el valor de p , el tamaño de la muestra se puede disminuir y el verdadero valor de p se deduce de la expresión $\sigma = p(1-p)$.

Con respecto al tamaño de las muestras destinadas a averiguar una determinada característica de una población, supuesta numerosa, el ejemplo anterior está incluido en el siguiente teorema general: "Si una población tiene una determinada característica en la proporción p , la proporción de la misma característica en las muestras de tamaño n es una variable aleatoria aproximadamente normal, de valor medio p y varianza $[p(1 - p)/n]^{1/2n}$."

APENDICE I

COMBINATORIA

Resumimos en este Apéndice algunas definiciones y fórmulas de análisis combinatorio que son fundamentales para muchos problemas de probabilidades en conjuntos finitos.

1. Permutaciones

Una *permutación* de n elementos, es una disposición de los mismos en un determinado orden.

Por ejemplo, con dos elementos a, b se pueden formar las permutaciones

$$ab, ba$$

y con tres elementos a, b, c las seis permutaciones siguientes

$$abc, bac, cab, bac, acb, cba.$$

En general, dados n elementos, para formar una permutación se elige uno cualquiera de ellos (n posibilidades), después se elige uno cualquiera de los $n - 1$ restantes ($n - 1$ posibilidades) y así sucesivamente. Resulta que *el número de permutaciones con n elementos es*

$$P_n = n(n-1)(n-2)\dots 2.1 = n! \quad [1]$$

Este número se llama *factorial* de n y está definido para todo número natural. Por convención, se define $0! = 1$.

A veces interesan las *permutaciones de r elementos elegidos entre n elementos dados* ($n \geq r$), llamadas también "variaciones". Para formar una de ellas se elige primero un elemento de los n dados (n posibilidades), después un segundo elemento entre los $n - 1$ restantes ($n - 1$ posibilidades) y así sucesivamente hasta el r -ésimo elemento, que será elegido entre los $n - r + 1$ elementos restantes. Resulta así que *el número de permutaciones que se pueden formar de r elementos elegidos entre los de un conjunto dado de n elementos* ($n \geq r$) es

$$P_{n,r} = n(n-1)(n-2)\dots(n-r+1) = \frac{n!}{(n-r)!} \quad [2]$$

2. Combinaciones

Una *combinación* de r elementos de un conjunto de n elementos (llamada combinación de tipo (n, r)) es todo subconjunto de r elementos del conjunto de los n dados.

Por ejemplo, si los n elementos son las letras a, b, c, d, e las combinaciones de $r = 3$ elementos son las 10 siguientes

$$abc, abd, abe, bcd, bce, acd, ace, bde, ade, cde.$$

El número de combinaciones de r elementos entre n dados se representa por $C_{n,r}$. Para encontrarlo, observemos que si en cada combinación cambiamos de todas las maneras posibles el orden de los elementos, tendremos $r!$ permutaciones de r elementos elegidos entre los n dados. Por tanto debe ser $C_{n,r} \cdot r! = P_{n,r}$. Teniendo en cuenta [2] resulta que el *número de combinaciones de r elementos elegidos entre los de un conjunto de n elementos es*

$$C_{n,r} = \frac{n!}{(n-r)!r!} = \binom{n}{r}. \quad [3]$$

El símbolo $\binom{n}{r}$ se llama el *número combinatorio de n objetos tomados r a r* .

3. Combinaciones con Repetición

Una *combinación con repetición* de tipo (n, r) es un conjunto de r elementos, distintos o repetidos, elegidos entre los de un conjunto dado de n elementos.

Por ejemplo, con 3 elementos a, b, c se pueden formar las siguientes combinaciones con repetición de 3 elementos

$$aaa, aab, aac, abb, abc, acc, bbb, bbc, bec, ccc.$$

Para calcular el número $C_{n,r}'$ de combinaciones con repetición de tipo (n, r) se puede proceder de la manera siguiente. Supóngase que los n elementos son a_1, a_2, \dots, a_n . Cada combinación con repetición se puede ordenar con los índices en orden creciente y luego (para distinguir los índices repetidos) añadir a cada índice el número de orden que tiene menos uno. Por ejemplo, si $n = 5$, $r = 4$, la combinación $a_2a_3a_3a_4$ se escribiría $a_2a_4a_5a_7$. Se tiene así siempre una combinación sin repetición de tipo $(n+r-1, r)$. Recíprocamente, dada una combinación de este tipo, rebajando cada índice, una vez ordenada, de un entero igual al número de orden

