

Master en Estadística Aplicada y Estadística para el Sector Público

CIFF

Teoría de la Probabilidad

Primer Tomo

Rosa M. Ramos
Domínguez



Universidad
de Alcalá



Instituto
Nacional de
Estadística



CIFF

Centro Internacional
de Formación Financiera

2008
2009

**Máster Universitario en
“Estadística Aplicada y Estadística
para el Sector Público”
6^a edición**

Teoría de la Probabilidad
Rosa M. Ramos Domínguez

Índice general

1. Fenómenos aleatorios. Sucesos y Probabilidades	1
1.1. Fenómenos aleatorios.	1
1.2. Espacios de Probabilidad	3
1.2.1. Espacio muestral	3
1.2.2. Tipos de espacios muestrales asociados a un experimento aleatorio	4
1.2.3. Sucesos	5
1.2.4. σ -álgebras. σ -álgebra de Borel en \mathbb{R}	10
1.2.5. Frecuencia de un suceso. Propiedades	13
1.3. Axiomas	14
1.3.1. Probabilidad	14
1.4. Propiedades	15
1.4.1. Primeras propiedades de la probabilidad	15
1.5. Construcción de espacios de probabilidad. Caso Discreto	18
1.6. Construcción de espacios de probabilidad. Caso Continuo	20
2. Probabilidad Condicionada	23
2.1. Introducción	23
2.2. Probabilidad Condicionada.	23
2.3. Teoremas de la Probabilidad Condicionada.	26
2.4. Independencia de sucesos.	26
2.5. Teorema de la Probabilidad Total.	29
2.6. Teorema de Bayes.	31
3. Variables aleatorias unidimensionales	34
3.1. Introducción.	34
3.2. Variable aleatoria unidimensional.	34

3.2.1.	Aplicación medible.	35
3.2.2.	Operaciones con variables aleatorias.	37
3.3.	Probabilidad inducida por una variable aleatoria.	39
3.4.	Función de Distribución.	40
3.5.	Distribuciones Discretas y absolutamente continuas.	42
3.5.1.	Función de masa o cuantía.	42
3.5.2.	Variable aleatoria discreta.	43
3.5.3.	Distribuciones absolutamente continuas. Función de densidad sobre \mathbb{R}	46
3.6.	Cambio de variable en las distribuciones unidimensionales.	51
3.6.1.	Cambios de variable discreta.	52
3.6.2.	Cambios de variable absolutamente continuos.	53
4.	Distribuciones Unidimensionales	58
4.1.	Distribuciones Unidimensionales.	58
4.1.1.	Introducción	58
4.2.	Esperanza Matemática.	59
4.3.	Propiedades de la Esperanza.	65
4.4.	Desigualdad de Tchebychev.	71
4.5.	Momentos de una distribución.	71
4.5.1.	Varianza	74
4.5.2.	Propiedades de la Varianza	74
4.6.	Indicadores de posición y dispersión.	76
4.6.1.	Características de las distribuciones unidimensionales	76
4.6.2.	Indicadores de posición	77
4.6.3.	Medidas de forma	82
4.6.4.	Coeficiente de asimetría	83
4.6.5.	Coeficiente de curtosis o apuntamiento	83
5.	Función Generatriz	85
5.1.	Función generatriz de probabilidad.	85
5.1.1.	Función generatriz de momentos	87
5.2.	Introducción a la función característica.	91
5.2.1.	Variables aleatorias complejas.	91
5.2.2.	Función característica.	92

5.2.3. Propiedades.	93
5.3. Teoremas.	94
5.3.1. La función característica y el problema de los momentos.	96

Capítulo 1

Fenómenos aleatorios. Sucesos y Probabilidades

1.1. Fenómenos aleatorios.

El propósito de cualquier ciencia es enunciar las *leyes* que rigen los fenómenos observados que, por su naturaleza, forman parte de su campo de estudio. En un buen número de casos, son leyes de carácter *necesario* que predicen sin ambigüedad las consecuencias que se presentan, *necesariamente*, cada vez que se observa el fenómeno.

Frente a los fenómenos de esta clase, se presentan otros en los que la concurrencia de unas circunstancias fijas no permite prever cual será el efecto producido. Ejemplos de esta categoría son los siguientes:

1. Si se colocan diversas bolas numeradas, con las mismas propiedades físicas, en una bolsa y se extrae una bola a ciegas, necesariamente la bola llevará uno de los números marcados, pero es imposible predecir cual será.
2. Si se realiza todos los días, en la misma ciudad y a la misma hora, el mismo trayecto entre dos puntos alejados, nunca se tardará exactamente lo mismo. La duración está gobernada por demasiados “imponderables” para que su valor esté determinado por un conjunto cuantificable de causas.
3. Si una moneda cae al suelo de una habitación, no se puede prever el punto al que irá a parar. Aunque se lanzasen diversas monedas idénticas, poniendo cuidado en hacerlo de idéntica manera en todas las ocasiones, no acabarían todas en el mismo punto.

Tales experimentos se denominan *aleatorios*; en ellos, las consecuencias no están predeterminadas por sus causas y el resultado se atribuye al azar.

En los experimentos aleatorios, el carácter imprevisible de las consecuencias del azar hace inútil cualquier intento de hallar reglas determinísticas que rijan la aparición de los resultados individuales. Sin embargo, el azar también está sometido a *leyes*. Lo que ocurre es que tales leyes no determinan el resultado de cada experiencia, sino que sólo atañen a la frecuencia relativa de los resultados que se obtienen cuando el fenómeno se repite un gran número de veces. Así “normalmente”, una moneda no se aleja mucho de la posición en la que cae y, sólo “ocasionalmente”, rueda hasta una esquina de la habitación.

Por ejemplo, si se lanza mil veces una moneda equilibrada, lo único que se puede afirmar *con certeza* es que se obtendrán entre 0 y 1000 caras; afirmación banal carente totalmente de interés. Sin embargo, el cálculo de probabilidades nos permitirá afirmar que aproximadamente en el 95 % de los casos que hagamos este experimento, el número de caras estará comprendido entre 468 y 532. Despreciar esta información, aduciendo que no siempre es correcta, no parece muy inteligente; lo procedente es aprender a hacer este tipo de valoraciones de situaciones no determinísticas, precisar su sentido, y extraer de ellas normas de conducta para enfrentarse al azar. Este es el propósito del Cálculo de Probabilidades.

El objeto de estudio del Cálculo de Probabilidades lo constituye lo que en términos genéricos hemos llamado fenómenos aleatorios o experimentos aleatorios, que como se ve resulta difícil definir con precisión.

Por esta razón, en lugar de una definición, daremos una serie de propiedades que suelen caracterizarlos; éstas son:

1. En las mismas condiciones iniciales pueden dar lugar a diferentes resultados finales, en contraposición con los fenómenos determinísticos, los cuales conducen siempre a los mismos efectos finales.
2. Todos los resultados posibles se conocen por anticipado.
3. No se puede predecir el resultado en cada experiencia particular.
4. En general, pueden repetirse en las mismas condiciones indefinidamente.
5. Si uno de estos experimentos se repite bajo condiciones completamente idénticas un número determinado, n , de veces, y anotamos el número, n_i , de veces que aparece un cierto resultado, el cociente n_i/n tiende a estabilizarse, cuando n aumenta, en un valor fijo.

Por ello, ante fenómenos de azar, la tendencia natural es tratar de medir el grado de verosimilitud de los diversos acontecimientos posibles, asignando una probabilidad a cada uno de ellos, que informe de la frecuencia con que hay que esperar que se presente cada uno, después de numerosas observaciones del fenómeno.

De las consideraciones anteriores, puede concluirse: En un fenómeno aleatorio, la probabilidad de cada uno de los acontecimientos posibles es un número entre 0 y 1, que expresa la frecuencia ideal con la que dicho acontecimiento se presentaría en una serie indefinidamente larga, de repeticiones del experimento, realizadas en idénticas condiciones.

Esta interpretación frecuentista considera que la probabilidad es una cualidad objetiva de los fenómenos aleatorios. En una moneda se asemeja a una característica más del objeto: su masa, su diámetro, su probabilidad de salir cara, etc... Pero, igual que no tendría sentido hablar de masa si no existiese el hecho empírico de la atracción de la Tierra sobre la moneda, tampoco tendría sentido esta interpretación de la Probabilidad, si no se aceptase la siguiente ley empírica de regularidad de las frecuencias a la que se somete el azar: en una sucesión ilimitada de repeticiones, en idénticas condiciones, de un fenómeno aleatorio, las frecuencias de cada uno de los acontecimientos posibles, observadas tras cada repetición, tienden a aproximarse hacia ciertos valores límites. El valor límite de la frecuencia de un acontecimiento es lo que se desea expresar mediante su probabilidad.

1.2. Espacios de Probabilidad

1.2.1. Espacio muestral

Dado un experimento aleatorio, llamaremos espacio muestral Ω al conjunto de todos los resultados posibles distintos de dicho experimento aleatorio. Los elementos de Ω se denominan sucesos elementales.

Obsérvese que asociados a un mismo experimento aleatorio pueden considerarse diferentes espacios muestrales; así, en el ejemplo, antes considerado, del lanzamiento de un dado, un posible espacio muestral podría ser $\Omega_1 = \{par, impar\}$, otro $\Omega_2 = \{\text{menor que } 3, \text{ mayor o igual que } 3\}$, y otro $\Omega_3 = \{1, 2, 3, 4, 5, 6\}$. Todo depende de lo que queramos estudiar en nuestro experimento aleatorio y de la facilidad que tengamos en asignar la probabilidad a los sucesos elementales del espacio muestral que consideremos.

Lo primero que debemos hacer, por tanto, es decidir cuales son los posibles

resultados del experimento aleatorio. No es demasiado serio admitir más cosas de las que realmente puedan ocurrir, pero debemos estar seguros de no olvidar ninguna que pueda ocurrir.

Ejemplos de espacios muestrales hay muchos. Si consideramos el experimento aleatorio consistente en el lanzamiento de dos monedas al aire, el espacio muestral asociado será el $\Omega = \{(C, C), (C, F), (F, C), (F, F)\}$ en el caso de que se distinguiesen las dos monedas. Si consideramos el experimento aleatorio consistente en introducir dos bolas en dos celdas A y B , el espacio muestral correspondiente será el $\Omega_1 = \{AA, AB, BA, BB\}$ si las bolas son distinguibles; si las bolas fueran indistinguibles en cuanto a su orden de introducción, el espacio muestral sería el $\Omega_2 = \{AA, AB, BB\}$.

1.2.2. Tipos de espacios muestrales asociados a un experimento aleatorio

Consideremos los siguientes tipos de espacios muestrales:

1. *Espacio muestral finito*: Este tipo de espacio muestral se da cuando Ω está formado por un número finito de elementos. Son ejemplos de este espacio muestral, el asociado al experimento del lanzamiento de un dado, o el correspondiente al lanzamiento de una moneda.
2. *Espacio muestral infinito numerable*: Consideremos el experimento consistente en lanzar una moneda hasta obtener cara por primera vez. Veamos que *a priori* un experimento de tal tipo puede dar lugar a un conjunto infinito de eventualidades que podemos escribir en la forma

$$C, FC, FFC, FFFC, \dots, F \dots FC, \dots$$

En tal situación, una formalización razonable deberá ser aquella en la que el espacio muestral Ω contenga como elementos todas aquellas sucesiones finitas de la forma $FF \dots FC$. Sin embargo, también debería poseer este conjunto un posible resultado que fuese “la cara no aparece jamás”, ya que por una parte *a priori* no debería rechazarse tal resultado, aunque por el contrario la realización de infinitas tiradas sea materialmente imposible. Estamos aquí pues, en una situación común a todas las que quieren representar formalmente un fenómeno real: toda formalización implica una elección y no es posible en la mayoría de los casos justificar una elección más que otra, solamente la eficacia y la coherencia de la teoría construida permite concluir en la validez de tal o cual esquema.

Definir el espacio muestral Ω , nos lleva a tener a priori un conjunto de posibles resultados que parecen razonables y conducen a resultados de acuerdo con la experiencia. Así, en el ejemplo anterior debemos considerar como espacio muestral el construido por todas las sucesiones finitas o infinitas de la forma $FF \dots FC$, es decir, un conjunto de elementos que es infinito numerable.

3. *Espacio muestral continuo*: Supongamos el experimento aleatorio consistente en el lanzamiento de una partícula en un plano y supongamos también, que estamos interesados en la posición que dicha partícula ocupa en el citado plano. En esta situación Ω es todo el plano y por lo tanto un conjunto continuo.

Si en el mismo ejemplo anterior, en vez de considerar la posición de la partícula consideramos la trayectoria de ésta, nuevamente estaríamos ante el caso de un espacio muestral continuo.

Si consideramos el nivel de agua de un pantano entre los tiempos t_1 y t_2 , Ω sería el conjunto de todas las funciones no negativas $\Omega = \{w_t, t \in [t_1, t_2]\}$.

En cualquiera de los casos, Ω no es más que un conjunto abstracto de puntos, y a causa de esta interpretación, el lenguaje y los conceptos de la teoría de conjuntos suministran un contexto natural para el desarrollo de la Teoría de la Probabilidad.

1.2.3. Sucesos

En muchas ocasiones, lo que interesa en la realización de un fenómeno aleatorio no es tanto el resultado individual que se ha producido, sino el hecho de que dicho resultado tenga o no, una determinada característica. Por ejemplo, al extraer una carta de un baraja, puede que haya un gran interés de que sea un as, o de que no sea un basto, sin que importe mucho la carta en concreto que se obtenga. Así, con frecuencia, los subconjuntos que expresan las características relevantes del resultado, juegan un papel más importante que los elementos del espacio muestral.

Los subconjuntos del espacio muestral Ω se denominan sucesos. Los sucesos simbolizan los acontecimientos que se pueden observar como resultado de un experimento de azar. Los subconjuntos con un único elemento representan uno sólo de los resultados posibles del fenómeno y se denominan sucesos simples; los subconjuntos con más de un elemento se denominan sucesos compuestos y son uniones de sucesos simples.

Ejemplo 1:

En el experimento que consiste en lanzar un dado y observar su cara superior, el espacio muestral adecuado es $\Omega = \{1, 2, 3, 4, 5, 6\}$; los sucesos simples son $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$ y $\{6\}$.

Ejemplos de sucesos compuestos son obtener resultado par, que coincide con el subconjunto $\{2, 4, 6\}$; obtener resultado inferior a 3, que equivale al subconjunto $\{1, 2\}$; obtener resultado primo: $\{2, 3, 5\}$, etc.

Nótese que cada suceso tiene dos descripciones: una verbal, que enuncia una propiedad característica de sus elementos, y otra extensiva que enumera todos los resultados que lo componen. La descripción verbal suele ser más sugerente de su significado, aunque hay que asegurarse de que expresa sin ambigüedad el subconjunto de resultados que define.

En particular, el propio espacio muestral Ω es un suceso compuesto, que contiene como elementos a todos los resultados posibles del fenómeno y recibe por ello el nombre de suceso seguro, mientras que el subconjunto vacío, \emptyset , se considera el suceso que no ocurre nunca, ni simple ni compuesto, denominado suceso imposible.

Naturalmente, identificar los sucesos relativos a un cierto fenómeno de azar con los subconjuntos de su espacio muestral, Ω , permite disponer de operaciones para formar nuevos sucesos a partir de otros dados; así:

1. Si A y B son sucesos, su unión: $A \cup B$, representa otro suceso que puede escribirse como “ocurre alguno de los sucesos A o B ”, que sucede siempre que el resultado pertenezca a A , a B , o a ambos simultáneamente.
2. La intersección, $A \cap B$, es también un nuevo suceso: “ocurren A y B simultáneamente”, que se presenta siempre que el resultado pertenezca a A y a B .
Si la intersección de dos sucesos A y B es vacía, $A \cap B = \emptyset$, ambos sucesos no pueden ocurrir simultáneamente; se dice entonces que A y B son incompatibles.
3. El complementario, $A^c = \Omega - A$, de cualquier suceso A , es el suceso contrario de A y significa que el resultado del fenómeno no pertenece al suceso A .
4. Dos sucesos A y B pueden estar relacionados de modo que siempre que ocurre A , ocurre B ; lo que equivale a que se verifique la relación: $A \subset B$; esto es, cualquier elemento de A pertenece también a B .

Observemos que dado cualquier suceso A , siempre ocurre

$$\emptyset \subset A \subset \Omega$$

Diferencia de sucesos. Dados dos sucesos A y B de un cierto experimento aleatorio, se define la diferencia de los sucesos A y B , y se representa por $A - B$, el suceso que denotamos por G , que ocurra A y no ocurra B

$$A - B = G$$

En definitiva, la familia de los sucesos que pueden considerarse en relación con un fenómeno aleatorio, de espacio muestral finito Ω , tiene la misma estructura que la familia, $\mathcal{P}(\Omega)$, de los subconjuntos de Ω . Cualquier relación u operación que se realice entre subconjuntos de Ω tiene su interpretación en términos de sucesos.

Si se considera un espacio muestral finito, el conjunto de sucesos, $\mathcal{P}(\Omega)$, será también finito; concretamente, si Ω tiene n elementos, habrá 2^n sucesos.

En ocasiones, el espacio muestral, Ω , que describe los posibles resultados de un fenómeno aleatorio es un conjunto numerable. Esta situación se produce, por ejemplo, en los casos siguientes:

Ejemplo 2:

Una moneda es lanzada hasta que aparece la primera cara, de manera que el resultado del experimento queda caracterizado por el número total de lanzamientos realizados.

“Normalmente” la primera cara aparecerá al cabo de unos pocos lanzamientos y el resultado obtenido será un número relativamente pequeño. Sin embargo, en teoría, no se puede excluir la posibilidad de que la primera cara tarde en aparecer más que cualquier número prefijado de lanzamientos. Negarlo, sería admitir que hay cierto número n_0 de lanzamientos tal que, si ha salido cruz en los n_0 primeros, es forzoso que aparezca cara en el lanzamiento $n_0 + 1$.

Así pues, los resultados posibles del fenómeno son tantos como números naturales. Es decir, $\Omega = \mathcal{N}$.

En el ejemplo si w describe el número de lanzamientos de una moneda hasta obtener la primera cara, para cada $w \in \mathcal{N}$, $\{w\}$ es un suceso simple. En cambio, son sucesos compuestos:

La primera cara aparece antes del décimo lanzamiento: $\{w < 10\}$.

La primera cara aparece en un lanzamiento par: $\{w = 2\}$.

En espacios muestrales numerables, no es necesario modificar la noción de suceso: Cualquier subconjunto de Ω es un suceso; simple o compuesto, según que contenga uno o más de los resultados posibles.

En casos como los anteriores, en que Ω es infinito numerable, el conjunto de sucesos, $\mathcal{P}(\Omega)$, no es numerable, sino que su cardinal es aún mayor (2^{\aleph} exactamente). A diferencia con el caso de espacios muestrales finitos, es posible entonces considerar sucesiones de sucesos

$$A_1, A_2, A_3, \dots, A_n, \dots$$

mediante las cuales pueden formarse nuevos sucesos. En primer lugar

$$\bigcap_{n=1}^{\infty} A_n \text{ y } \bigcup_{n=1}^{\infty} A_n$$

que describen, respectivamente, la ocurrencia de todos o alguno de los sucesos A_n . Pero también se puede considerar

$$\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n$$

que expresa que ocurren todos los A_n , a partir de algún m o, dicho más simplemente todos los A_n excepto un número finito de ellos. Y, así mismo,

$$\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$$

formado por aquellos resultados w que, para cualquier m dado, pertenecen a un A_n con $n \geq m$; es decir, los que implican la ocurrencia de un número infinito de los sucesos A_n .

Una vez establecido el concepto matemático de suceso asociado a un fenómeno aleatorio con espacio muestral discreto: finito o numerable, Ω , el propósito del Cálculo de Probabilidades es atribuir una probabilidad a cualquier subconjunto de $\mathcal{P}(\Omega)$.

Mientras que Ω es finito o numerable, no hay ninguna dificultad teórica en asignar probabilidad a todos y cada uno de los subconjuntos del espacio muestral.

La aparición de los modelos continuos supone un cambio considerable en tal estado de cosas.

Por un lado aunque pueda resultar sorprendente, no es posible que un modelo continuo asigne probabilidad a todos los subconjuntos de \mathbb{R} ; es decir, es imposible considerar la colección $\mathcal{P}(\mathbb{R})$ de todos los subconjuntos de \mathbb{R} , como la familia de sucesos. Pues, si se intenta definir una probabilidad

$$P : \mathcal{P}(\mathbb{R}) \longrightarrow [0,1]$$

que no asigne probabilidad 1 a un subconjunto numerable de \mathbb{R} , se producen incompatibilidades. La razón es que existen subconjuntos de \mathbb{R} de una irregularidad tal que no se les puede atribuir probabilidad que resulte coherente con la del resto de los conjuntos.

No hay pues otro remedio que restringir la clase de sucesos que pueden considerarse. Esto es, seleccionar una subfamilia de $\mathcal{P}(\mathbb{R})$ ($\mathcal{F} \subset \mathcal{P}(\Omega)$), compuesta por subconjuntos de \mathbb{R} que sean útiles a efectos prácticos, de la que se excluyan los conjuntos

anómalos. Sin duda, lo más lógico es considerar como sucesos los intervalos de \mathbb{R} y aquellos conjuntos que puedan formarse (mediante operaciones conjuntistas numerables) a partir de intervalos.

Casos como los anteriores ponen de relieve que, en general, con un espacio muestral arbitrario, Ω , la familia $\mathcal{P}(\Omega)$ de todos los subconjuntos de Ω puede no ser el ámbito adecuado para definir la probabilidad. Con frecuencia, será necesario restringir su dominio a una familia de subconjuntos de Ω , (\mathcal{F}) que ofrezca garantías de coherencia. Dicho de otro modo, es obligatorio que todo modelo probabilístico, antes de especificar la probabilidad, incluya un dato previo: la familia, \mathcal{F} , de subconjuntos de Ω , sobre la que estará definida la probabilidad.

Esto supone que se renuncia a hablar de la probabilidad de cualquier subconjunto del espacio muestral que no pertenezca a \mathcal{F} . Y, puestas así las cosas, parece razonable que \mathcal{F} cumpla determinados requisitos. Concretamente:

1. Si $A \in \mathcal{F}$, también debe estar en \mathcal{F} el subconjunto $A^c = \Omega - A$.

En caso contrario, podría considerarse la probabilidad de que A se realice, pero carecería de sentido la probabilidad de que no ocurra A .

2. Si $\{A_n\}$ es una colección finita o numerable de conjuntos de \mathcal{F} , su unión $\bigcup_n A_n$ debe también pertenecer a \mathcal{F} .

Ello supone que, si se asigna probabilidad a cada A_n , también hay que otorgar probabilidad al hecho de que se realice alguno de ellos. Mientras se trate de un número finito de conjuntos tal exigencia parece incontestable. La extensión al caso de una cantidad numerable de conjuntos es una necesidad matemática importante, ligada con las propiedades de la probabilidad. En cambio, no procede ir más allá e incluir en \mathcal{F} las uniones arbitrarias de sus elementos ya que, en la mayor parte de las ocasiones, \mathcal{F} contiene a los conjuntos constituidos por un único punto de Ω , en cuyo caso cualquier subconjunto de Ω es unión de conjuntos de \mathcal{F} .

En Cálculo de Probabilidades, la exigencia habitual que haremos a la clase \mathcal{F} es que sea **σ -álgebra**.

A estudiar las características de esta estructura y de otras más sencillas dedicaremos las dos siguientes secciones que estudiaremos desde un punto de vista abstracto, considerando al espacio muestral Ω como un conjunto de puntos fijos al que se denomina espacio total.