menos uno, tendremos una combinación con repetición de tipo (n, r) . Por tanto: *el número de las combinaciones con repetición de tipo (n, r) es igual al de las combinaciones sin repetición del tipo $(n + r - 1, r)$, o sea*

$$C'_{n,r} = C_{n+r-1,r} = \binom{n+r-1}{r} = \frac{(n+r-1)!}{(n-1)!r!}. \quad [4]$$

4. Particiones

Interesa a veces el número de maneras en que un conjunto de n elementos se puede repartir en l subconjuntos de r_h ($h = 1, 2, \dots, l$) elementos cada uno ($l \leq n; r_1 + r_2 + \dots + r_l = n$). El número de estas *particiones* se representa por $P_{r_1, r_2, \dots, r_l}^n$, y para obtenerlo basta observar que tomando una de las $n!$ permutaciones de los n elementos y permutando entre sí los r_1 primeros (lo que da $r_1!$ casos), luego los r_2 siguientes y así sucesivamente, siempre se obtendrá una misma partición. Encambio, partiendo de otra permutación, las mismas operaciones dan lugar a una partición distinta. Por tanto

$$P_{r_1, r_2, \dots, r_l}^n = \frac{n!}{r_1! \dots r_l!}. \quad [5]$$

5. Fórmula de Stirling

El cálculo del valor exacto de $n! = n(n-1)(n-2) \dots 2 \cdot 1$, resulta pesado cuando n es un tanto grande. Algunos valores se dan en la Tabla I del Apéndice III. Para el cálculo con logaritmos es útil la siguiente fórmula aproximada de Stirling (1730)

$$n! \sim \sqrt{2\pi n} n^n e^{-n}. \quad [6]$$

Decir que el segundo miembro es un valor *aproximado* del primero, cuando n es grande, quiere decir que el límite del segundo miembro dividido por el primero tiende a la unidad para $n \rightarrow \infty$, aunque los valores de ambos miembros de [6] pueden diferir en cantidades grandes.

6. El Binomio de Newton

Los números combinatorios $\binom{n}{r}$ aparecen como coeficientes de la fórmula del binomio de Newton.

$$(a + b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \dots + \binom{n}{r} a^{n-r} b^r + \dots + \binom{n}{n} b^n. \quad [7]$$

En particular, para $a = 1, b = 1$ y para $a = 1$ y $b = -1$ resultan las fórmulas

$$\binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n} = 2^n, \quad \binom{n}{0} - \binom{n}{1} + \dots + (-1)^n \binom{n}{n} = 0.$$

APENDICE II

ALGUNAS FORMULAS DE CALCULO

1. El Número e

El número $e = 2,71828\dots$ se define por

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n. \quad [1]$$

Más generalmente, para cualquier x se tiene

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n. \quad [2]$$

La fórmula de Mac Laurin da inmediatamente el siguiente desarrollo en serie

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + \dots \quad [3]$$

En particular

$$e = 2 + \frac{1}{2!} + \frac{1}{3!} + \dots, \quad e^{-1} = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots$$

2. Integrales de la Función de Densidad Normal

Sea

$$f(x) = a e^{-b(x-c)^2}, \quad a, b > 0.$$

Queremos calcular

$$I = \int_{-\infty}^{\infty} f(x) dx = a \int_{-\infty}^{\infty} e^{-b(x-c)^2} dx \quad [4]$$

o sea, por el cambio de variables $x - c = y$

$$I = a \int_{-\infty}^{\infty} e^{-by^2} dy \quad [5]$$

Para ello se sigue el clásico artificio siguiente:

$$I^2 = \left(a \int_{-\infty}^{\infty} e^{-by^2} dy \right) \left(a \int_{-\infty}^{\infty} e^{-bz^2} dz \right) = a^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-b(y^2+z^2)} dy \cdot dz$$

y pasando a coordenadas polares ($y = r \cos \theta$, $z = r \sin \theta$),

$$I^2 = a^2 \int_0^{2\pi} \int_0^{\infty} e^{-br^2} r dr d\theta = 2\pi a^2 \int_0^{\infty} r e^{-br^2} dr = \pi a^2 / b.$$

Por tanto

$$I = a \sqrt{\frac{\pi}{b}}. \quad [6]$$

En particular, para $a = 1/\sqrt{2\pi}$, $b = 1/2$, $c = 0$, resulta

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1. \quad [7]$$

Si queremos que $f(x)$ sea una función de probabilidad, debe ser $I = 1$ y por tanto

$$f(x) = \sqrt{\frac{b}{\pi}} e^{-b(x-c)^2}, \quad b > 0. \quad [8]$$

La esperanza matemática correspondiente a esta función será

$$\alpha = E(X) = \int_{-\infty}^{\infty} x f(x) dx = \sqrt{\frac{b}{\pi}} \int_{-\infty}^{\infty} x e^{-b(x-c)^2} dx = c.$$

La varianza es

$$\sigma^2(X) = \int_{-\infty}^{\infty} (x-c)^2 f(x) dx = \sqrt{\frac{b}{\pi}} \int_{-\infty}^{\infty} (x-c)^2 e^{-b(x-c)^2} dx,$$

que se calcula integrando por partes, y resulta

$$\sigma^2(X) = \frac{1}{2b}.$$

Por tanto, poniendo de manifiesto la media α y la varianza σ^2 , se tiene

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\alpha)^2/2\sigma^2}.$$

La función generatriz de momentos es

$$\psi(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-(x-\alpha)^2/2\sigma^2} dx.$$

Para calcular esta integral se observa que el exponente se puede escribir

$$tx - \frac{(x - \alpha)^2}{2\sigma^2} = \frac{-1}{2\sigma^2}((x - \alpha - t\sigma^2)^2 - 2t\alpha\sigma^2 - t^2\sigma^4)$$

y haciendo el cambio de variables $x - \alpha - t\sigma^2 = z\sigma$ y aplicando [7], resulta

$$\psi(t) = e^{\alpha t + \sigma^2 t^2 / 2}.$$

Los momentos centrados de [8] son

$$\mu_{2k+1} = 0, \quad \mu_{2k} = \frac{(2k)!}{2^k (k!)} \sigma^{2k}. \quad [9]$$

3. La Función Gamma

Se define por la integral

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

Esta función es positiva y tiende a ∞ para $x \rightarrow 0$ y para $x \rightarrow \infty$. Presenta un mínimo para $x_0 = 1,4616$, en el cual la función vale $\Gamma(x_0) = 0,8856$. Para valores naturales de x , sea $x = n + 1$, la función gamma da el factorial de n , o sea

$$\Gamma(n + 1) = n! = n(n - 1)(n - 2) \dots 2 \cdot 1. \quad [10]$$

APENDICE III

Tabla 1

Tabla de Factoriales $n!$

n	$n!$
1	1
2	2
3	6
4	24
5	120
6	720
7	5040
8	40320
9	362880
10	3 628800
11	39 916800
12	479 001600
13	6227 020800
14	87178 291200
15	1 307674 368000
16	20 922789 888000
17	355 687428 096000
18	6402 373705 728000
19	121645 100408 832000
20	2 432902 008176 640000

Tabla 2
Valores de la Función de Probabilidad de Poisson $P_r = \frac{\lambda^r}{r!} e^{-\lambda}$

$\frac{\lambda}{r}$	0,2	0,4	0,6	0,8	1	2	3	4	5	7	10
0	0,819	0,670	0,549	0,449	0,368	0,135	0,050	0,018	0,007	0,001	0,000
1	0,164	0,268	0,329	0,359	0,368	0,271	0,149	0,073	0,034	0,006	0,000
2	0,016	0,054	0,099	0,144	0,184	0,271	0,224	0,146	0,084	0,022	0,002
3	0,001	0,007	0,020	0,038	0,061	0,180	0,224	0,195	0,140	0,052	0,008
4		0,001	0,003	0,008	0,015	0,090	0,168	0,195	0,175	0,091	0,019
5				0,001	0,003	0,036	0,101	0,156	0,175	0,128	0,038
6						0,012	0,050	0,104	0,146	0,149	0,063
7						0,003	0,022	0,059	0,104	0,149	0,090
8						0,001	0,008	0,030	0,065	0,130	0,113
9							0,003	0,013	0,036	0,101	0,125
10							0,001	0,005	0,018	0,071	0,125
11								0,002	0,008	0,045	0,114
12								0,001	0,003	0,026	0,095
13									0,001	0,014	0,073
14										0,007	0,052
15										0,003	0,035