Definición 1.1. Si Ω es un conjunto cualquiera y \mathcal{F} una σ -álgebra de subconjuntos de Ω , se dice que (Ω, \mathcal{F}) es un espacio medible o probabilizable. Los subconjuntos que

pertenecen a \mathcal{F} se denominan sucesos y, por tanto, \mathcal{F} recibe el nombre de σ -álgebra de sucesos.

La introducción de la σ -álgebra de sucesos no afecta en absoluto a los modelos probabilísticos discretos. Cuando Ω es discreto (finito o numerable), la norma es tomar $\mathcal{F} = \mathcal{P}(\Omega)$ y atribuir probabilidad a todos los subconjuntos de Ω . No hay ninguna razón para hacer una elección distinta.

En cambio, en los modelos probabilísticos continuos, cuyo espacio muestral, Ω , es un subconjunto no numerable de un espacio euclídeo \mathbb{R}^k , siempre se elige como σ -álgebra de sucesos la σ -álgebra de Borel en Ω ; es decir la restricción de \mathbb{B}^k al conjunto Ω .

1.2.4. σ -álgebras. σ -álgebra de Borel en \mathbb{R}

El objetivo de esta sección es analizar la estructura que pueden tener algunas familias de subconjuntos de un conjunto y las propiedades correspondientes. El motivo es que, tanto en teoría de la medida como en cálculo de probabilidades, juega un papel importante la estructura de σ -álgebra que se define a continuación

Definición 1.2. Dado un conjunto arbitrario Ω , una familia \mathcal{F} de subconjuntos de Ω tal que

1. para cualquier $A \in \mathcal{F}$ es $A^c \in \mathcal{F}$,
2. si $\{A_n\}$ es una colección numerable de subconjuntos de Ω que verifican $A_n \in \mathcal{F}$ para cada n , se cumple $\bigcup_n A_n \in \mathcal{F}$,

recibe el nombre de σ -álgebra de subconjuntos de Ω .

La definición anterior tiene algunas consecuencias inmediatas

Proposición 1.1. Si \mathcal{F} es una σ -álgebra de subconjuntos de Ω , se cumple

1. $\Omega \in \mathcal{F}$ y $\emptyset \in \mathcal{F}$.
2. Si $\{A_n\}$ es una colección numerable de subconjuntos de Ω tal que $A_n \in \mathcal{F}$ para cada n , se cumple $\bigcap_n A_n \in \mathcal{F}$.

En efecto, dado cualquier $A \in \mathcal{F}$ es $A^c \in \mathcal{F}$ y, por tanto, $\Omega = A \cup A^c \in \mathcal{F}$. Luego, también, $\emptyset = \Omega^c \in \mathcal{F}$. Por otra parte, es bien conocida la ley de De Morgan:

$$\bigcap_n A_n = \left(\bigcup_n A_n^c \right)^c$$

En caso de que los conjuntos A_n pertenezcan a \mathcal{F} , lo mismo ocurre con sus complementarios, A_n^c , y con $\bigcup_n A_n^c$. En consecuencia, $\bigcap_n A_n \in \mathcal{F}$.

La definición y las propiedades anteriores muestran que las σ -álgebras son aquellas clases de subconjuntos de Ω , cerradas por cualquiera de las operaciones conjuntistas (unión, intersección y complementario) siempre que se realicen con un número numerable de términos; en el sentido de que cualquier operación llevada a cabo con conjuntos que pertenecen a \mathcal{F} , sigue siendo un elemento de \mathcal{F} .

Ejemplo 3:

Dentro de un conjunto cualquiera Ω , existen muchas σ -álgebras de un mayor o menor interés. La más grande es sin duda, la familia $\mathcal{P}(\Omega)$ de todos los subconjuntos de Ω que se denomina a veces la σ -álgebra discreta.

La más pequeña, es la colección $\{\emptyset, \Omega\}$, formada exclusivamente por el conjunto vacío y el conjunto total que suele denominarse la σ -álgebra trivial.

Para cada conjunto $A \subset \Omega$, existe una σ -álgebra mínima que lo contiene; es concretamente

$$\{\emptyset, A, A^c, \Omega\}$$

De forma similar, dado otro subconjunto B de Ω , que no sea disjunto con A ni con A^c , la menor σ -álgebra que contiene a A y a B es

$$\{\emptyset, A, B, A^c, B^c, A \cup B, A \cup B^c, A^c \cup B, A^c \cup B^c, A \cap B, A \cap B^c,$$

$$A^c \cap B, A^c \cap B^c, (A \cap B) \cup (A^c \cap B^c), (A \cap B^c) \cup (A^c \cap B), \Omega\}$$

(Los conjuntos A y B dividen el espacio en cuatro piezas: $A \cap B$, $A \cap B^c$, $A^c \cap B$ y $A^c \cap B^c$; entre los conjuntos de la σ -álgebra hay 6 formados por dos piezas: A , A^c , B , B^c , $(A \cap B) \cup (A^c \cap B^c)$ y $(A \cap B^c) \cup (A^c \cap B)$; más otros 4 formados por tres piezas: $A \cup B$, $A \cup B^c$, $A^c \cup B$ y $A^c \cup B^c$. En total 2^4 conjuntos.

Al añadir más conjuntos la complejidad de la σ -álgebra crece a un ritmo desorbitado; con n conjuntos cualesquiera puede constar hasta de $2^{(2^n)}$ elementos.

Puesto que hay múltiples σ -álgebras en cualquier conjunto, tiene interés la siguiente propiedad.

Proposición 1.2. Sea $\{\mathcal{F}_i\}_{i \in I}$ una colección arbitraria de σ -álgebras en un conjunto Ω , entonces

$$\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i$$

es una σ -álgebra de subconjuntos de Ω .

En efecto, si $A \in \mathcal{F}$, es $A \in \mathcal{F}_i$ para cada $i \in I$ y, por consiguiente, $A^c \in \mathcal{F}_i$ para cada $i \in I$. Luego, $A^c \in \mathcal{F}$.

Análogamente, si $\{A_n\} \subset \mathcal{F}$, será $\{A_n\} \subset \mathcal{F}_i$ para cada $i \in I$; o bien, puesto que cada \mathcal{F}_i es σ -álgebra, $\bigcup_n A_n \in \mathcal{F}_i$ para cada $i \in I$. Con lo cual $\bigcup_n A_n \in \mathcal{F}$.

El resultado anterior permite dar la siguiente definición

Definición 1.3. Dada una familia cualquiera, \mathcal{C} , de subconjuntos de un conjunto Ω , la intersección de todas las σ -álgebras que contienen a \mathcal{C} :

$$\sigma(\mathcal{C}) = \bigcap_{\mathcal{F} \supset \mathcal{C}} \mathcal{F}$$

es la mínima σ -álgebra que contiene a \mathcal{C} , ya que está contenida en cualquier σ -álgebra \mathcal{F} que contenga a \mathcal{C} . Se denomina a $\sigma(\mathcal{C})$ la σ -álgebra engendrada por \mathcal{C} .

En el ejemplo 3 se ha construido directamente la σ -álgebra engendrada por una familia de subconjuntos \mathcal{C} que esté formada por un único subconjunto, A , o por dos de ellos, A y B . Pero, en el caso $\Omega = \mathbb{R}$, la definición tiene sobre todo interés cuando la familia \mathcal{C} está formada por todos los intervalos.

Ejemplo 4:

Sea $\Omega = \{1, 2, 3, 4, 5\}$ y consideremos la familia $\mathcal{C} = \{\{1\}, \{3, 4\}\}$ construir la sigma álgebra engendrada por \mathcal{C} .

$$\begin{aligned} \mathcal{C}_1 &= \{\emptyset, \Omega, \{1\}, \{3, 4\}, \{2, 3, 4, 5\}, \{1, 2, 5\}\} \\ \mathcal{C}_2 &= \{\emptyset, \Omega, \{1\}, \{3, 4\}, \{2, 3, 4, 5\}, \{1, 2, 5\}, \{2, 5\}\} \\ \mathcal{C}_3 &= \{\emptyset, \Omega, \{1\}, \{3, 4\}, \{2, 3, 4, 5\}, \{1, 2, 5\}, \{2, 5\}, \{1, 3, 4\}\} = \sigma(\mathcal{C}) \end{aligned}$$

Definición 1.4. Sea $\mathcal{I} = \{(a, b], (-\infty, b], (a, +\infty) / a \leq b; a, b \in \mathbb{R}\}$ la familia constituida por todos los intervalos de \mathbb{R} (abiertos, cerrados o semiabiertos, finitos o infinitos). La σ -álgebra $\sigma(\mathcal{I})$, engendrada por \mathcal{I} , se denomina la σ -álgebra de Borel en \mathbb{R} y suele denotarse por \mathbb{B} . Los elementos de \mathbb{B} se denominan conjuntos de Borel.

No existe una caracterización explícita de los conjuntos de Borel. La propia definición indica que es de Borel cualquier conjunto que pueda formarse mediante una secuencia de operaciones conjuntistas a partir de un número numerable de intervalos. Así que lo difícil es imaginar un conjunto que no pertenezca a la σ -álgebra de Borel. Por ejemplo, se cumple:

Proposición 1.3. Cualquier conjunto abierto de \mathbb{R} es un conjunto de Borel. Lo mismo sucede con cualquier conjunto cerrado, con cualquier intersección numerable de abiertos, con cualquier unión numerable de cerrados, etc.

En efecto, es conocido que cualquier abierto A de \mathbb{R} se puede expresar como unión numerable de intervalos abiertos disjuntos. Por consiguiente, cualquier abierto es un conjunto de Borel.

Así mismo, si A es cerrado, A^c es un abierto y, por consiguiente, de Borel. Lo mismo sucede entonces con su complementario, A .

□

En ocasiones, se plantean problemas relativos a la longitud que no involucran a toda la recta real sino a un subconjunto fijo, A , que será normalmente un intervalo o un conjunto de Borel. En tal caso, los elementos de \mathbb{B} disjuntos con A son inútiles y por ello se considera la “parte” de la σ -álgebra contenida en A . En general, la idea responde a la siguiente definición.

Definición 1.5. Sea \mathcal{F} una σ -álgebra de subconjuntos de Ω y A un subconjunto cualquiera de Ω , la familia

$$\mathcal{F}_A = \{B \subset A \mid B = A \cap C \text{ para algún } C \in \mathcal{F}\}$$

recibe el nombre de σ -álgebra restringida de \mathcal{F} a A . Si $\Omega = \mathbb{R}$ y \mathcal{F} es la σ -álgebra de Borel, \mathbb{B} , se denomina \mathbb{B}_A la σ -álgebra de Borel en A .

La comprobación de que \mathcal{F}_A es una σ -álgebra de subconjuntos de A es muy simple; si $B \in \mathcal{F}_A$ es $B = A \cap C$ con $C \in \mathcal{F}$, luego

$$B^c = A - B = A - A \cap C = A \cap (\Omega - C)$$

que pertenece a \mathcal{F}_A por ser $\Omega - C \in \mathcal{F}$. Análogamente, si para cada n es $B_n = A \cap C_n$ con $C_n \in \mathcal{F}$, será

$$\cup_n B_n = A \cap (\cup_n C_n)$$

que pertenece a \mathcal{F}_A por ser $\cup_n C_n \in \mathcal{F}$.

□

1.2.5. Frecuencia de un suceso. Propiedades

Nuestro objetivo es definir una función que nos indique una medida de la certeza o incertidumbre en la ocurrencia de los sucesos del experimento aleatorio. Para ello vamos a considerar las propiedades de las frecuencias relativas de un suceso en la realización de un experimento aleatorio.

Dado un suceso A, es decir, un suceso perteneciente al conjunto de sucesos del experimento aleatorio, la frecuencia absoluta del suceso A es el número de veces que ocurre el suceso A en una serie de n repeticiones similares del experimento; se la representa por n_A .

La frecuencia relativa del suceso A es la frecuencia absoluta dividida por el número de veces que se realiza el experimento; se la representa por f_A :

$$f_A = \frac{n_A}{n}$$

- 1) Observamos que para cualquier suceso A, resulta:

$$0 \leq f_A \leq 1$$

ya que será siempre cociente de dos números positivos, en el que el numerador es siempre menor o igual que el denominador.

- 2) Para el suceso seguro Ω , $f_\Omega = 1$ ya que Ω ocurre siempre

- 3) Para el suceso imposible, $f_\emptyset = 0$ ya que el \emptyset no ocurre nunca.

Si dos sucesos A y B son incompatibles, tendremos: sea n_A la frecuencia absoluta de A, n_B la frecuencia absoluta de B y sea $n_{A \cup B}$ la frecuencia absoluta del suceso $A \cup B$, por ser incompatibles

$$n_{A \cup B} = n_A + n_B$$

tendremos pues

$$f_{A \cup B} = \frac{n_{A \cup B}}{n} = \frac{n_A + n_B}{n} = \frac{n_A}{n} + \frac{n_B}{n} = f_A + f_B$$

es decir, la frecuencia relativa de la unión de sucesos incompatibles es la suma de las frecuencias relativas de cada uno de los sucesos.

- 4) $f_{A \cup B} = f_A + f_B$ si $A \cap B = \emptyset$

1.3. Axiomas

1.3.1. Probabilidad

El tercero de los elementos a considerar en relación con un experimento aleatorio es, como dijimos, la probabilidad de los sucesos de la σ -álgebra de subconjuntos del espacio muestral Ω .

Existen varias interpretaciones del concepto de probabilidad; aquí partiremos de unos axiomas como propiedades a cumplir por toda función probabilidad.

Axiomática de Kolmogorov

Definición 1.6. En un espacio medible (Ω, \mathcal{F}) , una probabilidad (o medida de probabilidad) es una aplicación $P : \mathcal{F} \rightarrow \mathbb{R}$ que verifica

(a) $P(A) \geq 0$ para todo $A \in \mathcal{F}$

(b) Para cualquier colección numerable de sucesos $\{A_n\} \subset \mathcal{F}$, disjuntos entre sí, se cumple

$$P\left(\bigcup_n A_n\right) = \sum_n P(A_n)$$

(c) $P(\Omega) = 1$

Si P es una probabilidad en (Ω, \mathcal{F}) , se denomina espacio de probabilidad a la terna (Ω, \mathcal{F}, P) y el valor $P(A)$ asociado con cada suceso $A \in \mathcal{F}$ recibe el nombre de probabilidad de A .

Observemos que estos axiomas son producto de una motivación puramente realista al considerar la probabilidad de un suceso como una abstracción de la frecuencia relativa del mismo en una larga serie de realizaciones del experimento.

1.4. Propiedades

1.4.1. Primeras propiedades de la probabilidad

1. $P(\emptyset) = 0$. En efecto, sea la sucesión $A, \emptyset, \emptyset, \emptyset, \dots$, con $A \in \mathcal{F}$, evidentemente $\bigcup_{n=1}^{\infty} A_n = A \cup \emptyset \cup \emptyset \dots = A$ con lo que por el axioma b)

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \leftrightarrow P(A) + \sum_{n=2}^{\infty} P(A_n) = P(A) \mapsto P(\emptyset) = 0$$

2. Se cumple la aditividad finita. Sea la sucesión $A_1, A_2, \dots, A_n, \emptyset, \emptyset, \dots$, donde $A_i \in \mathcal{F} \quad \forall i = 1, 2, \dots, n$ y $A_i \cap A_j = \emptyset \quad \forall i \neq j$ con $i, j \leq n$

Evidentemente $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^n A_i$, con lo que por el axioma b)

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \Leftrightarrow P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

como queríamos demostrar

3. Intuitivamente es evidente que, si el suceso B ocurre siempre que se presenta el suceso A, debe ser $P(A) \leq P(B)$. Formalmente, esta propiedad se enuncia:

$$\text{Si } A \subset B \subset \Omega \text{ entonces } P(A) \leq P(B) \quad (1.1)$$

Sin embargo, éste no es un requisito nuevo, sino que se deduce de los impuestos en la definición. En efecto, si $A \subset B$, es $B = A \cup (B \cap A^c)$; como A y $(B \cap A^c)$ son sucesos disjuntos, se obtiene de (2)

$$P(B) = P(A) + P(B \cap A^c)$$

y como $P(B \cap A^c)$ es un número positivo, resulta $P(B) \geq P(A)$.

4. El razonamiento anterior es más preciso que la propiedad enunciada. Según la última igualdad, el suceso $B - A = B \cap A^c$, que puede describirse de palabra como ocurre B pero no sucede A, tiene probabilidad $P(B \cap A^c) = P(B) - P(A)$. Es decir:

$$\text{Si } A \subset B \subset \Omega \text{ entonces } P(B - A) = P(B) - P(A) \quad (1.2)$$

5. Si en la propiedad anterior, se toma $B = \Omega$, el suceso $B \cap A^c$ es sencillamente A^c , el suceso contrario de A; de forma que se verifica:

$$\text{Para cualquier } A \subset \Omega, \quad P(A^c) = 1 - P(A) \quad (1.3)$$

propiedad sumamente útil en múltiples ocasiones.

6. Como generalización del razonamiento llevado a cabo en 3), si A y B son sucesos cualesquiera, se puede expresar

$$B = (B \cap A) \cup (B \cap A^c)$$

de donde, puesto que $B \cap A$ y $B \cap A^c$ son disjuntos, se obtiene

$$P(B) = P(B \cap A) + P(B - A) \quad (1.4)$$

lo cual enseña a descomponer la probabilidad de un suceso B en dos casos: aquél en que también sucede A y aquél en que A no ocurre.

7. La propiedad de aditividad 2) no permite obtener directamente la probabilidad de $A \cup B$, salvo en el caso en que $A \cap B = \emptyset$. Si no es así, se puede poner

$$A \cup B = A \cup (B - A)$$

y, dado que A y $B - A$ son disjuntos, será

$$P(A \cup B) = P(A) + P(B - A)$$

resultado que, combinado con 1.4 proporciona

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (1.5)$$

la interpretación en términos de frecuencias es sencilla: si A y B pueden ocurrir simultáneamente, al sumar la frecuencia con la que ocurre A más la frecuencia con la que ocurre B , habrán sido contados dos veces los casos en que sucede $A \cap B$. Para obtener la frecuencia con que sucede A ó B , es preciso descontar la frecuencia de $A \cap B$.

8. Como generalización de la propiedad anterior, podemos expresar la probabilidad de la unión de n elementos no disjuntos de la σ -álgebra \mathcal{F} de la forma

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \\ &+ \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n) \end{aligned}$$

como fácilmente se demuestra por inducción

9. P es subaditiva, ya que $\forall A, B \in \mathcal{F}$ es $P(A \cup B) \leq P(A) + P(B)$, como fácilmente se deduce de la propiedad 7), al ser $P(A \cap B) \geq 0$.
10. Como generalización de la propiedad anterior se tiene que

$$\forall A_1, \dots, A_n \in \mathcal{A} \quad P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$$

como se demuestra fácilmente por inducción.

Proposición 1.4. Sea (Ω, \mathcal{F}) un espacio medible para cualquier sucesión $\{A_n\} \subset \mathcal{F}$ se verifica

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n) \quad (\text{desigualdad de Boole}) \quad (1.6)$$

Ejemplos

- 5.- Sean A, B y C tres sucesos de un espacio probabilístico (Ω, \mathcal{F}, P) tales que $P(A) = 0,2$, $P(B) = 0,4$, $P(C) = 0,3$, $P(A \cap B) = 0,1$ y $(A \cup B) \cap C = \emptyset$. Calcular las probabilidades de los siguientes sucesos: *Solamente ocurre A. Los tres sucesos ocurren. Ocurren A, y B, pero no C. Por lo menos dos ocurren. Ocurren dos y no más. No ocurren más de dos. Ocurre por lo menos uno. Ocurre sólo uno. No ocurre ninguno.*
- 6.- Decir si son ciertas o falsas las siguientes afirmaciones, demostrándolas o poniendo un contraejemplo:
- a) Si $P(A) + P(B) > 1$, entonces $A \cap B \neq \emptyset$.
 - b) Si $P(A) = P(B) = p$, entonces $P(A \cap B) \leq P^2$.
 - c) Si $P(A) = P(B^c)$, entonces $A^c = B$.
 - d) Si $P(A) = 0$, entonces $A = \emptyset$.
 - e) Si $P(A) = 0$ y $P(B) = 1$, entonces $P(A \cap B) = 0$.

1.5. Construcción de espacios de probabilidad. Caso Discreto

Ahora, es fácil saber qué datos son suficientes para determinar la probabilidad de cualquier suceso en un espacio de probabilidad finito: Si $\Omega = \{w_1, w_2, \dots, w_n\}$ es el espacio muestral de un espacio de probabilidad finito (Ω, P) y $A = \{w_{i_1}, w_{i_2}, \dots, w_{i_k}\}$ es un suceso, se cumple:

$$P(A) = P(\{w_{i_1}\}) + P(\{w_{i_2}\}) + \dots + P(\{w_{i_k}\}) \quad (1.7)$$

En consecuencia, basta conocer la probabilidad de cada uno de los sucesos simples para poder determinar la probabilidad de cualquier suceso. Para que dichos datos sean coherentes, tendría que ser

$$P(\{w_1\}) + P(\{w_2\}) + \dots + P(\{w_n\}) = 1 \quad (1.8)$$

Evidentemente la conclusión se obtiene de la propiedad de aditividad

$$A = \{w_{i_1}\} \cup \{w_{i_2}\} \cup \dots \cup \{w_{i_k}\}$$

siendo disjuntos cada par de términos de la unión. La igualdad 1.8 es el caso particular de 1.7 correspondiente a $A = \Omega$.

En un espacio de probabilidad numerable $\Omega = \{w_1, w_2, \dots, w_n, \dots\}$, cualquier suceso $A = \{w_{i_1}, w_{i_2}, \dots, w_{i_k}, \dots\}$ tiene probabilidad

$$P(A) = \sum_{k=1}^{\infty} P(\{w_{i_k}\}) \quad (1.9)$$

Así pues, la probabilidad de cada uno de los sucesos simples es un conjunto de datos suficientes para poder determinar la probabilidad de cualquier suceso. Para que dichos datos sean correctos, tendrá que ser

$$\sum_{n=1}^{\infty} P(\{w_n\}) = 1 \quad (1.10)$$

Teorema 1.1. *Sea el espacio muestral Ω finito o infinito numerable, y el conjunto $\mathcal{P}(\Omega)$ que sabemos es una σ -álgebra. Una función $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ es una probabilidad sobre $(\Omega, \mathcal{P}(\Omega))$ si y sólo si*

1. $P(\{w\}) \geq 0 \quad \forall w \in \Omega, \quad \sum_{w \in \Omega} P(\{w\}) = 1.$
2. $P(\emptyset) = 0.$
3. $\forall A \in \mathcal{P}(\Omega), \quad A \neq \emptyset \text{ es } P(A) = \sum_{w \in A} P(\{w\}).$

La regla de Laplace

Para atribuir probabilidad a los sucesos relativos a un fenómeno aleatorio con espacio muestral finito, existe una norma que resulta útil si todos los sucesos elementales tienen la misma probabilidad de ocurrir en una realización del experimento. En este caso se dice que el experimento está dotado de simetría. Si los posibles sucesos elementales son e_1, e_2, \dots, e_n , y todos ellos son igualmente probables se tendrá que $P(e_1) = P(e_2) = \dots = P(e_n)$; como $\bigcup_{i=1}^n e_i = \Omega$ y son incompatibles, esto nos lleva a que

$$P(\Omega) = 1 = \sum_{i=1}^n P(e_i) \text{ y } P(e_i) = \frac{1}{n}$$

de donde para cualquier suceso $A \in \mathcal{P}(\Omega)$, como

$$A = \bigcup_{e_i \in A} e_i, \quad P(A) = \frac{\text{número de } e_i \in A}{n}$$

es decir,

$$P(A) = \frac{\text{número de casos favorables al suceso } A}{\text{número de casos posibles}}$$

y obtenemos la llamada definición clásica de Laplace de Probabilidad.

Observemos que hay muchos experimentos aleatorios en los que no se cumple la condición de simetría y además esta definición y concepción de la probabilidad es muy atacable desde el punto de vista lógico y desde el punto de vista aplicado.

Observemos, no obstante, que hemos llegado a ella a partir de nuestra definición axiomática y, por tanto, la llamada regla de Laplace es un caso particular de la definición axiomática de la probabilidad.

Ejemplos

- 7.- Se escoge un número al azar entre los que forman el conjunto $\{0, 1, 2, \dots, 10^n - 1\}$. Calcular la probabilidad P_k de que en el sistema decimal este número esté formado por k cifras.
- 8.- Distribuimos al azar n bolas blancas y n bolas negras en n urnas. Hallar:
 - La probabilidad de que una urna especificada contenga i bolas blancas y j bolas negras.
 - La probabilidad de que en cada urna haya una bola de cada color.

1.6. Construcción de espacios de probabilidad. Caso Continuo

La probabilidad de un suceso $A \subset \mathbb{R}$ no puede obtenerse sumando la probabilidad de los puntos contenidos en A ; por la sencilla razón de que cada punto tendrá probabilidad cero y sólo la presencia en A de un número no numerable de puntos permite que su probabilidad no sea nula. Así, en los ejemplos 2 y 3 del epígrafe 1.1 parece natural que una duración o una distancia exactas no tienen ninguna probabilidad de ocurrir entre un número continuo de alternativas posibles. No obstante, puede haber una probabilidad positiva de que la duración del trayecto esté entre 18 y 20 segundos, o de que la distancia al centro de la moneda, medida en metros, pertenezca al intervalo $(4,5)$.

Puestas así las cosas, es necesario encontrar un procedimiento alternativo para especificar las probabilidades de los sucesos.

Una analogía física puede aclarar las dificultades descritas y la solución que se adopta: Un conjunto de masas puntuales aisladas constituye un modelo físico discreto, válido para cosas tales como el choque de bolas de billar, o el movimiento del sistema solar. Sin embargo, no todos los sistemas de masas son así. Un alambre o un hilo metálico tiene una masa repartida de forma continua; cada punto es demasiado pequeño para contar con masa distinta de cero, pero cualquier intervalo del alambre tiene una masa positiva.

La descripción física de este tipo de cuerpos consiste en introducir la noción de densidad puntual. En principio, la densidad es un cociente: el cociente de la masa dividida por la longitud. Pero, para un alambre no homogéneo, de composición variable, la densidad se convierte en una característica puntual. La densidad $\rho(x)$ en un punto x del alambre es el cociente entre la masa y la longitud de un segmento infinitesimal de alambre alrededor del punto. Dicho con precisión matemática

$$\rho(x) = \lim_{h \downarrow 0} \frac{\text{masa de } I_x(h)}{2h}$$

donde $I_x(h)$ representa el intervalo de longitud $2h$ alrededor del punto x . Así pues, la densidad puntual tiene el carácter de derivada de la masa respecto a la longitud. Por consiguiente, la expresión

$$\text{masa de } I = \int_I \rho(x) dx$$

permite calcular la masa de cualquier intervalo I del alambre; después, si se recortan diversos trozos del alambre, la masa total será la suma de cada uno de ellos (incluso en el caso hipotético de que se hubiesen tomado un número numerable de trozos). La densidad $\rho(x)$ es, por consiguiente, el dato adecuado para describir el alambre, en todo lo relativo a las masas que contiene y a la manera en que están distribuidas a lo largo de él.

La misma idea sirve en el contexto de la probabilidad. Se pueden caracterizar probabilidades sobre \mathbb{R} mediante una densidad de probabilidad, $f(x)$, con la cual calcular la probabilidad de cualquier intervalo, I , en la forma

$$P(I) = \int_I f(x) dx$$

Desde luego, la función de densidad de probabilidad no puede ser una función arbitraria.

Definición 1.7. Una función $f : \mathbb{R} \rightarrow \mathbb{R}$ (con un número finito de discontinuidades) se denomina densidad de probabilidad si verifica

1. $f(x) \geq 0$ para todo $x \in \mathbb{R}$.

$$2. \int_{\mathbb{R}} f(x)dx = 1.$$

El modelo probabilístico asociado a la función de densidad f asigna probabilidad

$$P(I) = \int_I f(x)dx \tag{1.11}$$

a cualquier intervalo I de \mathbb{R}

El primer requisito es indispensable para que la probabilidad $P(I)$, de cualquier intervalo I , sea no negativa. La segunda condición se impone para conseguir que sea $P(\mathbb{R}) = 1$. La expresión 1.11 indica que $P(I)$ coincide con el área situada bajo la gráfica de $f(x)$, entre los extremos del intervalo I ; de manera que lo que interesa sobre todo de una función de densidad es como reparte el área total (igual a 1) bajo su gráfica.

Capítulo 2

Probabilidad Condicionada

2.1. Introducción

En el capítulo anterior, las probabilidades han tenido un carácter estático: antes de que un experimento aleatorio se lleve a cabo, se puede emitir un juicio sobre su resultado, que indique la frecuencia de cada uno de los sucesos que pueden ocurrir, una vez que el experimento haya finalizado.

Sin embargo, es muy útil considerar situaciones intermedias en las que el experimento no ha concluido, pero ya se dispone de alguna información acerca de la actuación del azar en la realización del fenómeno que se está llevando a cabo. Es indudable que tal información puede modificar los juicios y previsiones sobre los resultados finales.

Se requiere para ello del concepto de probabilidad condicionada que se introduce a continuación.

2.2. Probabilidad Condicionada.

Un ejemplo simple en el que se dispone de alguna información sobre el resultado de un fenómeno aleatorio es el siguiente:

Ejemplo 1:

Una persona extrae una carta al azar de una baraja española, la mira y la coloca boca abajo sobre la mesa, sin permitir que se vea. Si pregunta la probabilidad de que

sea un rey, la respuesta es inmediata

$$P(\text{Rey}) = \frac{4}{40} = \frac{1}{10}$$

Ahora bien, si como ayuda para identificar la carta, proporciona la información de que es una figura (sota, caballo o rey), esta nueva información obliga a modificar la asignación de probabilidades: la carta extraída sólo puede ser ahora una de las 12 figuras y en 4 casos será un rey; así que la probabilidad de que sea un rey es $4/12 = 1/3$. No hay nada sorprendente en que se obtenga un resultado distinto, puesto que la información de la que se parte en ambos casos es diferente. Pero, para no inducir a error, conviene escribir este nuevo resultado en la forma

$$P(\text{Rey} / \text{Figura}) = \frac{1}{3}$$

incluyendo no sólo el suceso del cual se obtiene la probabilidad, sino también la información de la que se dispone al hacerlo.

Análogamente, si se nos hubiese comunicado que la carta no es una figura, no hay ninguna oportunidad de que sea un rey; de manera que se concluiría

$$P(\text{Rey}/\text{No es una figura}) = 0$$

Tercer resultado que se diferencia de los anteriores en la información disponible y no en el suceso cuya probabilidad se quiere evaluar.

De manera general, al llevar a cabo un experimento aleatorio se puede saber que se ha producido un cierto suceso B; es decir, que el resultado del experimento pertenece al subconjunto B de Ω . Esta información modifica, en principio, la probabilidad de que haya ocurrido cualquier otro suceso A; así que se representa por $P(A/B)$ la probabilidad que tiene el suceso A en tales circunstancias y se denomina probabilidad de A condicionada por B.

En espacio muestrales finitos, tal y como ocurre en el ejemplo anterior, es frecuente que la probabilidad condicionada de A por B pueda calcularse directamente, enumerando los nuevos “casos posibles” aquellos en los que sucede B y los “casos favorables” aquellos en los que sucede también A para obtener

$$P(A/B) = \frac{\text{número de casos favorables a A y B}}{\text{número de casos en los que ocurre B}} \quad (2.1)$$

Esta regla, útil en espacios muestrales finitos, bajo el supuesto de que todos los resultados sean igualmente probables, da la pauta para lograr una definición general, válida en cualquier circunstancia. En efecto, si se divide numerador y denominador de la fracción 2.1 por el número total de casos posibles (es decir, por el cardinal de Ω) se obtiene el cociente entre $P(A \cap B)$ y $P(B)$. Por ello, cualquiera que sea el espacio muestral, se define:

Definición 2.1. En un espacio de probabilidad (Ω, \mathcal{F}, P) , si $B \in \mathcal{F}$ es un suceso con $P(B) > 0$, para cada $A \in \mathcal{F}$, se denomina probabilidad condicionada por B a

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

La interpretación es que la frecuencia con que se espera que se presente A , entre las repeticiones en las que sucede B , coincide con la frecuencia con la que ocurren ambos simultáneamente, dividida por la frecuencia con la que se da B .

Es bien sabido que en un espacio de probabilidad (Ω, \mathcal{F}, P) , la información de que el resultado del fenómeno aleatorio asociado pertenece a un determinado suceso $B \in \mathcal{F}$, afecta a la probabilidad de todos los sucesos: en lugar de su probabilidad original $P(A)$, debe asignarse a cada uno de ellos su probabilidad condicionada por B ; $P(A/B)$.

De esta forma, saber que se ha verificado un cierto suceso B transforma el espacio de probabilidad original en un nuevo espacio de probabilidad $(\Omega, \mathcal{F}, P(. / B))$. Incluso, puesto que se sabe que B^c no se ha realizado, lo relevante es el espacio de probabilidad (B, \mathcal{F}_B, P_B) , en el cual

$$\mathcal{F}_B = \{A \cap B / A \in \mathcal{F}\}$$

es la nueva σ -álgebra de sucesos; cada uno de los cuales $C \in \mathcal{F}_B$, recibe probabilidad

$$P_B(C) = \frac{P(C)}{P(B)} = \frac{P(A \cap B)}{P(B)} \text{ si } C = A \cap B \text{ con } A \in \mathcal{F}$$

Las comprobaciones formales necesarias constituyen la demostración del siguiente Teorema

Teorema 2.1. Dado un espacio de probabilidad (Ω, \mathcal{F}, P) , si $B \in \mathcal{F}$ es un suceso con $P(B) > 0$, entonces $(\Omega, \mathcal{F}, P(. / B))$ y (B, \mathcal{F}_B, P_B) son espacios de probabilidad.

Desde luego es $P(A/B) \geq 0$ para cualquier $A \in \mathcal{F}$ y $P(\Omega/B) = 1$. Además para cualquier colección numerable $\{A_n\} \subset \mathcal{F}$ de sucesos disjuntos

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n / B\right) &= \frac{P\left(\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_{n=1}^{\infty} (A_n \cap B)\right)}{P(B)} \\ &= \sum_n \frac{P(A_n \cap B)}{P(B)} = \sum_n P(A_n / B) \end{aligned}$$

Luego $P(. / B)$ es una probabilidad sobre el espacio medible (Ω, \mathcal{F}) .

Por otra parte, \mathcal{F}_B es la σ -álgebra restringida a B y, en el espacio medible (B, \mathcal{F}_B) , la función P_B cumple $P_B(A) \geq 0$, $P_B(B) = 1$ y

$$P_B\left(\bigcup_n A_n\right) = \frac{P\left(\bigcup_n A_n\right)}{P(B)} = \sum_n \frac{P(A_n)}{P(B)} = \sum_n P_B(A_n)$$

siempre que $\{A_n\} \subset \mathcal{F}_B$ y $A_n \cap A_m = \emptyset$ para $n \neq m$.

2.3. Teoremas de la Probabilidad Condicionada.

Los resultados siguientes se basan en el concepto de probabilidad condicionada.

Proposición 2.1. (Regla de la multiplicación) Si $A_1, A_2, \dots, A_n \in \mathcal{F}$ son sucesos en un espacio de probabilidad (Ω, \mathcal{F}, P) y $P(\bigcap_{i=1}^{n-1} A_i) > 0$ se cumple

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \dots P(A_n/A_1 \cap A_2 \cap \dots \cap A_{n-1}) \quad (2.2)$$

En efecto, la aplicación reiterada de la igualdad $P(A \cap B) = P(B)P(A/B)$ proporciona

$$\begin{aligned} P\left(\bigcap_{i=1}^n A_i\right) &= P\left(\bigcap_{i=1}^{n-1} A_i\right)P\left(A_n / \bigcap_{i=1}^{n-1} A_i\right) \\ &= P\left(\bigcap_{i=1}^{n-2} A_i\right) \cdot P\left(A_{n-1} / \bigcap_{i=1}^{n-2} A_i\right) \cdot P\left(A_n / \bigcap_{i=1}^{n-1} A_i\right) \end{aligned}$$

y así sucesivamente hasta llegar a 2.2

2.4. Independencia de sucesos.

Como ha quedado de manifiesto, la realización de un suceso B, puede alterar las probabilidades de los restantes, ya que obliga a asignarles su probabilidad condicionada por B en lugar de su probabilidad inicial.

Un suceso puede influir, favorablemente o desfavorablemente, en la aparición de otro. Por ejemplo, al extraer una carta de una baraja de 40 naipes, el suceso B: *no es un as*, favorece al suceso A: *es figura*. En el sentido de que la probabilidad a posteriori de A es mayor que la probabilidad a priori.

$$P(A/B) = \frac{1}{3} > P(A) = \frac{3}{10}$$

En cambio, el suceso C: *la carta tiene un número impar*, es desfavorable al suceso A, puesto que entre las veinte cartas de numeración impar sólo hay cuatro figuras, con lo cual se tiene:

$$P(A/C) = \frac{1}{5} < P(A) = \frac{3}{10}$$

Si un suceso influye, favorable o desfavorablemente, en otro, se dice que el segundo es *dependiente* del primero. Esta influencia no significa, en modo alguno, que haya una relación causal entre ambos, sino sólo que $P(A/B) \neq P(A)$, de forma que A se da con una frecuencia distinta, en las ocasiones en las que se produce B, que en el conjunto de todas las realizaciones.

No siempre la probabilidad condicionada es distinta de la inicial. Por ejemplo, si se sabe que la carta elegida pertenece al suceso D: *es un basto*, la probabilidad de que sea una figura sigue siendo 3/10. Esto es:

$$P(A/D) = \frac{3}{10} = P(A)$$

Así, puede concluirse que el palo de la baraja que resulte elegido, no altera la probabilidad de que la carta sea una figura. O, dicho de otro modo, que el hecho de que la carta sea una figura, es *independiente* del palo obtenido, en el sentido de que el palo no añade información sobre si se trata o no de una figura.

Esta idea da lugar a la siguiente definición:

Definición 2.2. *En un fenómeno aleatorio, el suceso A se dice independiente del suceso B, supuesto $P(B) > 0$, si se cumple:*

$$P(A/B) = P(A) \tag{2.3}$$

La ecuación anterior equivale a que se cumpla:

$$\frac{P(A \cap B)}{P(B)} = P(A)$$

o bien

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

La simetría en esta última relación pone de relieve que B es independiente de A si A lo es de B. Más vale entonces expresar la definición en la forma:

Definición 2.3. *Dos sucesos A y B de un espacio de probabilidad se dicen independientes si se cumple:*

$$P(A \cap B) = P(A)P(B) \tag{2.4}$$

Con la ventaja adicional de que la definición sirve incluso si $P(A) = 0$ ó $P(B) = 0$. Se evita así, tener que añadir que un suceso de probabilidad cero es independiente de cualquier otro, puesto que ambos miembros de la igualdad 2.4 se anulan.

Obsérvese que, si A y B son sucesos independientes, también lo son A y B^c . En efecto,

$$\begin{aligned} P(A \cap B^c) &= P(A - A \cap B) = P(A) - P(A \cap B) \\ &= P(A)[1 - P(B)] = P(A)P(B^c) \end{aligned}$$

Lo mismo sucede, por consiguiente, con A^c y B o con A^c y B^c .

La idea de independencia de tres sucesos A_1, A_2 y A_3 , exige que cualquier conocimiento que se tenga sobre dos de los sucesos no aporte información sobre el tercero. Por ello, la independencia de los tres sucesos, se define en la forma siguiente:

Definición 2.4. En un espacio de probabilidad, tres sucesos A_1, A_2 y A_3 son independientes si se cumplen las condiciones

1. $P(A_1 \cap A_2) = P(A_1)P(A_2)$
2. $P(A_1 \cap A_3) = P(A_1)P(A_3)$
3. $P(A_2 \cap A_3) = P(A_2)P(A_3)$
4. $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$

Las tres primeras condiciones definen la *independencia dos a dos* de los tres sucesos. Sin embargo, no implican la cuarta; de forma que tres sucesos pueden ser independientes dos a dos, sin ser independientes. Esto es lo que se pone de manifiesto en el ejemplo siguiente.

Ejemplo 2: Se escoge al azar una de las siguientes combinaciones de tres letras

$$\Omega = \{aaa, bbb, ccc, abc, acb, bac, bca, cab, cba\}$$

Los sucesos:

$$\begin{aligned} A_1 &= \text{“ hay una } a \text{ en el primer lugar”} \\ A_2 &= \text{“ hay una } a \text{ en el segundo lugar”} \\ A_3 &= \text{“ hay una } a \text{ en el tercer lugar”} \end{aligned}$$

cumplen

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$$

Por otra parte, hay un único caso favorable tanto a $A_1 \cap A_2$, como a $A_1 \cap A_3$ como a $A_2 \cap A_3$: el caso aaa . Se obtiene así

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{9}$$

Por tanto, los sucesos A_1, A_2 y A_3 son independientes dos a dos. Sin embargo, aaa es, en realidad, favorable a $A_1 \cap A_2 \cap A_3$. Luego

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{9} \neq P(A_1)P(A_2)P(A_3)$$

y los tres sucesos no son independientes.

La generalización para una colección arbitraria de sucesos, se formula en los siguientes términos:

Definición 2.5. En un espacio de probabilidad (Ω, \mathcal{F}, P) , los sucesos $\{A_i\}_{i \in I}$ se denominan sucesos independientes si

$$P\left(\bigcap_{i \in F} A_i\right) = \prod_{i \in F} P(A_i) \quad (2.5)$$

para cualquier subconjunto finito $F \subset I$.

Ejemplo 3: Sea A y B dos sucesos tales que

$$P(A) = 1/4 \quad P(B/A) = 1/2 \quad P(A/B) = 1/4$$

Razonar si son ciertas o falsas las siguientes afirmaciones

- $A \subset B$
- A y B son independientes
- A y B son incompatibles
- $P(A^c/B^c) = 1/2$
- $P(A/B) + P(A/B^c) = 1$

2.5. Teorema de la Probabilidad Total.

Sea (Ω, \mathcal{A}, P) un espacio probabilístico y sea $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$ un sistema completo de sucesos, es decir, una sucesión de sucesos disjuntos,

$$A_i \cap A_j = \emptyset \quad \forall i \neq j \quad \text{tal que} \quad \bigcup_{n=1}^{\infty} A_n = \Omega$$

Sea $B \in \mathcal{A}$ un suceso para el que se conocen las probabilidades condicionadas $P(B/A_i)$, y supongamos, por último, que se conocen también las probabilidades $P(A_i)$, con $P(A_i) > 0, \forall i = 1, 2, \dots$. En estas condiciones,

$$P(B) = \sum_{i=1}^{\infty} P(B/A_i)P(A_i)$$

Es decir, si el suceso B puede ocurrir por alguna de las causas A_i , la probabilidad de que ocurra es la suma de las probabilidades de las causas $P(A_i)$ por la probabilidad del suceso B condicionado a la causa $A_i : (P(B/A_i))$.

En efecto,

$$\begin{aligned} P(B) &= P(B \cap \Omega) = P(B \cap [\bigcup_{i=1}^{\infty} A_i]) = P(\bigcup_{i=1}^{\infty} [B \cap A_i]) = \\ &= \sum_{i=1}^{\infty} P(B \cap A_i) = \sum_{i=1}^{\infty} P(B/A_i) \cdot P(A_i) \end{aligned}$$

por ser

$$\forall i \in \mathbb{N}, P(B \cap A_i) = P(B/A_i) \cdot P(A_i)$$

Ejemplo 4:

Cierto análisis clínico se emplea en el diagnóstico de tres enfermedades, A_1, A_2, A_3 . En la población, la proporción de enfermos de cada una es 3 %, 2 % y 1 % respectivamente. Por otra parte el análisis da un resultado positivo en el 85 % de los enfermos de A_1 , el 92 % de los enfermos de A_2 y el 78 % de los enfermos de A_3 ; mientras que sólo lo da en un 0,5 % de las personas no afectadas por ninguna de las tres enfermedades. ¿Cuál es la probabilidad de que, al efectuar el análisis sobre una persona elegida al azar, el resultado sea positivo?

Si B representa el suceso de que el análisis tenga un resultado positivo y A_0 el que la persona elegida no padezca ninguna de las tres enfermedades, será:

$$\begin{aligned} P(B) &= P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + P(A_3)P(B/A_3) + P(A_0)P(B/A_0) \\ &= 0,03 \cdot 0,85 + 0,02 \cdot 0,92 + 0,01 \cdot 0,78 + 0,94 \cdot 0,005 = 0,0564 \end{aligned}$$

de acuerdo con los datos suministrados en el enunciado. Nótese que, para que la aplicación de la fórmula de las probabilidades totales sea correcta, siempre ha de ser $P(A_1) + P(A_2) + \dots + P(A_n) = 1$.

En muchos problemas del tipo anterior, lo más importante es valorar las probabilidades de las distintas circunstancias en que pueden ocurrir un suceso, cuando se sabe que éste se ha producido. Así, en el caso anterior, el interés se centrará, sin duda, en la probabilidad de que el paciente pertenezca a cada grupo, cuando se ha obtenido un resultado positivo del análisis.

2.6. Teorema de Bayes.

Consideremos ahora, bajo las mismas condiciones del teorema anterior, que estamos interesados en conocer la probabilidad de que ocurrido el suceso B la causa que lo haya producido sea la A_i . Expresado analíticamente, queremos calcular

$$P(A_i/B)$$

Teorema 2.2. *Sea (Ω, \mathcal{A}, P) un espacio probabilístico y sea $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$ un sistema completo de sucesos tal que $P(A_i) > 0, \forall i \in \mathbb{N}$. Sea $B \in \mathcal{A}$ un suceso con $P(B) > 0$, para el que se conocen las probabilidades $P(B/A_i)$. Entonces,*

$$P(A_i/B) = \frac{P(A_i) \cdot P(B/A_i)}{\sum_{i=1}^{\infty} P(A_i) \cdot P(B/A_i)}, \quad i \in \mathbb{N}$$

En efecto, por ser $P(B) > 0$, se tendrá

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} \quad \forall i \in \mathbb{N}$$

Por el teorema del producto,

$$P(A_i \cap B) = P(A_i)P(B/A_i), \quad \forall i \in \mathbb{N}$$

Por el teorema de la probabilidad total

$$P(B) = \sum_{i=1}^{\infty} P(A_i)P(B/A_i)$$

y sustituyendo los valores obtenidos en la expresión de $P(A_i/B)$ nos queda

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{\infty} P(A_i)P(B/A_i)} \quad \forall i \in \mathbb{N}$$

como queríamos demostrar.

A las probabilidades $P(A_i) > 0$ se las suele llamar probabilidades *a priori*, a las probabilidades $P(A_i/B)$ probabilidades *a posteriori*, y a las $P(B/A_i)$ *verosimilitudes*.

Ello permite obtener las probabilidades *a posteriori* de las diversas circunstancias, una vez que se ha observado el efecto B , en función de las probabilidades *a priori* de las “causas” y de las probabilidades de que en cada circunstancia se produzca el efecto B .