Tabla 3

Valores de la Función de Densidad de la Distribución Normal $N(0, 1)$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

x	$f(x)$	x	$f(x)$
0	0,3989	2	0,0540
0,1	0,3969	2,1	0,0440
0,2	0,3910	2,2	0,0355
0,3	0,3814	2,3	0,0283
0,4	0,3683	2,4	0,0224
0,5	0,3521	2,5	0,0175
0,6	0,3332	2,6	0,0136
0,7	0,3122	2,7	0,0104
0,8	0,2897	2,8	0,0079
0,9	0,2661	2,9	0,0059
1	0,2420	3	0,0044
1,1	0,2178	3,1	0,0033
1,2	0,1942	3,2	0,0024
1,3	0,1714	3,3	0,0017
1,4	0,1497	3,4	0,0012
1,5	0,1295	3,5	0,0009
1,6	0,1109	3,6	0,0006
1,7	0,0940	3,7	0,0004
1,8	0,0789	3,8	0,0003
1,9	0,0656	3,9	0,0002

Tabla 4

Tabla de la Función de Distribución Normal

$$\Phi(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx$$

x	$\Phi(x)$	x	$\Phi(x)$
0,0	0,5000	2,0	0,9772
0,1	0,5398	2,1	0,9821
0,2	0,5793	2,2	0,9861
0,3	0,6179	2,3	0,9893
0,4	0,6554	2,4	0,9918
0,5	0,6915	2,5	0,9938
0,6	0,7257	2,6	0,9953
0,7	0,7580	2,7	0,9965
0,8	0,7881	2,8	0,9974
0,9	0,8159	2,9	0,9981
1,0	0,8413	3,0	0,9987
1,1	0,8643	3,1	0,9990
1,2	0,8849	3,2	0,9993
1,3	0,9032	3,3	0,9995
1,4	0,9192	3,4	0,9997
1,5	0,9332	3,5	0,9998
1,6	0,9452		
1,7	0,9554		
1,8	0,9641		
1,9	0,9713		

Tabla 5

Distribución de χ^2 : $P(\chi^2 > \chi_0^2) = \int_{\chi_0^2}^{\infty} k_{\chi^2}(x) dx$

Valores de χ_0^2

$\frac{p}{n}$	0,99	0,95	0,90	0,10	0,05	0,025	0,01	0,001
1	0,00	0,00	0,02	2,71	3,84	5,02	6,63	10,83
2	0,02	0,10	0,21	4,61	5,99	7,38	9,21	13,82
3	0,12	0,35	0,58	6,25	7,81	9,35	11,34	16,27
4	0,30	0,71	1,06	7,78	9,49	11,14	13,28	18,47
5	0,55	1,15	1,61	9,24	11,07	12,83	15,09	20,52
6	0,87	1,64	2,20	10,64	12,59	14,45	16,81	22,46
7	1,24	2,17	2,83	12,02	14,07	16,01	18,47	24,32
8	1,65	2,73	3,49	13,36	15,51	17,53	20,09	26,13
9	2,09	3,33	4,17	14,68	16,92	19,02	21,67	27,88
10	2,56	3,94	4,87	15,99	18,31	20,48	23,21	29,59
11	3,05	4,57	5,58	17,27	19,67	21,92	24,72	31,26
12	3,57	5,23	6,30	18,55	21,03	23,34	26,22	32,91
13	4,11	5,89	7,04	19,81	22,36	24,74	27,69	34,53
14	4,66	6,57	7,79	21,06	23,68	26,12	29,14	36,12
15	5,23	7,26	8,55	22,31	25,00	27,49	30,58	37,70
16	5,81	7,96	9,31	23,54	26,30	28,84	32,00	39,25
17	6,41	8,67	10,08	24,77	27,59	30,19	33,41	40,79
18	7,01	9,39	10,86	25,99	28,87	31,53	34,80	42,31
19	7,63	10,12	11,65	27,20	30,14	32,85	36,19	43,82
20	8,26	10,85	12,44	28,41	31,41	34,17	37,57	45,32
25	11,52	14,61	16,47	34,38	37,65	40,65	44,31	52,62
30	14,95	18,49	20,60	40,26	43,77	46,98	50,89	59,70