Ejemplo 5:

Con los datos del ejemplo 4, si el análisis ha dado un resultado positivo (B), las probabilidades de que el paciente padezca cada una de las tres enfermedades son:

$$\begin{aligned} P(A_1/B) &= \frac{P(A_1)P(B/A_1)}{P(B)} = \frac{0,03 \cdot 0,85}{0,0564} \simeq 0,452 \\ P(A_2/B) &= \frac{P(A_2)P(B/A_2)}{P(B)} = \frac{0,02 \cdot 0,92}{0,0564} \simeq 0,326 \\ P(A_3/B) &= \frac{P(A_3)P(B/A_3)}{P(B)} = \frac{0,01 \cdot 0,78}{0,0564} \simeq 0,138 \end{aligned}$$

donde se ha aprovechado el valor de $P(B)$ calculado anteriormente. Por otra parte la probabilidad de que el paciente no esté realmente afectado por ninguna de las tres enfermedades es

$$P(A_0/B) = \frac{P(A_0)P(B/A_0)}{P(B)} = \frac{0,94 \cdot 0,005}{0,0564} \simeq 0,083$$

De manera que, mientras que en el conjunto de la población las tres enfermedades se presentan con frecuencias 3 %, 2 % y 1 % respectivamente, entre los individuos que dan positivo en el análisis tales frecuencias se transforman en 45,2 %, 32,6 % y 13,8 % respectivamente. Así mismo, la proporción de sanos es 94 % en el conjunto de la población, pero sólo del 8,3 % entre los que dan positivo en el análisis.

Ejemplos

- 6.- Se consideran dos urnas U_1 y U_2 . En la U_1 hay cinco bolas rojas y cuatro bolas verdes, mientras que la U_2 está inicialmente vacía. Se extraen cuatro bolas de U_1 y se introducen en U_2 . De esta segunda urna U_2 se extrae una bola y resulta ser verde. ¿Cuál es la probabilidad de sacar una bola roja en una segunda extracción?
- 7.- Se tienen N bolas numeradas del 1 al N y N urnas también numeradas. Se introduce una bola en cada urna. Se pide:

- Probabilidad de al menos una coincidencia.
 - Probabilidad de exactamente m coincidencias.
 - Comportamiento límite de las dos probabilidades cuando $N \rightarrow \infty$.
- 8.- Sean $N + 1$ urnas cada una con N bolas rojas y blancas. La urna U_k contiene k bolas rojas y $N - k$ blancas, $k = 0, 1, 2, \dots, N$. Se elige una urna al azar y se realizan a extracciones con reemplazamiento. Supóngase que las a bolas han resultado ser rojas. Hallar la probabilidad de que en la siguiente extracción resulte bola roja.

Capítulo 3

Variables aleatorias unidimensionales

3.1. Introducción.

Frecuentemente el interés del observador de un fenómeno aleatorio está centrado en alguna consecuencia numérica derivada de su realización. Por ejemplo, para un jugador sólo cuentan las cantidades que puede ganar o perder y las respectivas probabilidades de que estos acontecimientos ocurran. Para el jugador es indiferente apostar 6 euros por cara en el lanzamiento de una moneda, que apostar la misma cantidad por número par en el lanzamiento de un dado. En ambos casos espera ganar o perder 6 euros con igual probabilidad. Lo de menos es el mecanismo concreto de que se vale el azar para producir la incertidumbre. Lo esencial es el conjunto de valores numéricos que pueden darse y las probabilidades con que es previsible que se den. Es decir, habitualmente el observador de un fenómeno aleatorio está interesado no en el espacio probabilístico concreto, sino en unos valores numéricos asociados a cada elemento del espacio de probabilidades. Al estudio de estas magnitudes gobernadas por el azar va dedicado este capítulo.

3.2. Variable aleatoria unidimensional.

Consideremos un experimento aleatorio cuyos resultados no sean numéricos. Para fijar ideas supongamos que se lanzan dos monedas equilibradas. Se tendrá el siguiente espacio de probabilidad $\Omega = \{CC, CX, XC, XX\}$ y $\mathcal{F} = P(\Omega)$ $P(CC) = P(CX) =$

$P(XC) = P(XX) = 1/4$. Para que el estudio de este fenómeno aleatorio sea más tratable matemáticamente, necesitaremos datos numéricos como resultados de los experimentos. Así pues asignaremos a cada elemento del espacio muestral un valor numérico; y el modo de hacer dicha asignación dependerá del objetivo de nuestro estudio.

Intuitivamente, una variable real es una cantidad que se mide en conexión con un experimento aleatorio. Si Ω es el espacio muestral asociado a un experimento aleatorio y w es el resultado de una realización del mismo, se lleva a cabo un proceso de medida y se obtiene un número que denotamos por $X(w)$. De esta forma, una variable aleatoria real es una función real realizada sobre el espacio muestral

$$\begin{aligned} X : \quad \Omega &\rightarrow \mathbb{R} \\ w &\rightarrow X(w) \end{aligned}$$

Por otra parte, nos interesa que esta aplicación también conserve la probabilidad asociada al experimento aleatorio.

Así pues, es necesaria alguna condición para que las funciones definidas en un espacio de probabilidad merezcan la consideración de variables aleatorias.

3.2.1. Aplicación medible.

Sean Ω_1 y Ω_2 , dos conjuntos y f una aplicación de Ω_1 en Ω_2 . A partir de f se considera la aplicación f^{-1} de $\mathcal{P}(\Omega_2)$ de tal forma que $\forall C \in \mathcal{P}(\Omega_2)$,

$$f^{-1}(C) = \{w \in \Omega_1 : f(w) \in C\}$$

Es inmediato comprobar que $\forall A \subset \Omega_2$ y $\{A_i\}_{i \in I} \subset \Omega_2$, se tiene

$$\begin{aligned} f^{-1}(A^c) &= (f^{-1}(A))^c \\ f^{-1}\left(\bigcup_{i \in I} A_i\right) &= \bigcup_{i \in I} f^{-1}(A_i) \\ f^{-1}\left(\bigcap_{i \in I} A_i\right) &= \bigcap_{i \in I} f^{-1}(A_i) \end{aligned}$$

Por tanto, si \mathcal{F}_2 familia de conjuntos es un (álgebra o σ -álgebra) sobre Ω_2 ,

$$f^{-1}(\mathcal{F}_2) = \{f^{-1}(B) / B \in \mathcal{F}_2\}$$

es un (álgebra o σ -álgebra) sobre Ω_1 .

Obsérvese que en la teoría de la medida la imagen inversa es un objetivo mucho más deseable que la imagen directa ya que las operaciones no se conservan por imagen directa:

$$h\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} h(A_i)$$

pero

$$h\left(\bigcap_{i \in I} A_i\right) \subset \bigcap_{i \in I} h(A_i)$$

pudiendo ser propia la inclusión. $h(A^c)$ no es en general igual a $(h(A))^c$ y de hecho, si h es una función constante, los dos conjuntos son disjuntos.

Definición 3.1. Sea $(\Omega_1, \mathcal{F}_1)$ y $(\Omega_2, \mathcal{F}_2)$ dos espacios medibles. Una aplicación $f : \Omega_1 \rightarrow \Omega_2$ se llama medible si y sólo si $\forall B \in \mathcal{F}_2$ se verifica $f^{-1}(B) \in \mathcal{F}_1$.

Definición 3.2. Se denomina variable aleatoria definida en un espacio de probabilidad (Ω, \mathcal{F}, P) a cualquier función $X : \Omega \rightarrow \mathbb{R}$ tal que

$$X^{-1}(B) \in \mathcal{F} \text{ para cualquier conjunto } B \in \mathbb{B} \quad (3.1)$$

$X^{-1}(B)$ designa la contraimagen de B , es decir el subconjunto de Ω

$$\{w \in \Omega / X(w) \in B\}$$

que suele representarse más simplemente por $\{X \in B\}$.

La condición 3.1 exige que sean sucesos de \mathcal{F} todos los conjuntos $\{X \in B\}$ para cualquier $B \in \mathbb{B}$ y permite referirse a su probabilidad: $P\{X \in B\}$. Cuando así ocurre, se dice también que X es una *función medible* respecto a \mathcal{F} y \mathbb{B} .

$$X^{-1}(\mathbb{B}) = \{X^{-1}(B) / B \in \mathbb{B}\}$$

es una σ -álgebra; además, la condición 3.1 significa que $X^{-1}(\mathbb{B}) \subset \mathcal{F}$; de forma que $X^{-1}(\mathbb{B})$ es la menor de las σ -álgebras que consiguen que X sea una variable aleatoria.

Si la σ -álgebra de sucesos $\mathcal{F} = \mathcal{P}(\Omega)$, la condición 3.1 queda vacía de contenido. Ello explica que para espacios muestrales discretos, cualquier función $X : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria.

Proposición 3.1. Dado un espacio de probabilidad (Ω, \mathcal{F}, P) , una función X definida en Ω es una variable aleatoria si y sólo si

$$\{X \leq c\} \in \mathcal{F} \text{ para cada } c \in \mathbb{R} \quad (3.2)$$

Desde luego, si X es una variable aleatoria, $\{X \leq c\} = X^{-1}(-\infty, c] \in \mathcal{F}$ puesto que $(-\infty, c] \in \mathbb{B}$. Para probar el recíproco, considérese

$$\mathcal{B} = \{A \subset \mathbb{R} / X^{-1}(A) \in \mathcal{F}\}$$

\mathcal{B} es una σ -álgebra puesto que

1. Si $A \in \mathcal{B} \Rightarrow X^{-1}(A) \in \mathcal{F}$ y como \mathcal{F} es σ -álgebra $(X^{-1}(A))^c \in \mathcal{F} \Rightarrow X^{-1}(A^c) \in \mathcal{F} \Rightarrow A^c \in \mathcal{B}$
2. Si $A_i \in \mathcal{B} \Rightarrow X^{-1}(A_i) \in \mathcal{F} \Rightarrow \bigcup_i X^{-1}(A_i) \in \mathcal{F} \Rightarrow X^{-1}(\bigcup_i A_i) \in \mathcal{F} \Rightarrow \bigcup_i A_i \in \mathcal{B}$

Además, 3.2 obliga a \mathcal{B} a contener todos los intervalos de \mathbb{R} ; por ejemplo

$$\begin{aligned} X^{-1}(a, b] &= X^{-1}(-\infty, b] - X^{-1}(-\infty, a] \in \mathcal{F} \\ X^{-1}(a, b) &= X^{-1}\left(\bigcup_n (a, b - 1/n]\right) = \bigcup_n X^{-1}(a, b - 1/n] \in \mathcal{F} \end{aligned}$$

etc. Por consiguiente, \mathcal{B} contiene a la σ -álgebra de Borel, \mathbb{B} ; lo cual indica que $X^{-1}(\mathcal{B}) \in \mathcal{F}$ para cualquier $\mathcal{B} \in \mathbb{B}$.

Un razonamiento similar establece que para que X sea una variable aleatoria es condición necesaria y suficiente que pertenezcan a \mathcal{F} todos los conjuntos de la forma $\{X < c\}$; o bien $\{X > c\}$; o bien $\{X \geq c\}$.

Proposición 3.2. *Si X es una variable aleatoria definida en el espacio de probabilidad (Ω, \mathcal{F}, P) y $g : \mathbb{R} \mapsto \mathbb{R}$ es una función medible respecto a \mathbb{B} (es decir, tal que $g^{-1}(B) \in \mathbb{B}$ para cualquier $B \in \mathbb{B}$), entonces $g(X)$ es una variable aleatoria. En particular, la conclusión es cierta cuando g es una función continua.*

En efecto, para cualquier $c \in \mathbb{R}$, $\{g(X) \leq c\} = \{X \in g^{-1}(-\infty, c]\}$ pertenece a \mathcal{F} , supuesto que $g^{-1}(-\infty, c] \in \mathbb{B}$. Tal es el caso si g es una función medible respecto a \mathbb{B} y, también, si g es continua ya que $g^{-1}(-\infty, c]$ es, entonces, un conjunto cerrado de \mathbb{R} .

Según esto, si X es una variable aleatoria, también lo son $X + c$, cX y $|X|^c$ cualquiera que sea $c \in \mathbb{R}$, X^n para cualquier $n \in \mathbb{N}$, $\log |X|$, etc. Lo mismo ocurre con $1/X$, supuesto que $X \neq 0$, ya que $\{1/X \leq c\}$ coincide con $\{1/c \leq X < 0\}, \{X < 0\}$ ó $\{X < 0\} \cup \{X \geq 1/c\}$, según que c sea menor, igual o mayor que cero; y los tres pertenecen a \mathcal{F} .

3.2.2. Operaciones con variables aleatorias.

Sea (Ω, \mathcal{F}, P) un espacio probabilístico; sean X e Y dos variables aleatorias. Entonces:

1. $X + Y$ es una variable aleatoria si está definida la suma.
2. $X - Y$ es una variable aleatoria, si está definida la diferencia.

3. $X \cdot Y$ es una variable aleatoria.

Demostración:

1. Sea

$$X(w) + Y(w) > a \Rightarrow X(w) > a - Y(w) \Rightarrow$$

\Rightarrow existe un número racional r tal que

$$X(w) > r > a - Y(w)$$

con lo que

$$\begin{aligned} \{w/X(w) + Y(w) \leq a\}^c &= \{w/X(w) + Y(w) > a\} = \\ &= \bigcup_r \left[\{w/X(w) > r\} \cap \{w/Y(w) > a - r\} \right] \end{aligned}$$

y como esta unión es numerable, entonces pertenecerá a \mathcal{F} por ser X e Y variables aleatorias $\Rightarrow X + Y$ es v.a., por ser \mathcal{F} cerrada por complementarios.

2. $-Y$ es una variable aleatoria por poder ponerse como $(-1)Y$, y a continuación aplicándose el apartado anterior, $-Y$ es v.a. $\Rightarrow X + (-Y) = X - Y$ es variable aleatoria.

3. Por último,

$$X \cdot Y = \frac{1}{2}[(X + Y)^2 - X^2 - Y^2]$$

con lo cual $X \cdot Y$ será también variable aleatoria.

Definición 3.3. *Función indicatriz. Cualquier suceso $A \in \mathcal{F}$ de un espacio de probabilidad tiene asociada la variable aleatoria:*

$$I_A = \begin{cases} 1 & \text{si } w \in A \\ 0 & \text{si } w \notin A \end{cases}$$

que es medible, puesto que $\{I_A \leq c\}$ coincide con \emptyset , A^c o Ω , según el valor de c . Se llama a I_A función indicatriz del suceso A .

Ejemplo 1: Sea $\Omega = \{w_{ijk}\}$ el espacio muestral correspondiente al experimento aleatorio de lanzar 3 veces una moneda. Sea $\mathcal{A} = \{A_1, A_2, A_3, A_1 \cup A_2, A_1 \cup A_3, A_2 \cup A_3, \emptyset, \Omega\}$ una σ -álgebra donde $A_1 = \{w_{ccx}, w_{cxc}\}$, $A_2 = \{\Omega - \{A_1 \cup A_3\}\}$ y $A_3 = \{w_{xxx}\}$. Sea $X(w_{ijk}) =$ número de caras. ¿Es X una variable aleatoria?

3.3. Probabilidad inducida por una variable aleatoria.

En muchas ocasiones el interés acerca de un fenómeno aleatorio se centra en el valor que adopta una cierta variable aleatoria X , definida en el espacio de probabilidad (Ω, \mathcal{F}, P) . De ser cierto, esto supone una gran simplificación del problema pues, mientras que el espacio de probabilidad original puede tener una estructura muy compleja, la variable X transfiere el estudio a un espacio de probabilidad unidimensional.

De hecho, cualquier variable aleatoria X , definida en un espacio de probabilidad (Ω, \mathcal{F}, P) , induce una medida de probabilidad P_X en el espacio medible (\mathbb{R}, \mathbb{B}) . Basta definir

$$P_X(B) = P\{X \in B\} = P(X^{-1}(B))$$

para cada conjunto de Borel $B \in \mathbb{B}$

Vamos a ver que P_X es efectivamente una medida de probabilidad:

- 1) $P_X(B) \geq 0, \forall B \in \mathbb{B}$ por su propia definición.
- 2) $P_X(\mathbb{R}) = P\{X^{-1}(\mathbb{R})\} = P(\Omega) = 1$, por ser P una medida de Probabilidad.
- 3) Sea $\{B_i\}_{i \in \mathbb{N}} \subset \mathbb{B}$ una sucesión tal que $B_i \cap B_j = \emptyset, \forall i \neq j$. Entonces,

$$\begin{aligned} P_X\left(\bigcup_{i=1}^{\infty} B_i\right) &= P\left\{X^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right)\right\} = P\left\{\bigcup_{i=1}^{\infty} X^{-1}(B_i)\right\} = \\ &= \sum_{i=1}^{\infty} P\{X^{-1}(B_i)\} = \sum_{i=1}^{\infty} P_X(B_i) \end{aligned}$$

ya que $X^{-1}(B_i) \cap X^{-1}(B_j) = \emptyset, \forall i \neq j$, por ser X^{-1} un homomorfismo respecto a la unión, intersección y complementación en una σ -álgebra.

Definición 3.4. Dada una variable aleatoria X en un espacio de probabilidad (Ω, \mathcal{F}, P) , la medida de probabilidad en (\mathbb{R}, \mathbb{B}) definida por

$$P_X(B) = P\{X \in B\} \text{ para cada } B \in \mathbb{B},$$

se denomina la distribución de la variable aleatoria X y $(\mathbb{R}, \mathbb{B}, P_X)$ recibe el nombre de espacio de probabilidad inducido por X o espacio de probabilidad canónica de X .

Ejemplo 2: Consideremos $\Omega = \{w \in \mathbb{R} / 1 \leq w \leq 10\}$ $\mathcal{C} = \{A, B, C\}$ donde $A = \{w \in \mathbb{R} / 1 \leq w \leq 5\}$ $B = \{w \in \mathbb{R} / 2 \leq w \leq 8\}$ $C = \{w \in \mathbb{R} / 4 \leq w \leq 10\}$ y sea $\mathcal{F} = \sigma(\mathcal{C})$. Consideremos la medida de probabilidad dada por

$$P(a, b] = P(a, b) = P(a, b] = P[a, b) = \frac{b - a}{9}$$

Sea

$$X = \begin{cases} -1 & 1 \leq w < 2 \\ 0 & 2 \leq w < 4 \\ 1 & 4 \leq w < 10 \end{cases}$$

¿Es X una variable aleatoria? En caso afirmativo calcular su función de distribución y $\int_{\Omega} X dP$

3.4. Función de Distribución.

Afortunadamente, es posible caracterizar cualquier medida de probabilidad P en (\mathbb{R}, \mathbb{B}) mediante un tipo de dato muy familiar: una función real de variable real. Por supuesto, no se trata de una función arbitraria, sino de una cierta clase de funciones, denominadas funciones de distribución.

Definición 3.5. *Una función $F : \mathbb{R} \rightarrow [0, 1]$ recibe el nombre de función de distribución si:*

1. *es creciente; es decir $F(x_1) \leq F(x_2)$ siempre que sea $x_1 < x_2$.*
2. *es continua por la derecha; esto es, para cada $x \in \mathbb{R}$, se cumple*

$$F(x) = \lim_{y \rightarrow x, y > x} F(y).$$

3. *verifica*

$$\lim_{x \rightarrow -\infty} F(x) = 0 \text{ y } \lim_{x \rightarrow +\infty} F(x) = 1.$$

La condición 1. implica la existencia de los límites laterales

$$F(x^+) = \lim_{y \rightarrow x, y > x} F(y) \text{ y } F(x^-) = \lim_{y \rightarrow x, y < x} F(y).$$

Según 2., $F(x^+)$ coincide con $F(x)$. En cambio, en algunos puntos, $F(x^-)$ puede ser diferente de $F(x)$, de manera que la función presenta un salto en el punto x . La condición 3. se abrevia con frecuencia en la forma

$$F(-\infty) = 0 \text{ y } F(+\infty) = 1.$$

Cualquier medida de probabilidad P en (\mathbb{R}, \mathbb{B}) tiene asociada una determinada función de distribución. Basta considerar los conjuntos de Borel de la forma $(-\infty, x]$ y definir, para cada $x \in \mathbb{R}$,

$$F(x) = P((-\infty, x]) \quad (3.3)$$

Proposición 3.3. *Si P es una medida de probabilidad en (\mathbb{R}, \mathbb{B}) , la función F definida por 3.3 es una función de distribución.*

En efecto, si $x_1 < x_2$ es $(-\infty, x_1] \subset (-\infty, x_2]$, entonces $F(x_1) \leq F(x_2)$ y F es creciente. Además, si x_n es una sucesión en \mathbb{R} decreciente hacia x , la sucesión de intervalos $(-\infty, x_n]$ decrece hacia $(-\infty, x]$; así que, en virtud del punto 3. de la definición de F , se cumple

$$F(x) = P((-\infty, x]) = \lim_{n \rightarrow \infty} P((-\infty, x_n]) = \lim_{n \rightarrow \infty} F(x_n).$$

Según esto, F es continua por la derecha. Análogamente, si $x_n \downarrow -\infty$, el intervalo $(-\infty, x_n]$ decrece hacia el vacío, con lo cual

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P((-\infty, x_n]) = P(\emptyset) = 0 \text{ o bien } F(-\infty) = 0.$$

Y, si $x_n \uparrow \infty$, el intervalo $(-\infty, x_n]$ crece hacia \mathbb{R} , de forma que

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P((-\infty, x_n]) = P(\mathbb{R}) = 1 \text{ o bien } F(+\infty) = 1.$$

Pero lo más interesante es lo contrario: no sólo cualquier probabilidad P tiene asociada una función de distribución, sino que cualquier función de distribución F identifica, mediante la relación 3.3 una única medida de probabilidad P que asigna un valor a cada conjunto de Borel $B \in \mathbb{B}$.

Proposición 3.4. *Si $F : \mathbb{R} \mapsto [0, 1]$ es una función de distribución, existe una única probabilidad en (\mathbb{R}, \mathbb{B}) tal que*

$$P((-\infty, x]) = F(x) \text{ para cada } x \in \mathbb{R}.$$

La unicidad es consecuencia inmediata de la proposición 1.8 del capítulo 1. Supongamos que P_1 y P_2 son dos probabilidades en (\mathbb{R}, \mathbb{B}) asociadas a la función de distribución F . Entonces P_1 y P_2 coinciden sobre la π -clase de los conjuntos de la forma $(-\infty, x]$, la cual engendra la σ -álgebra \mathbb{B} . Luego P_1 y P_2 coinciden sobre \mathbb{B} .

Pese a que el proceso de extensión que prueba el resultado anterior no proporciona un método operativo para el cálculo de $P(B)$ para un conjunto de Borel arbitrario B , la proposición anterior es muy útil porque establece una correspondencia biyectiva entre las medidas de probabilidad en (\mathbb{R}, \mathbb{B}) y las funciones de distribución; así que, el

mejor procedimiento para caracterizar una de aquéllas es identificarla mediante la función de distribución asociada. Por eso es frecuente utilizar la notación P_F para referirse a la probabilidad sobre (\mathbb{R}, \mathbb{B}) asociada a la función de distribución F mediante 3.3

Como cualquier medida de probabilidad en (\mathbb{R}, \mathbb{B}) , la distribución P_X de una variable aleatoria ha de manejarse a través de la función de distribución asociada en virtud de la proposición anterior.

Definición 3.6. *Dada una variable aleatoria X definida en un espacio de probabilidad (Ω, \mathcal{F}, P) , la función de distribución F_X asociada a P_X :*

$$F_X(x) = P_X((-\infty, x]) = P\{X \leq x\}$$

se denomina función de distribución de la variable aleatoria X .

Naturalmente, cualquier función de distribución F en \mathbb{R} es función de distribución de alguna variable aleatoria. De hecho, F determina una medida de probabilidad P_F en (\mathbb{R}, \mathbb{B}) y la variable aleatoria $X(x) = x$, definida en $(\mathbb{R}, \mathbb{B}, P_F)$ tiene función de distribución F (pues $P_F\{X \leq x\} = P_F((-\infty, x]) = F(x)$).

3.5. Distribuciones Discretas y absolutamente continuas.

3.5.1. Función de masa o cuantía.

Sea X una v.a. sobre el espacio probabilístico (Ω, \mathcal{F}, P) y sea P_X la probabilidad inducida por X sobre (\mathbb{R}, \mathbb{B}) . Llamaremos función de masa o cuantía de la v.a. X a la aplicación $p_X : \mathbb{R} \mapsto [0, 1]$, definida por

$$\forall x \in \mathbb{R} \quad p_X(x) = P_X\{x\} = P\{X^{-1}(x)\} = P\{w \in \Omega / X(w) = x\}$$

Observemos que el punto x es un boreliano, ya que se puede poner como

$$\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x + \frac{1}{n} \right]$$

Es decir, como límite de una sucesión decreciente de borelianos.

Vamos a relacionar ahora la función p_X con la función de distribución F_X .

$$\begin{aligned} p_X(x) &= P_X\{x\} = P_X \left\{ \lim_{n \rightarrow \infty} \left(x - \frac{1}{n}, x + \frac{1}{n} \right] \right\} = \lim_{n \rightarrow \infty} P_X \left\{ \left(x - \frac{1}{n}, x + \frac{1}{n} \right] \right\} = \\ &= \lim_{n \rightarrow \infty} \left[F_X\left(x + \frac{1}{n}\right) - F_X\left(x - \frac{1}{n}\right) \right] = F_X(x) - F_X(x^-). \end{aligned}$$

Es decir, p_X es precisamente el salto que experimenta la función F_X en los puntos de discontinuidad, ya que si F_X es continua en el punto $x \in \mathbb{R} \Rightarrow p_X(x) = 0$.

Proposición 3.5. *Sea X una v.a. con función de masa p_X y sea $D_X = \{x \in \mathbb{R}/p_X(x) > 0\}$. Entonces, D_X es un conjunto numerable.*

En efecto: Para $n = 1, \dots$, sea el conjunto

$$E_n = \{x \in \mathbb{R}/p_X(x) \geq \frac{1}{n}\}.$$

Como p_X representa la altura del salto en la discontinuidad de la función $F_X(x)$, y $F_X(x)$ es una función no decreciente de variación igual a uno, entonces como máximo E_1 tendrá un sólo punto; E_2 dos; E_3 tres; etc., y como

$$D_X = \bigcup_{n=1}^{\infty} E_n$$

al ser la unión de conjuntos finitos o numerables un conjunto numerable, será D_X numerable como queríamos demostrar.

3.5.2. Variable aleatoria discreta.

Sea X una v.a. definida sobre el espacio probabilístico (Ω, \mathcal{F}, P) con función de masa p_X , y sea

$$D_X = \{x \in \mathbb{R}/p_X(x) > 0\}.$$

Si $D_X \neq \emptyset$ y si $\sum_{x \in D_X} p_X(x) = 1$, la variable aleatoria X se llama discreta y D_X recibe el nombre de soporte de dicha variable aleatoria.

En la práctica, el soporte es el conjunto de los valores observables de la v.a.

Proposición 3.6. *Sea X una v.a. discreta construida de (Ω, \mathcal{F}, P) en (\mathbb{R}, \mathbb{B}) y sea D_X su soporte. Observemos que $D_X \in \mathbb{B}$, es decir, es un boreliano, ya que se obtiene como unión numerable de conjuntos de Borel. Sea p_X la función de masa de la v.a. X . Entonces, $\forall B \in \mathbb{B}$ se tiene*

$$P_X(B) = P\{w \in \Omega/X(w) \in B\} = \begin{cases} 0 & \text{si } B \cap D_X = \emptyset \\ \sum_{x \in B \cap D_X} p_X(x) & \text{si } B \cap D_X \neq \emptyset \end{cases}$$

A partir de la proposición anterior, si X es una v.a. discreta con función de masa p_X , la función de distribución es

$$F_X(x) = P_X\{(-\infty, x]\} = \sum_{\{y \in (-\infty, x] \cap D_X\}} p_X(y)$$

En tal caso, la función de distribución F_X es una función escalonada, en la cual cada salto está separado del siguiente por un intervalo en el que F_X permanece constante.

Proposición 3.7. *Dado un conjunto $D \subset \mathbb{R}$ numerable de puntos y una función $p : \mathbb{R} \mapsto [0, 1]$ tal que $p(x) = 0$ si $x \notin D$ y $p(x) > 0$ si $x \in D$, con $\sum_{x \in D} p(x) = 1$. Entonces tenemos determinada una ley de probabilidad de una v.a. discreta X , sobre el espacio probabilizable (\mathbb{R}, \mathbb{B}) definida como*

$$\begin{aligned} P_X(B) &= \sum_{x \in B \cap D} p(x), \quad \forall B \in \mathbb{R} / B \cap D \neq \emptyset \\ & \text{y} \\ P_X(B) &= 0 \quad \text{si } B \cap D = \emptyset. \end{aligned}$$

En efecto $P_X(B)$ será, de acuerdo con el teorema 1.2, una probabilidad discreta sobre (\mathbb{R}, \mathbb{B}) de función de distribución asociada

$$F(y) = \sum_{(x \leq y) \cap D} p(x)$$

A esta función de distribución le corresponderá una v.a. X discreta de función de cuantía $p(x)$

$$p(x) = F(x) - F(x^-) \quad \forall x \in D$$

Ejemplos

- 3.- Sea una urna con 3 bolas blancas, 2 negras y 1 verde. Se extraen 3 bolas al azar y se considera $X(w) =$ número de bolas blancas extraídas. Construir la función de probabilidad inducida y la función de distribución.
- 4.- Cinco personas lanzan simultáneamente monedas para determinar quién ha de comprar los refrescos para todos. El sistema es el siguiente: el primero que obtenga un resultado (ya sea cara o cruz) distinto de cada uno de los resultados obtenidos por los demás, debe pagar los refrescos de todos. Sea X el número de ensayos requeridos para concluir el juego.
 - a) Determinar su función de masa de probabilidad.
 - b) Determinar su función de distribución.
 - c) Calcular $P(X > 2)$, $P(X < 2)$ y $P(3 \leq X \leq 6)$.
 - d) Calcular $P(X > 5/X > 2)$ y $P(X > 2/X > 5)$.

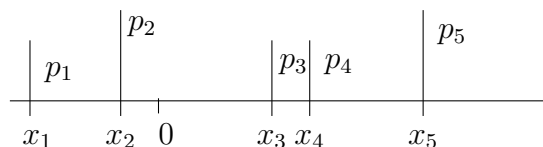


Figura 3.1: Función de probabilidad

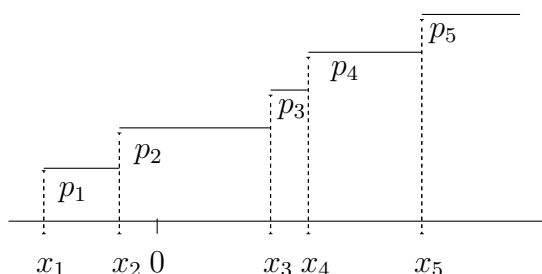


Figura 3.2: Función de distribución

Ejemplo 5: La función de probabilidad representada en la figura 3.1 anterior corresponde a los valores:

x_k	-3	-1	2	3	6
p_k	1/6	1/4	1/6	1/6	1/4

La función de distribución correspondiente al ejemplo anterior representada en la figura 3.2 se expresa

$$F(x) = \begin{cases} 0 & \text{si } x < -3 \\ 1/6 & \text{si } -3 \leq x < -1 \\ 5/12 & \text{si } -1 \leq x < 2 \\ 7/12 & \text{si } 2 \leq x < 3 \\ 9/12 & \text{si } 3 \leq x < 6 \\ 1 & \text{si } x \geq 6 \end{cases}$$

Las diferencias entre cada par de valores consecutivos son los valores de la función de probabilidad.

Se trata de una función con saltos de magnitud p_k , en cada uno de los valores x_k ; creciente de 0 a 1 y continua por la derecha. (Sería continua por la izquierda si se hubiese definido $F(x) = P\{X < x\}$).

3.5.3. Distribuciones absolutamente continuas. Función de densidad sobre \mathbb{R} .

Una función $f : \mathbb{R} \mapsto \mathbb{R}$ se llama función de densidad sobre \mathbb{R} si cumple:

1. $f(x) \geq 0, \forall x \in \mathbb{R}$.
2. f admite a lo más un número finito de discontinuidades sobre cada intervalo de \mathbb{R} , es decir, f es integrable Riemann.
3. $\int_{-\infty}^{\infty} f(x)dx = 1$, donde esta integral se entiende en el sentido Riemann.

Proposición 3.8. Si una función $f : \mathbb{R} \mapsto \mathbb{R}$ es función de densidad

$$F(x) = \int_{-\infty}^x f(t)dt \text{ para cada } x \in \mathbb{R} \quad (3.4)$$

F es función de distribución y, en todo punto x en que f sea continua, F es derivable y $F'(x) = f(x)$.

En efecto, si f cumple las condiciones impuestas, la función F definida por 3.4 cumple, cualquiera que sean $a < b \in \mathbb{R}$,

$$F(b) - F(a) = \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx = \int_a^b f(x)dx.$$

Desde luego F es creciente, puesto que $F(b) - F(a) \geq 0$ siempre que $a < b$. También es continua en cualquier punto $x \in \mathbb{R}$, pues si f está acotada por K en un entorno de x , es

$$F(x+h) - F(x) = \int_x^{x+h} f(t)dt \leq Kh, \quad F(x) - F(x-h) = \int_{x-h}^x f(t)dt \leq Kh$$

cuando h es suficientemente pequeño; en caso contrario, la propia definición de las integrales impropias

$$\int_a^x f(t)dt = \lim_{h \rightarrow 0} \int_a^{x-h} f(t)dt \text{ y } \int_x^b f(t)dt = \lim_{h \rightarrow 0} \int_{x+h}^b f(t)dt$$

muestra que F es continua en x , por la izquierda y por la derecha. Además

$$F(-\infty) = \lim_{x \rightarrow -\infty} \int_{-\infty}^x f(t)dt = 0 \text{ y } F(\infty) = \lim_{x \rightarrow +\infty} \int_{-\infty}^x f(t)dt = 1$$

En definitiva, F es una función de distribución continua y f es densidad de F , de forma que F es absolutamente continua.

La observación final constituye el teorema fundamental del cálculo.

Definición 3.7. Una variable aleatoria X se dice continua ó absolutamente continua si su función de distribución F_X puede ser representada para cada $x \in \mathbb{R}$ por:

$$F_X(x) = \int_{-\infty}^x f_X(t)dt$$

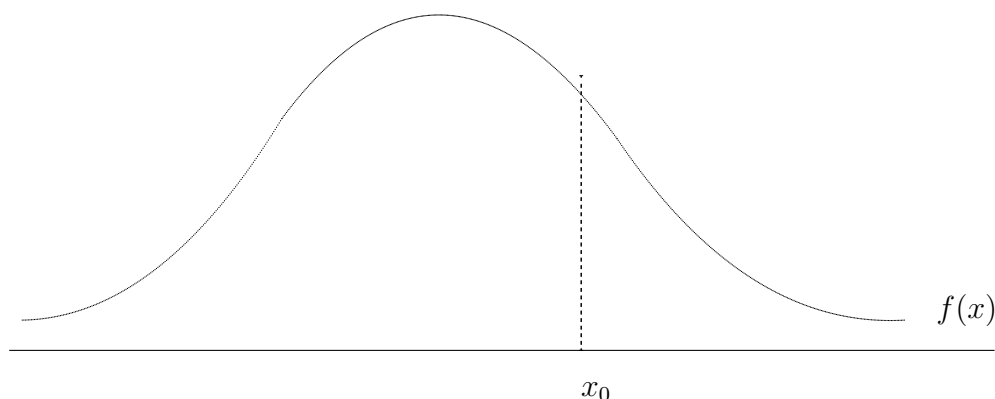
donde f_X es una función de densidad sobre \mathbb{R} . A esta función se la llama función de densidad de la variable aleatoria continua X , y al conjunto

$$C_X = \{x \in \mathbb{R} | f_X(x) > 0\}$$

soporte de la variable aleatoria. En la práctica, el soporte corresponde al conjunto de valores observables de la variable, y este conjunto es no numerable como veremos más adelante.

Suponiendo dada la función de densidad de una v.a., la función de distribución en un punto se obtiene sin más que calcular el área que hay debajo de la función de densidad a la izquierda de ese punto.

$$F_X(x_0) = \int_{-\infty}^{x_0} f(t)dt.$$



Teorema 3.1. Sea X una v.a. continua con función de densidad f_X y función de distribución F_X . Entonces se verifica que

1. F_X es continua.
2. Si f_X es continua en un punto x , F_X es derivable en ese punto y

$$F'_X(x) = f_X(x)$$

3. $D_X = \{x \in \mathbb{R} | p_X(x) > 0\} = \emptyset$ pues $p_X(x)$ representa el salto en x de F_X
4. Para todo intervalo $< a, b >$, donde por $< a, b >$ denotamos los intervalos de la forma $();$; $[];$ $();$ $]$, es

$$P\{X \in < a, b >\} = \int_a^b f(t)dt = F(b) - F(a)$$

Demostración:

3. El conjunto $D_X = \{x \in \mathbb{R} | p_X > 0\}$ sabemos que coincide con el conjunto de puntos de discontinuidad de F_X , pero como hemos visto, F_X es continua $\forall x \in \mathbb{R} \Rightarrow D_X = \emptyset$.
4. Al ser $D_X = \emptyset$ resultará que $P\{X \in < a, b >\}$, cualesquiera que sean los límites del intervalo (abierto o cerrado) no dependerá de $p_X(x)$ que vale cero $\forall x \in \mathbb{R}$, con lo que

$$P\{X \in < a, b >\} = F(b) - F(a) = \int_a^b f(t)dt$$

La interpretación de la función de densidad es la siguiente:

Sea $x \in \mathbb{R}$ un punto de continuidad de f_X . Entonces,

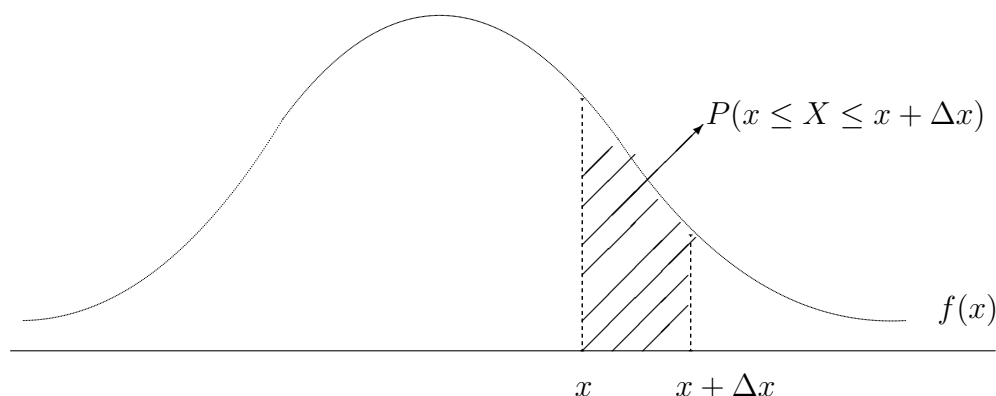
$$P\{x \leq X \leq x + \Delta x\} = \int_x^{x+\Delta x} f_X(t)dt = f_X(\xi)\Delta x, \text{ con } \xi \in [x, x + \Delta x].$$

Como $P\{x \leq X \leq x + dx\}$ representa una masa, si lo dividimos por Δx una longitud, obtenemos una densidad f_X . La densidad puntual $f(x)$ proporciona la masa de segmentos de longitud infinitesimal alrededor del punto x :

$$\text{masa de } (dx) \simeq f(x)dx;$$

puede expresarse simbólicamente

$$P_F(dx) \simeq f(x)dx \quad \text{o} \quad dF(x) \simeq f(x)dx.$$



Ejemplo 6: Sea la función

$$f_X(x) = \begin{cases} 3x^2/2 & \text{si } -1 \leq x \leq 1 \\ 0 & \text{resto} \end{cases}$$

Veamos que esta función es una función de densidad:

- a) $f_X(x) \geq 0, \forall x \in \mathbb{R}$.
- b) Posee solamente dos puntos de discontinuidad, el $x = 1$ y el $x = -1$.
- c) $\int_{-\infty}^{+\infty} f_X(x) dx = \int_{-1}^1 3x^2/2 dx = [x^3/2]_{-1}^1 = 1$.

Con lo cual f_X será una función de densidad. Su función de distribución correspondiente será:

$$\begin{aligned} \forall x < -1, \quad F_X(x) &= 0 \\ \forall -1 \leq x \leq 1, \quad F_X(x) &= \int_{-1}^x \frac{3t^2}{2} dt = \frac{x^3 + 1}{2} \\ \forall x > 1, \quad F_X(x) &= \int_{-\infty}^{-1} f_X(t) dt + \int_{-1}^1 f_X(t) dt + \int_1^x f_X(t) dt = 0 + 1 + 0 = 1 \end{aligned}$$

Así pues, es:

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq -1 \\ \frac{x^3 + 1}{2} & \text{si } -1 \leq x \leq 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

F_X continua y derivable salvo en $x = -1$ y $x = 1$. Por supuesto, también existen variables singulares, mixtas, etc.

En el caso de variables aleatorias de tipo mixto - las más generales que suelen presentarse en las aplicaciones - X puede tomar algunos valores con probabilidad no nula; concretamente

$$P\{X = x\} = F_X(x) - F_X(x^-) > 0$$

para todo x en que F_X presente un salto.

Ejemplo 7: La función de distribución que tiene por expresión

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x^2/16 & \text{si } 0 \leq x < 2 \\ 1/4 & \text{si } 2 \leq x < 4 \\ x/4 - 5/8 & \text{si } 4 \leq x < 5 \\ 1 - 5/(4x) & \text{si } 5 \leq x \end{cases}$$

Presenta discontinuidades de salto en los puntos 4 y 5, de tamaño $1/8$, en ambos casos; mientras que crece de manera continua en los intervalos $(0, 2)$, $(4, 5)$ y $(5, +\infty)$ y permanece constante en los intervalos $(-\infty, 0)$ y $(2, 4)$.

Ejemplo 8: La distribución de duración, en minutos, de las conversaciones telefónicas en un locutorio es, en principio, exponencial de parámetro $1/5$. Sin embargo, como los pasos del contador se producen cada tres minutos, una cuarta parte de los usuarios prolonga su llamada hasta el siguiente paso del contador. Determinar la función de distribución de la duración real de las llamadas; su parte discreta y su parte continua.

La duración original de las conversaciones tiene función de distribución

$$F(x) = 1 - e^{-x/5} \text{ para } x \geq 0.$$

Distribución absolutamente continua que proporciona la probabilidad de que una conversación acabe antes de x minutos.

Las tres cuartas partes de los usuarios respetan esta distribución, pero una cuarta parte de ellos la modifica, acumulando en el punto $3k$ la probabilidad de todo el intervalo $(3k - 3, 3k]$. Así pues, sustituyen F por una distribución discreta \bar{F} , tal que

$$P_{\bar{F}}(\{3k\}) = P_F((3k - 3, 3k]) = F(3k) - F(3k - 3) = e^{-3(k-1)/5}(1 - e^{-3/5})$$

para cada $k = 1, 2, 3, \dots$ será pues, para todo $x \geq 0$,

$$\bar{F}(x) = (1 - e^{-3/5}) \sum_{\{k \in \mathbb{N} / 3k \leq x\}} e^{-3(k-1)/5} = 1 - e^{-3[x/3]/5}$$

después de sumar $[x/3]$ primeros términos de la progresión geométrica de razón $e^{-3/5}$. La distribución real de la duración de las llamadas es, por tanto,

$$G(x) = \frac{3}{4}F(x) + \frac{1}{4}\bar{F}(x) = \frac{3}{4}(1 - e^{-x/5}) + \frac{1}{4}(1 - e^{-3[x/3]/5})$$

mixtura de la distribución absolutamente continua F y de la distribución discreta \bar{F} .

3.6. Cambio de variable en las distribuciones unidimensionales.

Sea X una variable aleatoria con función de distribución F_X , concentrada en un intervalo I (no hay pérdida de generalidad puesto que puede ser $I = \mathbb{R}$). Con frecuencia se plantea el problema de determinar la distribución de la variable aleatoria $Y = g(X)$, donde g es una función medible de I en \mathbb{R} . La solución teórica es sencilla. Basta observar que la función de distribución de Y se expresa

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = P\{X \in g^{-1}(-\infty, y]\} \quad (3.5)$$

resultado que puede determinarse a partir de F_X puesto que $g^{-1}(-\infty, y]$ es un subconjunto de Borel de I . Sin embargo, las dificultades prácticas provienen de la determinación de dicho conjunto que ha de hacerse con cuidado, incluso en situaciones simples como la siguiente.

Ejemplo 9: Si $g(x) = ax + b$ es una función lineal. La distribución de $Y = aX + b$ vale:

$$\triangleright \text{ Si } a > 0, F_Y(y) = P\{aX + b \leq y\} = P\left\{X \leq \frac{y-b}{a}\right\} = F_X\left(\frac{y-b}{a}\right)$$

$$\triangleright \text{ Si } a < 0, F_Y(y) = P\{aX + b \leq y\} = P\left\{X \geq \frac{y-b}{a}\right\} = 1 - F_X\left(\left(\frac{y-b}{a}\right)^-\right).$$

A medida que g se complica, las dificultades para obtener $g^{-1}(-\infty, y]$ crecen. Siempre conviene tener presente la gráfica de g y, en particular, su recorrido.

Ejemplo 10:

La función $g(x) = x^4 - 2x^2$ toma valores en el intervalo $[-1, \infty)$ como podría observarse si obtenemos su gráfica. Para $y \in (-1, 0)$ la ecuación $x^4 - 2x^2 = y$ tiene cuatro raíces:

$$\pm\sqrt{1 + \sqrt{1+y}} \quad \text{y} \quad \pm\sqrt{1 - \sqrt{1+y}}$$

las dos segundas comprendidas entre las dos primeras. En cambio, para $y > 0$, la ecuación sólo tiene las dos primeras raíces. Si graficamos la función g observamos:

▷ Si $y < -1$ es $g^{-1}(-\infty, y] = \emptyset$.

▷ Si $-1 \leq y < 0$ $g^{-1}(-\infty, y]$ coincide con

$$\left[-\sqrt{1+\sqrt{1+y}}, -\sqrt{1-\sqrt{1+y}}\right] \cup \left[\sqrt{1-\sqrt{1+y}}, \sqrt{1+\sqrt{1+y}}\right]$$

▷ Si $y \geq 0$, se tiene $g^{-1}(-\infty, y] = [-\sqrt{1+\sqrt{1+y}}, \sqrt{1+\sqrt{1+y}}]$. De forma que, si $Y = g(X)$ y X tiene función de distribución F_X , resulta

$$F_Y(y) = \begin{cases} 0 & \text{si } y < -1 \\ F_X(-\sqrt{1-\sqrt{1+y}}) - F_X((-\sqrt{1+\sqrt{1+y}})^-) \\ + F_X(\sqrt{1+\sqrt{1+y}}) - F_X((\sqrt{1-\sqrt{1+y}})^-) & \text{si } -1 \leq y < 0 \\ F_X(\sqrt{1+\sqrt{1+y}}) - F_X((-\sqrt{1+\sqrt{1+y}})^-) & \text{si } y \geq 0 \end{cases}$$

Por ejemplo, si $F_X(x) = x^2/4$ para $0 \leq x \leq 2$, la y varía entre -1 y 8, quedando

$$F_Y(y) = \begin{cases} 0 & \text{si } y < -1 \\ \frac{\sqrt{1+y}}{2} & \text{si } -1 \leq y \leq 0 \\ \frac{1+\sqrt{1+y}}{4} & \text{si } 0 \leq y \leq 8 \end{cases}$$

En los casos anteriores, $g^{-1}(\infty, y]$ es una unión de intervalos disjuntos. Cuando la función g sea muy irregular, el resultado puede ser un conjunto de Borel más complicado y habría entonces la dificultad añadida de calcular su probabilidad a partir de F_X . Afortunadamente, en la práctica es raro que haya que hacer cambios de variable raros. Por el contrario, el problema se simplifica cuando la variable original es discreta o absolutamente continua.