Tabla 6
Distribución t de Student: $P(|T| > t_0) = 2 \int_{t_0}^{\infty} s_n(x) dx$
Valores de t_0

$n \backslash p$	0,90	0,70	0,50	0,20	0,10	0,05	0,02	0,01
1	0,158	0,510	1,000	3,078	6,314	12,706	31,821	63,657
2	0,142	0,445	0,816	1,886	2,920	4,303	6,965	9,925
3	0,137	0,424	0,765	1,638	2,353	3,182	4,541	5,841
4	0,134	0,414	0,741	1,533	2,132	2,776	3,747	4,604
5	0,132	0,408	0,727	1,476	2,015	2,571	3,365	4,032
6	0,131	0,404	0,718	1,440	1,943	2,447	3,143	3,707
7	0,130	0,402	0,716	1,415	1,895	2,365	2,998	3,499
8	0,130	0,399	0,706	1,397	1,860	2,306	2,896	3,355
9	0,129	0,398	0,702	1,383	1,833	2,262	2,821	3,250
10	0,129	0,397	0,700	1,372	1,812	2,228	2,764	3,169
11	0,129	0,396	0,697	1,363	1,796	2,201	2,718	3,106
12	0,128	0,395	0,695	1,356	1,782	2,179	2,681	3,055
13	0,128	0,394	0,694	1,350	1,771	2,160	2,650	3,012
14	0,128	0,393	0,692	1,345	1,761	2,145	2,624	2,977
15	0,128	0,393	0,691	1,341	1,753	2,131	2,602	2,947
16	0,128	0,392	0,690	1,337	1,746	2,120	2,583	2,921
17	0,128	0,392	0,689	1,333	1,740	2,110	2,567	2,898
18	0,127	0,392	0,688	1,330	1,734	2,101	2,552	2,878
19	0,127	0,391	0,688	1,328	1,729	2,093	2,539	2,861
20	0,127	0,391	0,687	1,325	1,725	2,086	2,528	2,845
25	0,127	0,390	0,684	1,316	1,708	2,060	2,485	2,787
30	0,127	0,389	0,683	1,310	1,697	2,042	2,457	2,750

BIBLIOGRAFIA

La bibliografía sobre probabilidades y estadística es cuantiosa, y por ese motivo nos limitaremos a señalar algunos textos fundamentales y otros que, por estar en idioma castellano y ser de carácter elemental, pueden ser útiles como textos de iniciación.

Textos Fundamentales

- (1) FELLER, W. An Introduction to Probability Theory and its Applications, Wiley, Nueva York, Vol. 1 (1950).
- (2) RENY, A. Wahrscheinlichkeitsrechnung mit einem Anhang über Informationstheorie, Deutscher Verlag der Wissenschaften, Berlín (1966).
- (3) RIOS, S. Métodos Estadísticos, McGraw-Hill, Nueva York (1967).

131

Textos Intermedios

- (4) CRAMER, H. Elementos de la Teoría de Probabilidades y Aplicaciones, Aguilar, Madrid (1963).
- (5) FREEMAN, H. Introduction to Statistical Inference, Addison-Wesley, Reading, Mass. (1963).
- (6) LINDLEY, D. V. Introduction to Probability and Statistics, from a Bayesian Viewpoint; Part I. Probability, Part II. Inference, Cambridge University Press, Londres (1965).
- (7) TORANZOS, F. I. Estadística, Kapelusz, Buenos Aires (1962).

Textos Elementales

- (8) MASTROGIOVANI, M. de. Estadística y Probabilidad para Educadores, Estrada, Buenos Aires (1969).
- (9) NUÑEZ, J. A. y LENZI, R. V. Introducción al Cálculo de Probabilidades, Centro Argentino de Profesores de Enseñanza Media, Buenos Aires (1968).