3.6.1. Cambios de variable discreta.

Cuando X es una variable aleatoria discreta, concentrada en el conjunto numerable D , con función de probabilidad $p(x)$, cualquier transformación $Y = g(X)$ da lugar a una variable aleatoria discreta, concentrada en el conjunto $g(D)$.

Es posible, entonces, evitar el cálculo de la función de distribución de Y y obtener directamente su función de probabilidad: Para cada $y \in g(D)$,

$$P\{Y = y\} = P\{g(X) = y\} = P\{X \in g^{-1}(y)\} = \sum_{x \in D \cap g^{-1}(y)} p(x) \quad (3.6)$$

Ejemplo 11:

Sea X una variable aleatoria con distribución geométrica de parámetro p :

$$P\{X = n\} = p(1 - p)^{n-1} \text{ para } n = 1, 2, 3, \dots$$

Si $a \in \mathbb{N}$, la variable $Y = (X - a)^2$ toma valores en el conjunto de los cuadrados perfectos y es

$$P\{Y = m^2\} = \begin{cases} p(1 - p)^{a-1} & \text{si } m = 0 \\ p(1 - p)^{a+m-1} + p(1 - p)^{a-m-1} & \text{si } 0 < m < a \\ p(1 - p)^{a+m-1} & \text{si } m \geq a \end{cases}$$

Ejemplo 12:

Se considera la variable aleatoria X = suma de los puntos obtenidos al lanzar 2 dados, y la variable aleatoria Y ganancia de un jugador dada por

$$Y = \begin{cases} 5 & \text{si } X = 7, 11 \\ -3 & \text{si } X = 3, 12 \\ 0 & \text{resto} \end{cases}$$

Calcular la función de probabilidad de Y .

3.6.2. Cambios de variable absolutamente continuos.

Sea X una variable absolutamente continua, con función de densidad f , concentrada en el intervalo I (es decir, $f(x) = 0$ si $x \notin I$).

Como primera posibilidad, la función $g : I \mapsto \mathbb{R}$ puede ser escalonada y tomar sólo un número numerable de valores, que componen su imagen $g(I)$. Entonces la variable aleatoria $Y = g(X)$ es discreta, con función de probabilidad

$$p(y) = \int_{g^{-1}(y)} f(x) dx \quad \text{para cada } y \in g(I) \quad (3.7)$$

De hecho $g^{-1}(y)$ es la unión, a lo sumo numerable, de todos los intervalos en que g toma el valor y . Por tanto

$$P\{Y = y\} = P\{X \in g^{-1}(y)\} = \int_{g^{-1}(y)} f(x) dx$$

Nótese que $\bigcup_{y \in g(I)} g^{-1}(y) = I$, así que Y está concentrada en $g(I)$ pues

$$\sum_{y \in g(I)} p(y) = \sum_{y \in g(I)} \int_{g^{-1}(y)} f(x) dx = \int_I f(x) dx = 1.$$

Ejemplo 13:

Se elige X con distribución exponencial de densidad $f(x) = e^{-x}$ en $x > 0$.

La parte entera de X , $Y = [X]$, es una variable aleatoria que toma cada valor entero $y \geq 0$ cuando $X \in [y, y + 1)$. Así pues, la función de probabilidad de Y asigna a cada $y \in \mathbb{N} \cup \{0\}$, probabilidad

$$p(y) = P\{[X] = y\} = P\{y \leq X < y + 1\} = \int_y^{y+1} e^{-x} dx = e^{-y}(1 - e^{-1})$$

y se cumple $\sum_{y=0}^{\infty} p(y) = 1$. Así mismo, la variable aleatoria

$$Z = I_{\{X - [X] > 1/2\}} - I_{\{X - [X] < 1/2\}}$$

toma los valores -1, 0 y 1 e indica si la parte decimal de X es menor, igual o mayor que 1/2. Su función de probabilidad será

$$\begin{aligned} P\{Z = -1\} &= P\left\{X \in \bigcup_{y=0}^{\infty} [y, y + 1/2)\right\} = \sum_{y=0}^{\infty} P\{X \in [y, y + 1/2)\} \\ &= \sum_{y=0}^{\infty} \int_y^{y+1/2} e^{-x} dx = (1 - e^{-1/2}) \sum_{y=0}^{\infty} e^{-y} = \frac{1}{1 + e^{-1/2}} \simeq 0,622 \\ P\{Z = 1\} &= P\left\{X \in \bigcup_{y=0}^{\infty} (y + 1/2, y + 1)\right\} = \sum_{y=0}^{\infty} P\{X \in (y + 1/2, y + 1)\} \\ &= \sum_{y=0}^{\infty} \int_{y+1/2}^{y+1} e^{-x} dx = (1 - e^{-1/2}) \sum_{y=0}^{\infty} e^{-y-1/2} = \frac{1}{1 + e^{1/2}} \simeq 0,378. \end{aligned}$$

Naturalmente

$$P\{Z = 0\} = P\left\{X \in \bigcup_{y=0}^{\infty} \{y + 1/2\}\right\} = \sum_{y=1}^{\infty} P\{X = y + 1/2\} = 0$$

puesto que X es continua. Así pues, Z está concentrada en $\{-1, 1\}$.

Examinemos ahora la situación en que g es una función continuamente derivable. El caso más simple es que g sea estrictamente creciente o decreciente en I (y, por tanto, inyectiva), ya que se verifica:

Proposición 3.9. *Sea X una variable aleatoria absolutamente continua, con función de densidad f , concentrada en I . Si $g : I \mapsto \mathbb{R}$ es una función con derivada continua, estrictamente monótona, entonces $Y = g(X)$ es una variable aleatoria absolutamente continua, cuya densidad en $y \in g(I)$ vale*

$$f_Y(y) = f(g^{-1}(y))|(g^{-1})'(y)| = \frac{f(g^{-1}(y))}{|g'(g^{-1}(y))|} \quad (3.8)$$

En efecto, para cualesquiera $a < b \in g(I)$, si g es estrictamente creciente es

$$\begin{aligned} P\{a < Y \leq b\} &= P\{g^{-1}(a) < X \leq g^{-1}(b)\} = \int_{g^{-1}(a)}^{g^{-1}(b)} f(x)dx \\ &= \int_a^b f(g^{-1}(y))(g^{-1})'(y)dy \end{aligned}$$

después de hacer en la integral el cambio de variable $y = g(x)$ la variable aleatoria Y es absolutamente continua, con densidad $f(g^{-1}(y))(g^{-1})'(y)$. Además g^{-1} es estrictamente creciente y derivable, con lo cual $(g^{-1})'(y) > 0$.

Análogamente, si g es estrictamente decreciente, será

$$\begin{aligned} P\{a < Y \leq b\} &= P\{g^{-1}(b) \leq X \leq g^{-1}(a)\} = \int_{g^{-1}(b)}^{g^{-1}(a)} f(x)dx \\ &= \int_a^b f(g^{-1}(y))(g^{-1})'(y)dy \end{aligned}$$

Por tanto, la densidad de Y es $-f(g^{-1}(y))(g^{-1})'(y) = f(g^{-1}(y))|(g^{-1})'(y)|$, puesto que g^{-1} , derivable y decreciente, tiene derivada $(g^{-1})'(y) < 0$. Recuérdese, además, que $(g^{-1})'(y) = 1/g'(g^{-1}(y))$.

Ejemplo 14:

Sea X una variable continua con soporte $I = (0, 3)$ y función de densidad

$$f_X(x) = \begin{cases} \frac{x^2}{9} & 0 \leq x < 3 \\ 0 & \text{resto} \end{cases}$$

Se considera $Y = g(X) = \begin{cases} -1 & \text{si } X < 1 \\ 0 & \text{si } X = 1 \\ 1 & \text{si } X > 1 \end{cases}$ Hallar la distribución de probabilidad de la variable Y .

En realidad la proposición anterior indica que, si $y = g(x)$, la probabilidad de que X se encuentre en los alrededores de x es transferida por g a la probabilidad de que Y se encuentre en los alrededores de y ; esto es

$$f_Y(y)dy = f(x)dx$$

Pero $dy = |g'(x)|dx$ expresa como se contraen o dilatan los entornos de x al transformarse por g en entornos de y ; así que la concentración de la probabilidad alrededor de y , $f_Y(y)$, resulta de la concentración $f(x)$ alrededor de x , corregida por la tasa de contracción o dilatación: $dx/dy = 1/|g'(x)|$.

¿Qué ocurre cuando la transformación continuamente derivable, g , no es inyectiva? En tal caso, cada valor de $Y = g(X)$ puede ser transformado de diversos valores de X , de forma que en los alrededores del valor y se acumulan las probabilidades que provienen de los alrededores de cada valor x comprendido en $g^{-1}(y)$. El tamaño de dicho conjunto puede variar, según el valor de y , desde un único punto hasta todo un intervalo, lo cual supone diversos comportamientos posibles.

En muchos casos, el conjunto $\{x \in I / g'(x) = 0\}$ es numerable (de forma que g no permanece constante en ningún intervalo) y, por tanto, I se descompone en una sucesión de intervalos en cada uno de los cuales g es monótona y se puede aplicar el resultado 3.8. Entonces, Y es absolutamente continua con densidad en $y \in g(I)$:

$$f_Y(y) = \sum_{x \in g^{-1}(y)} \frac{f(x)}{|g'(x)|} \quad (3.9)$$

Teorema 3.2. Sea X una variable aleatoria continua con soporte C_X y función de densidad f_X . Sea $\varphi : \mathbb{R} \mapsto \mathbb{R}$ una función derivable $\forall x \in C_X$ tal que φ' es continua y $\varphi'(x) \neq 0$ salvo un número finito de puntos. Suponemos que para cada $y \in \mathbb{R}$ se cumple:

1. Existen exactamente $m(y)$ puntos ($m(y) \geq 1$); $x_1(y), x_2(y), \dots, x_{m(y)}(y)$ de C_X tales que para cada $k = 1, 2, \dots, m(y)$ es

$$\varphi(x_k(y)) = y \text{ y } \varphi'(x_k(y)) \neq 0$$

o bien

2. Si no existe ningún punto $x \in C_X$ tal que $\varphi(x) = y$ y $\varphi'(x) \neq 0$, ponemos en este caso $m(y) = 0$.

Entonces, $\varphi(X) = Y$ es una variable aleatoria continua cuya función de densidad viene dada por

$$f_Y(y) = \begin{cases} \sum_{k=1}^{m(y)} f_X(x_k(y)) \cdot |\varphi'(x_k(y))|^{-1} & \text{si } m(y) > 0 \\ 0 & \text{si } m(y) = 0. \end{cases}$$

Ejemplos:

15.- Sea X una variable aleatoria continua con soporte $I = (0, 3)$ y función de densidad

$$f(x) = \begin{cases} \frac{x^2}{9} & \text{si } 0 < X < 3 \\ 0 & \text{resto} \end{cases}$$

Sea $Y = g(X) = X^2$ Calcular la distribución de Y .

16.- Sea $(\mathbb{R}, \mathbb{B}, P)$ con P dada por la función de distribución:

$$F(x) = \begin{cases} 0 & x < -3 \\ \frac{x+3}{4} & -3 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Sean, asimismo, $g_1(X) = X^2$, $g_2(X) = X^3$, $g_3(X) = e^{-X}$ tres variables aleatorias. Calcular las funciones de distribución inducidas por cada una de ellas.

17.- Sea X una variable aleatoria con función de densidad

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & \text{resto} \end{cases}$$

y sea

$$g(X) = \begin{cases} X^2 & \text{si } 0 \leq X < 2 \\ 4 & \text{si } 2 \leq X < 3 \\ -4(X - 4) & \text{si } 3 \leq X < 4 \\ 0 & \text{resto} \end{cases}$$

Calcular la distribución de la variable aleatoria $g(X)$.

Capítulo 4

Distribuciones Unidimensionales

4.1. Distribuciones Unidimensionales.

4.1.1. Introducción

Cuando estudiábamos la Estadística Descriptiva, nos ocupábamos de la información correspondiente a una variable estadística, obteníamos su correspondiente distribución de frecuencias, y una serie de medidas descriptivas tales como la media, mediana, moda, cuantiles, varianza, etc., que nos resumían la información sobre la variable estadística en cuestión, pudiendo establecer comparaciones entre diferentes poblaciones, utilizando únicamente sus características descriptivas.

Una situación análoga nos encontramos ahora, pues toda la información sobre una variable aleatoria X está especificada en su función de probabilidad $p(x)$ o en su función de densidad $f(x)$, según que la variable aleatoria sea discreta o continua, respectivamente. Pero en muchas situaciones hay que realizar comparaciones entre dos o más distribuciones de probabilidad y éstas son más fáciles de realizar mediante los valores característicos de las variables aleatorias correspondientes a esas distribuciones, que directamente con las diferentes distribuciones de probabilidad o las correspondientes funciones de densidad.

Siguiendo un cierto paralelismo con la Estadística Descriptiva estudiaremos en este capítulo, tanto para variables discretas como para variables continuas, algunas características de las mismas: media, varianza, mediana, etc., y sus relaciones más importantes.

4.2. Esperanza Matemática.

El concepto de valor esperado o de esperanza matemática es fundamental en la teoría de la probabilidad. Como la misma probabilidad, surgió en el contexto de los juegos de azar, cuando Fermat y Pascal, hacia 1650, se proponían uno a otro problemas del siguiente tipo:

1. Se lanza un dado.
2. Si aparece un 1 el jugador gana 10 euros.
3. Si aparece 2, 3, 4 ó 5 el jugador pierde 3 euros.
4. Si aparece un 6 el jugador gana 2 euros.

Queremos saber si el juego es equitativo o justo, es decir, si en una larga serie de jugadas un jugador no gana más que el otro.

Para responder podemos calcular la equidad del juego de la siguiente forma: Sea X : cantidad que gana un jugador en el juego, siendo negativa cuando pierde.

Esta variable aleatoria, X , puede tomar los siguientes valores:

$$\begin{aligned} X = 10, & \quad \frac{1}{6} \text{ de las veces} \\ X = -3, & \quad \frac{2}{3} \text{ de las veces} \\ X = 2, & \quad \frac{1}{6} \text{ de las veces} \end{aligned}$$

Si calculamos la media ponderada de los posibles resultados utilizando como ponderación la proporción de veces que se presenta un resultado particular, tenemos que el resultado ponderado será:

$$10 \cdot \frac{1}{6} + (-3) \cdot \frac{2}{3} + 2 \cdot \frac{1}{6} = 0$$

y según este resultado podemos decir que no esperamos ni ganar ni perder en este juego, es decir que la ganancia esperada por juego es cero.

El promedio de las ganancias de un jugador, ponderadas por la probabilidad con que puede lograrlas, se denomina el valor esperado que tiene el juego para él, ó la esperanza matemática de su ganancia. De hecho, es lo que espera ganar en media, por partida, si el juego se repite indefinidamente, pues, según la ley empírica del capítulo 1, cada ganancia se obtendrá, a la larga, con una frecuencia igual a su probabilidad.

El ejemplo anterior establece el siguiente criterio de uso frecuente: Un juego es favorable, desfavorable, o equitativo para un jugador según que su valor esperado sea positivo, negativo o nulo.

Más concretamente y considerando el juego planteado anteriormente, que da lugar a una variable aleatoria discreta, X , cuya función de probabilidad es:

$$P(x_i) = P(X = x_i) = \begin{cases} \frac{1}{6} & \text{si } x_i = 10 \\ \frac{2}{3} & \text{si } x_i = -3 \\ \frac{1}{6} & \text{si } x_i = 2 \\ 0 & \text{en el resto} \end{cases}$$

y si designamos por $E[X]$ la ganancia esperada tenemos que:

$$E[X] = 10 \cdot \frac{1}{6} + (-3) \cdot \frac{2}{3} + 2 \cdot \frac{1}{6} = 0$$

Definición 4.1. Si X es una variable aleatoria definida en un espacio de probabilidad discreto $(\Omega, \mathcal{P}(\Omega), P)$, el valor esperado o esperanza matemática de X es

$$E[X] = \sum_{w \in \Omega} X(w)P(w) \quad (4.1)$$

supuesto que $\sum_{\{w|X(w)>0\}} X(w)P(w) < \infty$ y $\sum_{\{w|X(w)<0\}} X(w)P(w) > -\infty$.

Desde luego X será una variable aleatoria discreta, que toma valores en $X(\Omega) = \{x_1, x_2, \dots, x_k, \dots\}$. Su valor esperado, si existe, se puede expresar

$$E[X] = \sum_k \sum_{w \in X^{-1}(x_k)} X(w)P(w) = \sum_k x_k \sum_{w \in X^{-1}(x_k)} P(w) = \sum_k x_k P\{X = x_k\} \quad (4.2)$$

con $\sum_{k/x_k > 0} x_k P(X = x_k) < +\infty$ y $\sum_{k/x_k < 0} x_k P(X = x_k) > -\infty$.

El interés debe centrarse ahora en variables aleatorias de tipo más general, definidas en un espacio de probabilidad arbitrario (Ω, \mathcal{F}, P) y cuya distribución podría ser absolutamente continua, mixta

- Para empezar, supongamos que X es una variable aleatoria absolutamente continua, con función de densidad f ; de modo que X toma valores en cada intervalo

$(x, x + dx)$ con probabilidad $f(x) dx$. En tales circunstancias, es lógico definir, como valor esperado de X , el promedio análogo a 4.2:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx \quad (4.3)$$

en el supuesto de que $\int_0^{\infty} x f(x) dx < \infty$ y $\int_{-\infty}^0 x f(x) dx > -\infty$.

La esperanza matemática concuerda en gran parte con el concepto físico de centro de gravedad. De hecho, la suma $\sum_k x_k m_k / M$ define el centro de gravedad de un conjunto de partículas situadas en los puntos x_k de una recta y de masas m_k , cuya suma es M . Cuando en vez de un conjunto de partículas, se dispone de una varilla de longitud L que tiene densidad de masa $\rho(x)$, en cada punto x , el centro de gravedad está situado en el punto $\int_0^L x \rho(x) dx / M$, donde M es también la masa total de la varilla.

- El siguiente paso debe generalizar el concepto de esperanza matemática al caso de variables aleatorias con distribución mixta: $F = \alpha F_1 + (1 - \alpha) F_2$, donde F_1 es una distribución discreta, concentrada en $D = \{x_i\}_{i \in I}$, con función de probabilidad p_1 , y F_2 es absolutamente continua, con función de densidad f_2 . Así, $p = \alpha p_1$ y $f = (1 - \alpha) f_2$ designarán respectivamente la función de probabilidad y la función de densidad asociadas a F .

La definición del valor esperado de X debe englobar, ahora, las dos situaciones previas; de manera que se define

$$E[X] = \sum_{i \in I} x_i p(x_i) + \int_{-\infty}^{\infty} x f(x) dx \quad (4.4)$$

siempre que no aparezcan los valores $+\infty$ y $-\infty$ al calcular:

$$\sum_{\{i/x_i > 0\}} x_i p(x_i), \int_0^{\infty} x f(x) dx, \sum_{\{i/x_i < 0\}} x_i p(x_i) \text{ y } \int_{-\infty}^0 x f(x) dx.$$

Con el convenio de escribir el segundo miembro de 4.4 en la forma $\int_{-\infty}^{\infty} x F(dx)$, los tres casos anteriores se integran en una definición única.

Definición 4.2. *El valor esperado o esperanza matemática de una variable aleatoria X , con función de distribución F , es*

$$E[X] = \int_{-\infty}^{\infty} x F(dx) \quad (4.5)$$

supuesto que $\int_0^{\infty} x F(dx) < \infty$ y $\int_{-\infty}^0 x F(dx) > -\infty$.

En relación con la definición anterior debe observarse:

- (a) Cada variable aleatoria X puede expresarse: $X = X^+ - X^-$, donde

$$X^+ = X I_{\{X \geq 0\}} = \max(0, X) \text{ y } X^- = -X I_{\{X \leq 0\}} = -\min(0, X)$$

son ambas positivas. El cálculo de $E[X]$ equivale a calcular primero $E[X^+]$ y $E[X^-]$, para formar después $E[X] = E[X^+] - E[X^-]$, salvo en el caso en que tengan valor infinito. Se consigue así evitar casos de convergencia condicional en los cuales la suma $\sum_k x_k p_k$ depende del orden en que se introducen los términos, o casos en que el valor de $\int_{-\infty}^{\infty} x f(x) dx$ depende de cómo tienden los límites de la integral hacia $\pm\infty$

- (b) La esperanza matemática de una variable aleatoria X no es una función de X sino un número fijo y una propiedad de la distribución de probabilidad de la variable aleatoria. Por otra parte, el valor esperado puede no existir dependiendo de si la correspondiente suma o integral no converge a un valor finito. Además, el valor esperado no tiene por qué coincidir con alguno de los valores que puede tomar la variable.

Ejemplo 1:

En el juego del parchís un jugador no puede iniciar el avance con una ficha hasta que obtiene un cinco. El número N de veces que es necesario lanzar un dado hasta que aparece el primer cinco es una variable discreta. Sus valores posibles y probabilidades asociadas son los de la tabla siguiente. Como vemos esta variable aleatoria puede tomar un número de valores infinito numerable.

Valores	1	2	3	...	k	...
Probabilidades	$(\frac{1}{6})$	$(\frac{5}{6})(\frac{1}{6})$	$(\frac{5}{6})^2(\frac{1}{6})$...	$(\frac{5}{6})^{(k-1)}\frac{1}{6}$...

La tabla muestra el número de lanzamientos hasta obtener el primer cinco.

El valor esperado de la variable aleatoria definida en el juego del parchís es:

$$E[X] = \sum_{k=1}^{\infty} k P(N = k) = \frac{1}{6} \frac{1}{(1 - \frac{5}{6})^2} = 6$$

Ejemplo 2:

Sea una variable aleatoria X , con función de densidad

$$f(x) = \begin{cases} \frac{1}{x^2}, & x \geq 1 \\ 0, & \text{en el resto} \end{cases}$$

Obtener, si existe, la esperanza matemática de X .

Aplicando la definición de valor esperado para el caso continuo, tenemos:

$$\begin{aligned} E[X] &= \int_1^{\infty} x \cdot \frac{1}{x^2} dx = \int_1^{\infty} \frac{1}{x} dx = [\ln x]_1^{\infty} = \lim_{x \rightarrow \infty} \ln(x) - \ln 1 = \\ &= \lim_{x \rightarrow \infty} \ln(x) - 0 = \infty \end{aligned}$$

Luego no existe el valor esperado $E[X]$.

Ejemplo 3:

La distribución de la duración de las llamadas telefónicas, propuesta en el ejemplo 4 del capítulo anterior, tiene densidad $f(x) = e^{-x/5} 3/20$ en $(0, \infty)$ ($\int_0^{\infty} e^{-x/5} \frac{3}{20} dx = \frac{3}{4}$ no es una función de densidad) y asigna a cada entero $3k > 0$ probabilidad

$$\frac{e^{-3(k-1)/5}(1 - e^{-3/5})}{4}$$

($\sum_{k=0}^{\infty} e^{-3(k-1)/5}(1 - e^{-3/5}) \frac{1}{4} = \frac{1}{4}$ no es función de masa).