COLECCION DE MONOGRAFIAS CIENTIFICAS

Publicadas

Serie de matemática

- N° 1. La Revolución en las Matemáticas Escolares, por el Consejo Nacional de Maestros de Matemáticas de los Estados Unidos de América.
- N° 2. Espacios Vectoriales y Geometría Analítica, por Luis A. Santaló.
- N° 3. Estructuras Algebraicas I, por Enzo R. Gentile.
- N° 4. Historia de las Ideas Modernas en la Matemática, por José Babini.
- N° 5. Algebra Lineal, por Orlando Villamayor.
- N° 6. Algebra Linear e Geometria Euclidiana, por Alexandre Augusto Martins Rodrigues.
- N° 7. El Concepto de Número, por César A. Trejo.
- N° 8. Funciones de Variable Compleja, por José I. Nieto.
- N° 9. Introducción a la Topología General, por Juan Horváth.
- N° 10. Funções Reais, por Djairo G. de Figueiredo.
- N° 11. Probabilidad e Inferencia Estadística, por Luis A. Santaló.
- N° 12. Estructuras Algebraicas II (Algebra Lineal), por Enzo R. Gentile.
- N° 13. La Revolución en las Matemáticas Escolares (Segunda Fase), por Howard F. Fehr, John Camp y Howard Kellogg.
- N° 14. Estructuras Algebraicas III (Grupos Finitos), por Horacio O'Brien.

133

Serie de física

- N° 1. Concepto Moderno del Núcleo, por D. Allan Bromley.
- N° 2. Panorama de la Astronomía Moderna, por Félix Cernuschi y Sayd Codina.
- N° 3. La Estructura Electrónica de los Sólidos, por Leopoldo M. Falicov.
- N° 4. Física de Partículas, por Igor Saavedra.
- N° 5. Experimento, Razonamiento y Creación en Física, por Félix Cernuschi.
- N° 6. Semiconductores, por George Bemski.
- N° 7. Aceleradores de Partículas, por Fernando Alba Andrade.
- N° 8. Física Cuántica, por Onofre Rojo y H. McIntosh.
- N° 9. La Radiación Cósmica, por Gastón R. Mejía y Carlos Aguirre.

N° 10. Astrofísica, por Carlos Jaschek y Mercedes C. de Jaschek.

N° 11. Ondas, por Oscar J. Bressan y Enrique Gaviola.

Serie de química

N° 1. Cinética Química Elemental, por Harold Behrens Le Bas.

N° 2. Bioenergética, por Isaias Raw y Walter Colli.

N° 3. Macromoléculas, por Alejandro Paladini y M. Burachik.

N° 4. Mecanismos de las Reacciones Orgánicas, por Jorge A. Brioux.

N° 5. Elementos Encadenados, por Jacobo Gómez-Lara.

N° 6. Enseñanza de la Química Experimental, por Francisco Giral.

N° 7. Fotoquímica de Gases, por Ralf-Dieter Penzhorn.

N° 8. Introducción a la Geoquímica, por Félix González-Bonorino.

N° 9. Resonancia Magnética Nuclear de Hidrógeno, por Pedro Joseph-Nathan.

N° 10. Cromatografía Líquida de Alta Presión, por Harold M. McNair y Benjamín Esquivel H.

N° 11. Actividad Óptica, Dispersión Rotatoria Óptica y Dicroísmo Circular en Química Orgánica, por Pierre Crabbé.

N° 12. Espectroscopia Infrarroja, por Jesús Morcillo Rubio.

N° 13. Polarografía, por Alejandro J. Arviá y Jorge A. Bolzan.

134

Serie de biología

N° 1. La Genética y la Revolución en las Ciencias Biológicas, por José Luis Reissig.

N° 2. Bases Ecológicas de la Explotación Agropecuaria en la América Latina, por Guillermo Mann F.

N° 3. La Taxonomía y la Revolución en las Ciencias Biológicas, por Elías R. de la Sota.

N° 4. Principios Básicos para la Enseñanza de la Biología, por Oswaldo Frota-Pessoa.

N° 5. A Vida da Célula, por Renato Basile.

N° 6. Microorganismos, por J. M. Gutiérrez-Vázquez.

N° 7. Principios Generales de Microbiología, por Norberto J. Palleroni.

N° 8. Los Virus, por Enriqueta Pizarro-Suárez y Gamba.

N° 9. Introducción a la Ecología del Bentos Marino, por Manuel Vegas Vélez.

N° 10. Biosíntesis de Proteínas y el Código Genético, por Jorge E. Allende.

- N° 11. Fundamentos de Inmunología e Inmunoquímica, por Félix Córdoba Alva y Sergio Estrada-Parra.
- N° 12. Bacteriófagos, por Romilio Espejo T.
- N° 13. Biogeografía de América Latina, por Angel L. Cabrera y Abraham Willink.
- N° 14. Relación Huésped-Parásito. Mecanismo de Patogenicidad de los Microorganismos, por Manuel Rodríguez Leiva.

En preparación

Serie de matemática

- Estructuras Algebraicas IV (Algebra Multilineal), por Orlando Villamayor y Artibano Micali.
- Estructuras Algebraicas V (Teoría de Cuerpos), por Darío J. Picco.
- Estructuras Algebraicas VI (Estructura de Algebras), por Artibano Micali.
- Introducción al Análisis, por Manuel Balanzat.
- Introducción a la Integral de Lebesgue en la Recta, por Juan Antonio Gatica.
- Introdução à Análise Funcional: Espaços de Banach e Cálculo Diferencial, por Leopoldo Nachbin.
- Introducción a los Espacios de Hilbert, por Jose I. Nieto.
- Introdução às Equações Diferenciais Ordinárias, por Chaim S. Hönl.
- Introducción a la Computación, por Jaime Michelow.
- Teoría General de la Optimización, por Enrique Cansado.
- Teoría de Grafos, por Fausto Alfredo Toranzos.
- Programación Lineal, por Fernando L. Garagorry.

135

Serie de física

- El Láser, por Mario Garavaglia.
- Cálculo de Errores, Aplicación y Teoría, por Wolfgang Meckbach.
- Oceanografía Física, por Luis E. Herrera.

Serie de química

- Paramagnetismo Electrónico, por Juan A. McMillan.
- Productos Naturales Vegetales, por Venancio Deulofeu y colaboradores.
- Esteroquímica Orgánica, por Juan A. Garbarino.
- Fotometría de Llama por Emisión, por Juan Ramírez Muñoz.
- Fotometría de Llama por Absorción Atómica, por Juan Ramírez Muñoz.

Cromatografía en Papel y en Capa Delgada, por Jorge A. Domínguez.
Momento Polar, por Pedro Lehman.
Fluorescencia Atómica, por Juan Ramírez Muñoz.
Introdução à Espectrometria de Massa das Substâncias Orgânicas, por Otto R. Gottlieb.
Los Esteroides, por Josef E. Herz.
Terموquímica Moderna, por Jaime Cases.
Cromatografía de Gases, por Harold M. McNair.

Serie de biología

Micología, por Margarita Silva-Hutner, William C. Merz y Luiz R. Travassos.
Procesos Microbianos Aerobios de Importancia Industrial, por Carlos Casas-Campillo.
Microbiología de Suelos, por Luis Longeri Spada.
Ecología Fisiológica, por Ernesto Medina.
Etología: El Estudio del Comportamiento Animal, por Raúl Vaz-Ferreira.
Citogenética Básica y Biología de los Cromosomas, por F. A. Saez y H. Cardoso.
Citogenética Ultraestructural y la Biología Molecular de los Cromosomas, por R. Wettstein y J. Roberto Sotelo.
Análisis de Sistemas en Ecología, por Gilberto C. Gallopiñ.
Ecología de Poblaciones, por Jorge E. Rabinovich.
Sistemas Ecológicos y el Hombre, por Ariel E. Lugo.
Inventario de Vegetación de Biomas, por Jorge Morello.

Nota: Las personas interesadas en adquirir estas obras deben dirigirse a la Unidad de Ventas y Circulación, Organización de los Estados Americanos, Washington, D. C., 20006 o a las Oficinas de la Secretaría General de la OEA en el país respectivo.