La duración esperada es pues

$$E[X] = \frac{1 - e^{-3/5}}{4} \sum_{k=1}^{\infty} 3k e^{-3(k-1)/5} + \frac{3}{20} \int_0^{\infty} x e^{-x/5} dx = \frac{3}{4(1 - e^{-3/5})} + \frac{3}{4} 5.$$

Supera en 0,412 minutos a la duración esperada de las conversaciones originales (cuya distribución era exponencial de parámetro $1/5$).

Proposición 4.1. Sea X una variable aleatoria con distribución F de tipo mixto (sin componente singular) y $g : \mathbb{R} \mapsto \mathbb{R}$ una función, con derivada continua g' que sólo se anula en un número numerable de puntos, entonces la variable aleatoria $Y = g(X)$ tiene esperanza

$$E[Y] = \int_{-\infty}^{\infty} g(x) F(dx) \quad (4.6)$$

siempre que $\int_{\{g>0\}} g(x) F(dx) < \infty$ y $\int_{\{g<0\}} g(x) F(dx) > -\infty$.

En efecto separemos el caso discreto y continuo:

Si X v.a. discreta con $D = \{a_i\}_{i \in I}$ y $p_i = P(X = a_i) \forall i \in I$.

Sea $Y = g(X)$ entonces $E[Y] = \sum_{i \in I} g(a_i) p_i$

Demostración:

$$Y = g(X) \Rightarrow Y \text{ discreta con soporte } D' = g(D) \text{ y función de cuantía } q_j = P(Y = b_j) = P_X\{a_i/g(a_i) = b_j\} = \sum_{\{i|g(a_i)=b_j\}} p_i.$$

$$\text{Por definición de esperanza: } E[Y] = \sum_{j \in J} b_j q_j$$

Halleemos:

$$\begin{aligned} \sum_{i \in I} g(a_i) p_i &= \sum_{j \in J} \sum_{\{i|g(a_i)=b_j\}} g(a_i) p_i = \sum_{j \in J} \sum_{\{i|g(a_i)=b_j\}} b_j p_i = \sum_{j \in J} b_j \sum_{\{i \in g(a_i)=b_j\}} p_i = \\ &= \sum_{j \in J} b_j q_j = E[Y] \end{aligned}$$

Sea X absolutamente continua con función de densidad $f(x)$. Sea $Y = g(X)$ siendo g una función con derivada continua, estrictamente monótona, entonces

$$E[Y] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Demostración:

De acuerdo con el tema anterior Y es absolutamente continua con función de densidad

$$\bar{f}(y) = f(g^{-1}(y)) |(g^{-1}(y))'|$$

$$\text{Por definición de esperanza } E[Y] = \int_{-\infty}^{\infty} y \bar{f}(y) dy.$$

Pero si hacemos en la integral $\int_{-\infty}^{\infty} g(x) f(x) dx$ el cambio $x = g^{-1}(y)$ resulta

$$\int_{-\infty}^{\infty} g(x) f(x) dx = \int_{-\infty}^{\infty} g(g^{-1}(y)) f(g^{-1}(y)) |(g^{-1}(y))'| dy = \int_{-\infty}^{\infty} y \bar{f}(y) dy$$

La proposición anterior justifica la siguiente definición para el valor esperado de una función de variable aleatoria

Definición 4.3. *Valor esperado de una función de una variable aleatoria*

Sea X una variable aleatoria de tipo discreto con función de probabilidad $P(x_i)$ y sea $g(X)$ una función de la variable aleatoria X , entonces definimos $E[g(X)]$ como:

$$E[g(X)] = \sum_i g(x_i) P(x_i) \quad (4.7)$$

Si la variable aleatoria X es de tipo continuo con función de densidad $f(x)$, entonces definimos $E[g(X)]$ como:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \quad (4.8)$$

ó también

$$E[g(X)] = \int_{\mathbb{R}} g(x)F(dx) \quad (4.9)$$

siempre que $\int_{\{g>0\}} g(x)F(dx) < \infty$ y $\int_{\{g<0\}} g(x)F(dx) > -\infty$.

Ejemplos

4. Sea X una variable aleatoria con función de densidad

$$f(x) = \begin{cases} kx + 1/2, & \text{si } x \in [-1, 1] \\ 0, & \text{en otro caso} \end{cases}$$

- a) Determinar los valores de k para los cuales $f(x)$ es función de densidad.
 - b) Calcular la esperanza, la moda y la mediana de X .
 - c) ¿Para qué valores de k se minimiza $Var(X)$?
5. De una estación de cercanías sale un tren dejando alternativamente 10 y 20 minutos después de la última salida. Sea X la variable aleatoria que nos mide el tiempo de espera de un viajero que llega a la estación en un momento cualquiera para tomar el primer tren que salga:
- a) Determinar la distribución de la variable X .
 - b) Calcular la esperanza de X .
 - c) Calcular la probabilidad de que tenga que esperar menos de 8 minutos.
 - d) Comparar los resultados con una estación en la que las salidas son a intervalos de 15 minutos.

4.3. Propiedades de la Esperanza.

Las propiedades de la Esperanza Matemática son:

1. La esperanza de una constante es la propia constante. Es decir, si k es una constante entonces

$$E[k] = k$$

Demostración: Teniendo en cuenta la definición de valor esperado tendremos, para el caso discreto:

$$E[k] = \sum_i kP(x_i) = k \sum_i P(x_i) = k$$

pues

$$\sum_i P(x_i) = 1$$

Análogamente para el caso continuo

$$E[k] = \int_{-\infty}^{\infty} kf(x)dx = k \int_{-\infty}^{\infty} f(x)dx = k$$

pues

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

2. Si una variable aleatoria X está acotada, es decir, existen dos valores a y b tales que $a \leq X \leq b$, entonces se verifica que

$$a \leq E[X] \leq b$$

Demostración: Como la variable aleatoria X está acotada los posibles valores que toma, para todo i verifican que

$$a \leq x_i \leq b$$

y como la distribución de probabilidad de X es

$$P(x_i) = P(X = x_i), \quad i = 1, 2, \dots, \text{con } \sum_i P(x_i) = 1$$

se verifica que

$$a \cdot P(X = x_i) \leq x_i \cdot P(X = x_i) \leq b \cdot P(X = x_i)$$

y sumando para todos los valores de i , tendremos

$$\begin{aligned} \sum_i a \cdot P(X = x_i) &\leq \sum_i x_i \cdot P(X = x_i) \leq \sum_i b \cdot P(X = x_i) \\ a \cdot \sum_i P(X = x_i) &\leq E[X] \leq b \cdot \sum_i P(X = x_i) \end{aligned}$$

y como

$$\sum_i P(x_i) = \sum_i P(X = x_i) = 1$$

tendremos

$$a \leq E[X] \leq b$$

Análogamente se puede demostrar para el caso continuo.

3. Sea X una variable aleatoria y sean $g(X)$ y $h(X)$ dos funciones de X que, a su vez son variables aleatorias, cuyos valores esperados existen y sean a y b dos constantes cualesquiera, entonces:

$$E[a \cdot g(X) + b \cdot h(X)] = a \cdot E[g(X)] + b \cdot E[h(X)]$$

Demostración: Consideremos en primer lugar el caso continuo y designamos por $f(x)$ la función de densidad de X , entonces:

$$\begin{aligned} E[a \cdot g(X) + b \cdot h(X)] &= \int_{-\infty}^{\infty} [a \cdot g(x) + b \cdot h(x)f(x)]dx = \\ &= a \cdot \int_{-\infty}^{\infty} g(x)f(x)dx + b \cdot \int_{-\infty}^{\infty} h(x)f(x)dx = \\ &= a \cdot E[g(X)] + b \cdot E[h(X)] \end{aligned}$$

Análogamente se puede hacer para el caso discreto. Sea $P(x_i)$ la función de probabilidad de la variable aleatoria X , entonces,

$$\begin{aligned} E[a \cdot g(X) + b \cdot h(X)] &= \sum_i [a \cdot g(x_i) + b \cdot h(x_i)]P(x_i) = \\ &= a \cdot \sum_i g(x_i)P(x_i) + b \cdot \sum_i h(x_i)P(x_i) = \\ &= a \cdot E[g(X)] + b \cdot E[h(X)] \end{aligned}$$

Esta propiedad se puede extender al caso de n funciones de la variable aleatoria X y se conoce como propiedad de linealidad de la esperanza.

Como consecuencias inmediatas de esta propiedad podemos indicar las siguientes:

$$\begin{aligned} E[g(X) + h(X)] &= E[g(X)] + E[h(X)] \\ E[a \cdot X + b] &= a \cdot E[X] + b \\ E[a \cdot X] &= a \cdot E[X] \end{aligned}$$

4. Sea X una variable aleatoria y sean $g(X)$ y $h(X)$ dos funciones de X que a su vez son variables aleatorias cuyos valores esperados existen; si $g(X) \leq h(X)$ entonces

$$E[g(X)] \leq E[h(X)]$$

Demostración: La demostración es inmediata, bastaría aplicar la definición de valor esperado y comprobar que en el caso discreto cada sumando del primer miembro es menor o igual que cada sumando del segundo miembro, pues $\forall x_i$, se verifica

$$g(x_i) \leq h(x_i)$$

y consecuentemente $g(x_i) \cdot P(x_i) \leq h(x_i) \cdot P(x_i)$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \leq \int_{-\infty}^{\infty} h(x)f(x)dx = E[h(X)]$$

pues $\forall x$, se verifica que $g(x) \leq h(x)$.

5. $E[X]$ existe y es finita si y sólo si $E[|X|] < \infty$. Entonces $|E[X]| \leq E[|X|]$. De hecho, $E[X]$ existe y es finita cuando $E[X^+] < \infty$ y $E[X^-] < \infty$; puesto que $|X| = X^+ + X^-$, ello equivale a $E[|X|] < \infty$. Además

$$|E[X]| = |E[X^+] - E[X^-]| \leq E[X^+] + E[X^-] = E[|X|].$$

6. Sea X una variable aleatoria y sea $g(X)$ una función de X que a su vez es una variable aleatoria, cuyo valor esperado existe, entonces:

$$|E[g(X)]| \leq E[|g(X)|]$$

Demostración: Para el caso continuo tendremos:

$$|E[g(X)]| = \left| \int_{-\infty}^{\infty} g(x) \cdot f(x)dx \right| \leq \int_{-\infty}^{\infty} |g(x)| \cdot f(x)dx = E[|g(X)|]$$

pues el valor absoluto de una integral es menor o igual que la integral del valor absoluto.

Análogamente sucede en el caso discreto, pues el valor absoluto de una suma es menor o igual que la suma de los valores absolutos de cada uno de los sumandos.

7. Si X es una variable aleatoria con distribución simétrica respecto del valor c , entonces, si existe $E[X]$, su valor será igual a c .

Demostración: Para simplificar la demostración consideremos que la variable aleatoria X es de tipo continuo y que es simétrica respecto a $c = 0$, en consecuencia

$$f(x) = f(-x)$$

Teniendo en cuenta la definición del valor esperado $E[X]$, tendremos:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_{-\infty}^0 x \cdot f(x) dx + \int_0^{\infty} x \cdot f(x) dx = \\ &= - \int_0^{\infty} x \cdot f(x) dx + \int_0^{\infty} x \cdot f(x) dx = 0 \end{aligned}$$

Ahora bien si la variable aleatoria X es simétrica respecto del punto x , se verificará que

$$f(c - x) = f(c + x)$$

será simétrica respecto de cero, es decir:

$$E[X - c] = 0 \Rightarrow E[X] = c$$

Luego si una variable aleatoria es simétrica respecto de un punto c , la media de esa variable, si existe, es igual a c .

8. Si X es una variable aleatoria que toma sólo valores enteros no negativos, se cumple

$$E[X] = \sum_{m=1}^{\infty} P\{X \geq m\} \quad (4.10)$$

o lo que es lo mismo, $E[X] = \sum_{m=0}^{\infty} P\{X > m\}$. Por consiguiente, para una variable aleatoria X con valores enteros es

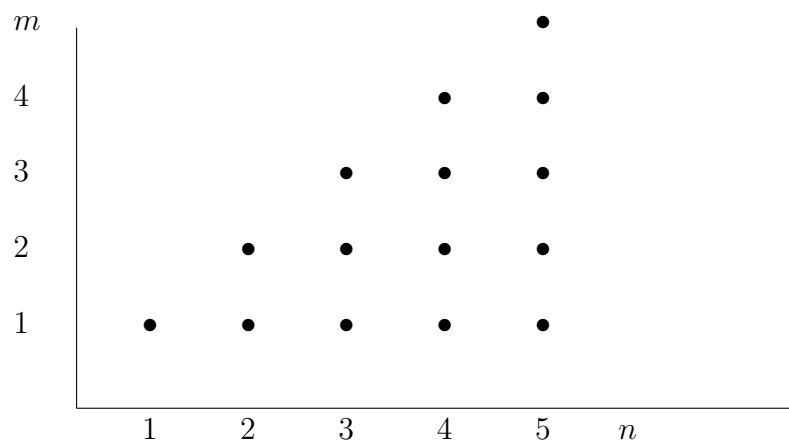
$$E[|X|] = \sum_{m=1}^{\infty} P\{|X| \geq m\} \quad (4.11)$$

En efecto,

$$E[X] = \sum_{n=1}^{\infty} nP\{X = n\} = \sum_{n=1}^{\infty} \sum_{m=1}^n P\{X = n\}$$

En la última suma, hay un sumando asociado a cada punto (n, m) del retículo adjunto y, como todos son positivos, lo mismo da sumarlos por verticales, una tras otra, como indica la expresión anterior, o bien sumarlos por horizontales, una tras otra, como indica la expresión siguiente:

$$E[X] = \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} P\{X = n\}$$



La suma interior es ahora $P\{X \geq m\}$, así que se obtiene la igualdad 4.10. Aplicando esta igualdad a $|X|$ se obtiene 4.11. Las probabilidades $P\{|X| \geq m\}$ se denominan a veces colas de la distribución, porque dan la probabilidad de que X esté fuera de la región central $(-m, m)$.

9. Para cualquier variable aleatoria $X \geq 0$

$$\sum_{n=1}^{\infty} P\{X \geq n\} \leq E[X] \leq 1 + \sum_{n=1}^{\infty} P\{X \geq n\} \quad (4.12)$$

Si ξ representa la parte entera, es evidente que $\xi(X) \leq X \leq \xi(X) + 1$ y, por consiguiente, $E[\xi(X)] \leq E[X] \leq E[\xi(X) + 1] = E[\xi(X)] + 1$. Además

$$E[\xi(X)] = \sum_{n=1}^{\infty} P\{\xi(X) \geq n\} = \sum_{n=1}^{\infty} P\{X \geq n\}.$$

10. Si $X \geq 0$ es una variable aleatoria con esperanza finita, se cumple

$$\lim_{x \rightarrow \infty} xP\{X > x\} = 0 \text{ y } E[X] = \int_0^{\infty} P\{X > x\} dx \quad (4.13)$$

Si F es la distribución de X , la función $G(y) = \int_{(0,y]} xF(dx)$ es creciente hacia $G(\infty) = E[X]$. Por tanto, dado cualquier $\epsilon > 0$, existe k tal que $y > k$; con lo cual, si $y > k$ y $c > 0$, se tiene

$$\epsilon > G(y+c) - G(y) = \int_{(y,y+c]} xF(dx) \geq yP\{y < X \leq y+c\}.$$

Al crecer c hacia infinito, resulta $yP\{X > y\} < \epsilon$ siempre que $y > k$, lo cual prueba la primera parte de 4.13. Supuesto que X tenga densidad f , se puede integrar por partes, con $F(x) - 1 = -P\{X > x\}$ como primitiva de $f(x)$, para obtener

$$E[X] = \int_0^\infty xf(x)dx = \int_0^\infty [1 - F(x)]dx.$$

Con frecuencia, las dos últimas propiedades se aplican a $|X|$, con objeto de relacionar la finitud de $E[|X|]$ - equivalente a la de $E[X]$ - con las probabilidades del tipo $P\{|X| > x\}$, denominadas “colas” de la distribución F .

4.4. Desigualdad de Tchebychev.

Desigualdades de Markov y Tchebychev Sea $g : [0, \infty) \mapsto [0, \infty)$ una función creciente y X una variable aleatoria tal que $E[g(|X|)] < \infty$. Entonces, para cualquier $c > 0$, se verifica

$$P\{|X| \geq c\} \leq \frac{E[g(|X|)]}{g(c)} \quad (4.14)$$

En particular, para cualquier $a \in \mathbb{R}$ y $r > 0$, si $E[|X - a|^r] < \infty$ se cumple

$$P\{|X - a| \geq c\} \leq \frac{E[|X - a|^r]}{c^r} \quad (4.15)$$

$$E[g(|X|)] = \int_{\mathbb{R}} g(|x|)F(dx) \geq \int_{\{|x| \geq c\}} g(|x|)F(dx) \geq g(c)P\{|X| \geq c\}.$$

Si $g(x) = x^r$, y se sustituye X por $X - a$ se obtiene 4.15.

Si $r > 0$, se verifica

$$E[|X - a|^r] = 0 \leftrightarrow P\{X = a\} = 1 \quad (4.16)$$

Si $E[|X - a|^r] = 0$, según 4.15 es $P\{|X - a| \geq 1/n\} = 0$ para todo $n \in \mathbb{N}$, y $P\{|X - a| > 0\} = \lim_{n \rightarrow \infty} P\{|X - a| \geq 1/n\} = 0$. El recíproco es trivial.

4.5. Momentos de una distribución.

Anteriormente hemos comentado la analogía existente entre una distribución de probabilidad de una variable aleatoria y un sistema físico de masas. Por ejemplo, para la

Física, un modo de resumir un sistema es considerar que toda la masa está concentrada en un punto, el centro de gravedad.

Para el Cálculo de Probabilidades una variable aleatoria puede resumirse mediante su valor esperado, concepto equivalente al de centro de gravedad de un sistema físico. La similitud ha contribuido incluso a traspasar la terminología: los momentos de una variable aleatoria representan características semejantes a las estudiadas por la Estática física.

Los momentos de una variable aleatoria son los valores esperados de ciertas funciones de la variable. Constituyen un conjunto de medidas descriptivas que pueden emplearse para caracterizar la distribución. Aunque pueden considerarse momentos referidos a cualquier valor, generalmente se definen con respecto al cero o con respecto al valor esperado.

Definición 4.4. *Momento de orden r o momento de orden r respecto al origen. Sea X una variable aleatoria, y r un número entero positivo. Definimos el momento de orden r de X , y lo notaremos por α_r como*

$$\alpha_r = E[X^r], \text{ para } r = 1, 2, \dots$$

Definición 4.5. *Momento central de orden r . Sea X una variable aleatoria y r un número entero positivo. Definimos el momento central de orden r de X y lo notaremos por μ_r como*

$$\mu_r = E[(X - E[X])^r], \text{ para } r = 2, 3, \dots$$

En ambos casos los momentos se definen como valores esperados, luego para que existan los momentos tendrán que existir los correspondientes valores esperados y, por tanto, las series o las integrales correspondientes tendrán que ser absolutamente convergentes.

El momento de orden uno es la media de la distribución de X , y se suele notar como:

$$\alpha_1 = E[X] = \mu$$

En general el momento central de orden r , también lo podemos expresar así:

$$\mu_r = E[(X - E[X])^r] = E[(X - \mu)^r]$$

Observemos que en la definición del momento central de orden r , indicamos desde $r = 2, 3, \dots$ pues el momento central de orden uno es igual a cero. En efecto, si existe $E[X]$,

tendremos:

$$\begin{aligned}
 \mu_1 &= E[(X - E[X])] \\
 &= E[X] - E[E[X]] \\
 &= E[X] - E[X] \\
 &= 0
 \end{aligned}$$

Proposición 4.2. *Dada una variable aleatoria X , si existe el momento respecto al origen de orden k , existen todos los momentos respecto al origen inferior a k .*

Si $s < k$ se tiene

$$\begin{aligned}
 E[|X|^s] &= \int_{|x| \leq 1} |x|^s F(dx) + \int_{|x| > 1} |x|^s F(dx) \\
 &\leq 1 + \int_{|x| > 1} |x|^k F(dx) \leq 1 + E[|X|^k]
 \end{aligned} \tag{4.17}$$

de forma que cuando es finito el momento de orden k son finitos todos los momentos de orden inferior a k .

Proposición 4.3.

$$\forall k, \mu_k = \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \alpha_1^{k-j} \alpha_j$$

supuestos existentes los momentos α_j , $j = 1, 2, \dots, k$.

$$\forall k, \alpha_k = \sum_{j=0}^k \binom{k}{j} \alpha_1^{k-j} \mu_j$$

supuestos existentes los momentos μ_j , $j = 1, 2, \dots, k$.

Como x^r es una función creciente en $[0, \infty)$, para todo $x; a \in \mathbb{R}$, se cumple

$$|x - a|^r \leq (|x| + |a|)^r \leq 2^r \max(|x|^r; |a|^r) \leq 2^r (|x|^r + |a|^r) \tag{4.18}$$

En consecuencia, $E[|X - a|^r] \leq 2^r (E[|X|^r] + |a|^r)$, así que cada momento es finito o no, independientemente del punto a respecto al cual se calcule.

Cuando r es entero, la relación que liga los momentos centrales con los momentos respecto al origen es fruto de la fórmula del binomio de Newton. En efecto, $(X - \mu)^r =$

$\sum_{i=0}^r \binom{r}{i} (-\mu)^i X^{r-i}$, con lo cual

$$\mu_r = \sum_{i=0}^r \binom{r}{i} (-\mu)^i \alpha_{r-i}. \tag{4.19}$$

En particular, habida cuenta que $\mu = \alpha_1$, es

$$\mu_2 = \alpha_2 - \mu^2 \quad (4.20)$$

$$\mu_3 = \alpha_3 - 3\mu\alpha_2 + 2\mu^3 \quad (4.21)$$

$$\mu_4 = \alpha_4 - 4\mu\alpha_3 + 6\mu^2\alpha_2 - 3\mu^4 \quad (4.22)$$

Al revés, como $X^r = (X - \mu + \mu)^r = \sum_{i=0}^r \binom{r}{i} \mu^i (X - \mu)^{r-i}$, resulta

$$\alpha_r = \sum_{i=0}^r \binom{r}{i} \mu^i \mu_{r-i} \quad (4.23)$$

De los momentos hay dos que merecen especial atención, el momento de orden uno, $\alpha_1 = E[X] = \mu$, media de la distribución, que es una medida de posición y que ya estudiamos. Y el otro es el momento central de orden dos, μ_2 , o varianza de la distribución.

4.5.1. Varianza

Se define como el momento central de orden dos, y se suele expresar como

$$Var(X) = \sigma_X^2 = \mu_2 = E[(X - E[X])^2]$$

refiriéndose a la varianza de la distribución.

La varianza es una medida de la dispersión de los valores de la variable respecto de su media, y nos permite conocer el grado de dispersión de los valores de la distribución, pudiendo establecer comparaciones con otras distribuciones. También la podemos utilizar como una cierta medida de cómo representa la media a la distribución.

La varianza se expresa en las mismas unidades que la variable X , pero elevadas al cuadrado, y para evitar esto se introduce la desviación estándar o desviación típica que se define como la raíz cuadrada positiva de la varianza y se expresa en la misma unidad que la variable X , la notaremos por σ_X .

4.5.2. Propiedades de la Varianza

Las propiedades de la varianza son:

1. La varianza se puede expresar como:

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \alpha_2 - \mu^2$$

Como valor esperado del cuadrado: $(X - \mu)^2$, de la distancia entre X y su media, siempre es $\sigma^2 \geq 0$; lo cual equivale a que $E[X^2] \geq E[X]^2$. Por la misma razón, σ^2 proporciona una medida de la dispersión de X alrededor de su media μ . Idéntico papel puede atribuirse a la desviación típica, con la ventaja adicional de que σ tiene las mismas unidades que X .

Cabe observar que

$$E[(X - a)^2] = E[X^2 - 2aX + a^2] = E[X^2] - 2aE[X] + a^2 \quad (4.24)$$

es un polinomio de segundo grado en a , que alcanza su mínimo cuando $a = E[X]$. Así pues, la dispersión de X es mínima cuando se mide alrededor de la media μ en vez de cualquier otro valor a . Esto justifica la elección de la varianza (o la desviación típica) como índice de la dispersión y, además, destaca el papel de la media μ como valor central de la distribución.

2. La varianza de una constante c es cero.

La demostración es inmediata.

3. Sea X una variable aleatoria cuya varianza existe y a y b dos constantes cualesquiera. Entonces:

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

Demostración: Aplicando de nuevo la definición de varianza:

$$\begin{aligned} \text{Var}(aX + b) &= E[((aX + B) - E[aX + b])^2] \\ &= E[(aX + b - aE[X] - b)^2] \\ &= a^2 E[(X - E[X])^2] \\ &= a^2 \text{Var}(X) \end{aligned}$$

y la desviación estándar será:

$$\sigma_{aX+b} = |a| \sigma_X$$

de modo que la desviación típica no se ve afectada por las traslaciones y se multiplica por el mismo factor — a — que dilata o contrae la variable.

4. Desigualdad de Tchebychev. Sea X una variable aleatoria con media μ y varianza σ^2 finita. Entonces para cualquier $k > 0$, se verifica:

$$P[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2}$$

De acuerdo con 4.16, $\sigma^2 = 0$ es equivalente a $P\{X = \mu\} = 1$. Así pues, sólo es nula la varianza de las distribuciones causales, concentradas en un único punto en el que, por supuesto, está la media. En general, la desigualdad 4.15, aplicada con $a = \mu$ y $r = 2$, da lugar a la desigualdad de Tchebychev:

$$P\{X - \mu \geq c\} \leq \frac{\sigma^2}{c^2} \text{ o bien } P\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2} \quad (4.25)$$

(al hacer $c = k\sigma$). Según ello, el intervalo $[\mu - k\sigma, \mu + k\sigma]$ contiene una proporción superior a $1 - 1/k^2$ de cualquier distribución.

Ejemplo 6:

Sea X una v.a. con función de distribución

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1+2x^2}{6} & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

Calcular $E[X]$ y $V[X]$.

4.6. Indicadores de posición y dispersión.

4.6.1. Características de las distribuciones unidimensionales

Existen unas medidas o características que sintetizan a las distribuciones unidimensionales, que sirven para describir su distribución, y que permiten cierto tipo de comparaciones.

Las características principales de las distribuciones podemos clasificarlas en,

- Medidas de posición o promedios.
- Medidas de dispersión.
- Medidas de concentración.

- Medidas de asimetría o deformación.
- Medidas de apuntamiento, curtosis, o concentración central.

El interés de este asunto queda de relieve si se piensa en la función de distribución de la renta de los habitantes de un país, tabulada de modo que cada línea contiene la proporción $F(x)$ de ciudadanos con una renta inferior a cada cantidad x (separadas unas de otras por un paso relativamente pequeño, puesto que se considera la renta como una cantidad continua). Ante tal abundancia de datos, parece imprescindible expresar la esencia de la distribución en unas pocas cifras de resumen, que permitan compararla con la de otros países o con las correspondientes a otras épocas, a efectos de poder concluir si la renta es más elevada en uno que en otro, si crece o decrece, si está mejor o peor repartida, etc.

Muchos de estos objetivos se consiguen mediante conceptos derivados de la noción de esperanza matemática y, más concretamente, mediante el valor esperado de alguna potencia de X o de $X - a$ (momentos de una distribución). Se examinan métodos alternativos, ajenos a la esperanza matemática.

La media y la desviación típica son parámetros comúnmente aceptados para medir la posición y la dispersión de una distribución. Sin embargo, si no existen, es preciso buscar alternativas razonables; e incluso si existen, la comparación entre diversas alternativas puede resultar provechosa.

4.6.2. Indicadores de posición

- La **Moda** de una distribución absolutamente continua, de densidad $f(x)$, corresponde al punto m en que f tiene un máximo; al igual que, en el caso discreto, está situada en el punto m en que se hace máxima la función de probabilidad $p(x)$.

Corresponde en ambos casos a la idea de valor más probable y, en este sentido, seguiría siendo el valor en que es máxima $p(x)$ incluso en el caso de distribuciones mixtas; aunque es raro emplearlo en estas circunstancias.

Por supuesto, la moda no es única cuando $f(x)$ o $p(x)$ alcanzan el mismo valor máximo en varios puntos (que, en el primer caso, incluso pueden ser todos los puntos de un intervalo).

El valor modal es muy significativo cuando $f(x)$ o $p(x)$ decrecen de forma relativamente rápida a ambos lados de m . En cambio, para la distribución uniforme en (a, b) , afirmar que cualquier punto de (a, b) es moda es tan inútil como correcto.

► La **Mediana** de una distribución F es cualquier valor M tal que

$$F(M^-) \leq 1/2 \text{ y } F(M) \geq 1/2 \quad (4.26)$$

es decir, al menos hay probabilidad $1/2$ tanto en $(-\infty, M]$ como en $[M, \infty)$.

Para una distribución continua, M es solución de la ecuación $F(x) = 1/2$, aunque la solución podría no ser única si es $F(x) = 1/2$ en todos los puntos de un intervalo. Cuando F tiene saltos, puede también ocurrir que $F(x) = 1/2$ no tenga solución y M cumpla $F(M^-) < 1/2$ y $F(M) > 1/2$. Lo mismo que la media μ es el valor en que se hace mínima $E[(X - a)^2]$, la mediana M minimiza $E[|X - a|]$. Para verificarlo, puede suponerse que es $M = 0$ (si no basta restar M a la variable X). Entonces, si es $a > 0$ y $g(a) = E[|X - a|]$, se tiene que

$$\begin{aligned} g(a) &= E[(a - X)I_{\{X < a\}}] + E[(X - a)I_{\{X > a\}}] \\ &= E[X(I_{\{X > a\}} - I_{\{X < a\}})] + a(P\{X < a\} - P\{X > a\}) \\ g(0) &= E[X(I_{\{X > 0\}} - I_{\{X < 0\}})] \end{aligned}$$

y, como

$$\begin{aligned} I_{\{X > a\}} - I_{\{X < a\}} - (I_{\{X > 0\}} - I_{\{X < 0\}}) &= \begin{cases} 0 & \text{si } x > a \\ -2 & \text{si } 0 < x < a \\ 0 & \text{si } x < 0 \\ -1 & \text{si } x = a \\ -1 & \text{si } x = 0 \end{cases} \\ &= -2I_{\{0 < X < a\}} - I_{\{X=0\}} - I_{\{X=a\}} \end{aligned}$$

$$\begin{aligned} g(a) - g(0) &= -E[X2I_{\{0 < X < a\}}] + a(P\{X < a\} - P\{X \geq a\}) \\ &\geq a(P\{X < a\} - P\{X \geq a\} - 2P\{0 < X < a\}) \\ &= a(P\{X \leq 0\} - P\{X > 0\}) \geq 0 \end{aligned}$$

pues si 0 es Mediana ($M = 0$)

$$P\{X \leq 0\} \geq 1/2 \Rightarrow 2P\{X \leq 0\} - 1 \geq 0.$$

Análogamente se prueba que $g(a) \leq g(0)$ si $a < 0$, en definitiva, g tiene un máximo en el origen.

Cuando se desea hacer un pronóstico \hat{x} del valor que tomará cierta variable aleatoria X , media, mediana y moda son los tres candidatos principales. Elegir $\hat{x} = \mu$

supone valorar los errores del pronóstico con un coste $(X - \hat{x})^2$ y tratar, por tanto, de minimizar el coste esperado. La elección $\hat{x} = M$ es indicada cuando el coste se valora mediante $|X - \hat{x}|$. La moda tiene la justificación, de índole diferente, de ser el valor más probable.

Ejemplo 7:

En la distribución exponencial la densidad $f(x) = \lambda e^{-\lambda x}$ ($x \geq 0$) es decreciente, de modo que la moda es $m = 0$.

La función de distribución $F(x) = 1 - e^{-\lambda x}$ vale $1/2$ en su mediana:

$$M = \frac{\log 2}{\lambda} \simeq 0,693\mu$$

donde $\mu = \frac{1}{\lambda}$ es la media.

El cálculo de la mediana puede ser trabajoso cuando no se dispone de una expresión explícita de la función de distribución. Una excepción son las distribuciones simétricas, en las que la mediana coincide con el punto de simetría, lo mismo que la media (si existe), y lo mismo que la moda si la distribución es también unimodal.

Si X tiene mediana M , $Y = X - M$ tiene mediana 0

$$\begin{aligned}\forall x \quad F_{X-M}(x) &= P(X - M \leq x) = P(X \leq M + x) = F_X(M + x) \\ F_{X-M}(0) &= F_X(M) \geq 1/2 \\ F_{X-M}(x^-) &= F_X((M + x)^-) \\ F_{X-M}(0^-) &= F_X(M^-) \leq 1/2\end{aligned}$$

visto que $g(a) = E[|Y - a|] \geq g(0) = E[|Y|]$ entonces $E[|X - M - a|] \geq E[|X - M|] \Rightarrow E[|X - b|] \geq E[|X - M|] \quad \forall b$

Si X es simétrica respecto de c , la mediana vale c pues por simetría $X - c = c - X$

$$\begin{aligned}F_X(c) &= P\{X \leq c\} = P\{X - c \leq 0\} = P\{c - X \leq 0\} = P\{X \geq c\} \\ &= 1 - F_X(c^-) \geq 1 - F_X(c) \Rightarrow F_X(c) \geq 1/2\end{aligned}$$

Análogamente

$$\begin{aligned}F_X(c^-) &\leq F_X(c) = 1 - F_X(c^-) \\ \Rightarrow F_X(c^-) &\leq 1/2 \Rightarrow c \text{ es mediana}\end{aligned}$$

- Los **Cuantiles** Dada una variable aleatoria X , definimos el cuantil de orden q , para $0 \leq q \leq 1$, como aquel valor x_q tal que

$$\begin{aligned} P(X \leq x_q) \geq q \quad \text{y} \quad P(X \geq x_q) \geq 1 - q \quad \text{si } X \text{ es discreta, o} \\ P(X \leq x_q) = q \quad \text{o} \quad F(x_q) = q \quad \text{si } X \text{ es continua} \end{aligned}$$

Los cuantiles x_q dividen la distribución en dos partes, así pues a la izquierda quedan todos los valores de la variable aleatoria que son menores o iguales que x_q y a la derecha quedan los valores que son mayores o iguales que x_q .

Los cuantiles en el caso de una variable aleatoria continua son fáciles de calcular pues bastará resolver la ecuación:

$$F(x_q) = q$$

cuya solución no será siempre única, y la solución será el cuantil que estamos buscando.

En el caso discreto el cuantil tampoco tiene por qué ser único y generalmente se obtienen por interpolación, ya que no siempre será posible obtener una solución exacta única.

Dentro de los cuantiles nos interesan: la mediana, los cuartiles, deciles y percentiles, que respectivamente serán aquellas cantidades:

$$\begin{aligned} M_e \\ Q_1, Q_3 \\ D_1, D_2, \dots, D_9 \\ P_1, P_2, \dots, P_{99} \end{aligned}$$

que dividen los valores de la variable en mitades, cuartas, décimas y centésimas partes, respectivamente.

- El **Coefficiente de Variación** Sabemos que la varianza y por tanto, la desviación típica, es una medida de dispersión de la variable respecto a su media y teniendo en cuenta su definición, vemos que la varianza de una variable está influenciada por el tamaño de los valores que toma y consecuentemente por la media, entonces para evitar esta influencia del tamaño de los valores de la variable, debemos de dar una medida relativa de dispersión que nos exprese la dispersión de la variable por su valor esperado o media. Esta medida relativa de dispersión es el coeficiente de variación, que se define como:

$$CV_X = \frac{\sigma_X}{E[X]} = \frac{\sigma_X}{\mu_X}$$

El coeficiente de variación no tendrá sentido cuando la variable X tome valores positivos y negativos, pues en este caso la media podría quedar compensada por los valores positivos y negativos y no nos reflejaría el tamaño de X . Luego el coeficiente de variación sólo tendrá sentido cuando X sea una variable que toma solo valores positivos.

Este coeficiente es adimensional e invariante frente a cambios de escala.

Ejemplo 8:

Sean X e Y dos variables aleatorias que cumplen $E(X) = 100$, $Var(X) = 64$, $E(Y) = 10$ y $Var(Y) = 4$. Si bien en términos absolutos la varianza de X es mucho mayor, la dispersión relativa de X es inferior, puesto que el coeficiente de variación de X

$$CV_X = \frac{8}{100} = 0,08$$

es inferior al coeficiente de variación de Y

$$CV_Y = \frac{2}{10} = 0,2$$

- **El Promedio de las desviaciones absolutas** Como medidas de dispersión alternativas a la desviación típica puede considerarse el promedio de las desviaciones absolutas: $E[|X - a|]$, donde a suele ser la media o la mediana.

Definición 4.6. *Desviación absoluta media respecto a la mediana. La definimos como un momento absoluto de primer orden, es decir,*

$$E[|X - M_e|] = \int_{-\infty}^{\infty} |x - M_e| f(x) dx.$$

- **El Recorrido Intercuartílico.** Como medida de dispersión asociada a los cuartiles tenemos el recorrido intercuartilístico o intercuartílico

Definición 4.7. *Recorrido intercuartílico. Definimos el recorrido intercuartílico como:*

$$R_Q = Q_3 - Q_1$$

Dentro de este intervalo intercuartílico se encuentra el 50 % de los valores centrales de la variable, prescindiendo del 25 % de los valores más pequeños y del 25 % de los valores más grandes.

El cuartil de orden $p \in (0, 1)$ es cualquier número c_p que verifique

$$F(c_p^-) \leq p \text{ y } F(c_p) \geq p$$

y puede no ser único cuando existe un intervalo en el que F vale p .

Todo intervalo $[c_p, c_{1-p}]$ con $p < 1/2$, contiene probabilidad $1 - 2p$ y su longitud es indicativa de la dispersión de la distribución. En particular, el intervalo entre los cuartiles, $c_{1/4}$ y $c_{3/4}$, contiene probabilidad $1/2$, aunque no tiene por qué ser simétrico alrededor de la mediana $c_{1/2}$.

Ejemplo 9:

Para la distribución exponencial de parámetro λ , $F(x) = 1 - e^{-\lambda x}$ ($x > 0$), el cuantil de orden p es solución de la ecuación

$$F(c_p) = p \text{ es decir } c_p = -\frac{\log(1-p)}{\lambda}$$

de modo que el intervalo $[c_p, c_{1-p}]$ tiene longitud $[\log(1-p) - \log(p)]/\lambda$; inversamente proporcional a λ , lo mismo que la desviación típica.

El promedio de las desviaciones absolutas a a es

$$E[|X - a|] = \int_0^a (a - x)\lambda e^{-\lambda x} dx + \int_a^\infty (x - a)\lambda e^{-\lambda x} dx = \frac{2e^{-\lambda a} + \lambda a - 1}{\lambda}$$

y para $a = \mu = 1/\lambda$ y $a = M = \log 2/\lambda$, toma los valores

$$E[|X - \mu|] = \frac{2e^{-1}}{\lambda} \simeq \frac{0,736}{\lambda} \text{ y } E[|X - M|] = \frac{\log 2}{\lambda} \simeq \frac{0,693}{\lambda}$$

ambos inversamente proporcionales a λ .

4.6.3. Medidas de forma

Hasta ahora y utilizando las medidas de posición hemos presentado sobre la recta real (escala de valores de la variable) las medidas de posición central más significativas de la distribución y mediante las medidas de dispersión representábamos la cercanía o alejamiento de los valores de la variable a las medidas de posición central. No obstante y aunque estas medidas nos proporcionan alguna información acerca de la forma de la distribución de la variable, la mayor parte de esa información la tendremos observando la representación gráfica de la misma. Por ello es necesario definir una serie de medidas que nos permitan cuantificar, en lo posible, la forma de la distribución, es decir, tendremos que definir otras medidas que nos den información sobre el perfil de la función de probabilidad o de la función de densidad, de tal manera que podamos cuantificar el grado de simetría y el de apuntamiento o de aplastamiento de la función de probabilidad o de la función de densidad.

Las medidas de forma serán adimensionales e invariantes ante cambios de origen y de escala.

4.6.4. Coeficiente de asimetría

El momento central de tercer orden $\mu_3 = E[(X - \mu)^3]$ se utiliza como medida de la asimetría de la distribución. Por ejemplo, si X tiene densidad $f(x)$, se puede expresar

$$\begin{aligned}\mu_3 &= \int_{\mu}^{\infty} (x - \mu)^3 f(x) dx + \int_{-\infty}^{\mu} (x - \mu)^3 f(x) dx \\ &= \int_0^{\infty} y^3 [f(\mu + y) - f(\mu - y)] dy\end{aligned}$$

(tras hacer $y = x - \mu$ e $y = \mu - x$ en cada integral), lo cual muestra que μ_3 es una combinación de las diferencias de los valores de f en puntos simétricos respecto a μ . Si $f(\mu + y) = f(\mu - y)$, la distribución es simétrica respecto a μ y a $\mu_3 = 0$. En general, μ_3 indica en qué medida predomina $f(\mu + y)$ sobre $f(\mu - y)$ o al revés: en un caso $\mu_3 > 0$ y en el otro $\mu_3 < 0$.

Lo mismo ocurre con el coeficiente de asimetría:

$$\gamma_3 = \mu_3 / \sigma^3$$

en el cual se divide por σ^3 para obtener un índice adimensional, independiente de la escala en que se exprese la variable. De hecho, si $b > 0$, tanto $\mu_3(bX) = b^3 \mu_3(X)$ como $\sigma^3(bX) = b^3 \sigma^3(X)$. Si:

$$\gamma_3 = \begin{cases} < 0 & \text{distribución asimétrica a la izda.} \\ = 0 & \text{distribución simétrica o casi simétrica respecto a la media} \\ & \text{(punto de simetría)} \\ > 0 & \text{distribución asimétrica a la derecha} \end{cases}$$

4.6.5. Coeficiente de curtosis o apuntamiento

Este coeficiente trata de medir el grado de agrupamiento de los valores centrales en torno de la media aritmética. Es decir, mide el grado de aplastamiento de la gráfica correspondiente a la función de densidad.

Fisher define este coeficiente de curtosis como:

$$\gamma_4 = \frac{\mu_4}{\sigma^4} - 3$$

y toma como referencia la función de densidad de una distribución normal.

Este coeficiente puede tomar los siguientes valores:

- $\gamma_4 = 0$ Entonces la distribución tiene el mismo tipo de concentración que la distribución normal y se dice que es mesocúrtica.
- $\gamma_4 < 0$ Entonces diremos que la distribución tiene los datos centrales menos concentrado que la distribución normal, y en consecuencia será menos apuntada que la normal, y se dirá que es platicúrtica
- $\gamma_4 > 0$ Entonces la distribución tiene los datos centrales más concentrados que la normal, por tanto, será más apuntada que la normal, y se dirá que es leptocúrtica.

Capítulo 5

Función Generatriz

5.1. Función generatriz de probabilidad.

En el caso de distribuciones de probabilidad con valores no negativos, la función generatriz resulta un instrumento de cálculo de gran utilidad. La idea, que se remonta a Euler, es definir una función g , de una variable auxiliar z , por el procedimiento siguiente:

Definición 5.1. Si X es una variable aleatoria con valores enteros no negativos, con $P\{X = n\} = p_n$ para cada n , se llama función generatriz de X a la función $g : \mathbb{R} \mapsto \mathbb{R}$ tal que $\forall z \in \mathbb{R}$

$$g_X(z) = E[z^X] = \sum_{n=0}^{\infty} z^n p_n \quad (5.1)$$

Se trata de una serie de potencias de z convergente para $|z| < 1$, término a término e indefinidamente; es decir, que se cumple

$$\begin{aligned} g'_X(z) &= \sum_{n=1}^{\infty} n z^{n-1} p_n \\ g''_X(z) &= \sum_{n=2}^{\infty} n(n-1) z^{n-2} p_n \end{aligned}$$

y, en general,

$$g_X^{(k)}(z) = \sum_{n=k}^{\infty} n(n-1) \dots (n-k+1) z^{n-k} p_n = \sum_{n=k}^{\infty} \binom{n}{k} k! z^{n-k} p_n \quad (5.2)$$

En particular, para $z = 0$ todos los sumandos se anulan excepto el primero, así que

$$g_X^{(k)}(0) = k!p_k$$

Esto, además de asegurar que $\sum_{n=0}^{\infty} z^n p_n$ es el desarrollo de Taylor de $g_X(z)$, garantiza que los términos p_n de la distribución de probabilidad pueden recuperarse a partir de $g_X(z)$:

$$P(X = n) = \frac{g_X^{(n)}(0)}{n!}$$

Ejemplo 1:

Para una distribución geométrica de parámetro p , si $q = 1 - p$, se tiene

$$p_n = pq^{n-1}, \quad n = 1, 2, \dots$$

La función generatriz es entonces

$$g_X(z) = \sum_{n=1}^{\infty} z^n pq^{n-1} = \frac{pz}{1 - qz}$$

Teorema 5.1. Sea X una variable aleatoria no negativa y supongamos que $E[|X|^k] < \infty \quad \forall k = 1, 2, \dots$. Entonces

$$E[X(X-1)\dots(X-k+1)] = g_X^{(k)}(1)$$

pues por ejemplo:

$$\begin{aligned} g_X'(1) &= \sum_{n=1}^{\infty} np_n = E[X] \\ g_X''(1) &= \sum_{n=2}^{\infty} n(n-1)p_n = E[X(X-1)] \\ g_X^{(k)}(1) &= E[X(X-1)\dots(X-k+1)] \end{aligned}$$

Para $k = 1$ y $k = 2$ tenemos, en particular, el siguiente resultado:

Corolario 5.1. Sea X una variable aleatoria no negativa. Entonces

1. $E[|X|] < \infty \Rightarrow E[X] = g_X'(1), y$
2. $E[X^2] < \infty \Rightarrow Var(X) = g_X''(1) + g_X'(1) - (g_X'(1))^2$

5.1.1. Función generatriz de momentos

Los momentos de una distribución, tanto respecto al origen como respecto a la media, pueden determinarse directamente sin más que aplicar sus definiciones respectivas. Sin embargo, algunos momentos de órdenes elevados pueden resultar muy laboriosos de calcular y es a través de la función generatriz de momentos que ahora veremos que dichos momentos pueden ser calculados más fácilmente.

Otro de los aspectos de la función generatriz de momentos y de la función característica que más tarde veremos es el de permitir caracterizar una distribución a través de ella. Es decir, a partir de ahora podremos dar la distribución de una variable aleatoria sin más que dar su función generatriz de momentos o su función característica, aunque esta última tiene una ventaja sustancial sobre aquella, siempre existe.

A pesar de su utilidad, la función generatriz de probabilidad tiene uso limitado, en el sentido de que sólo están definidas para variables aleatorias no negativas con valores enteros. Distribuciones importantes, tales como la normal o la exponencial no pueden ser manejadas con esta función. Este inconveniente se resuelve como sigue:

Definición 5.2. Sea X una variable aleatoria. La función generatriz de momentos de X es

$$\psi_X(t) = E[e^{tX}]$$

supuesto que $\exists h > 0$, tal que la esperanza exista y sea finita $\forall |t| < h$.

Nota 5.1. Como primera observación mencionamos la relación existente entre la funciones generatriz de momentos y la transformada de Laplace de funciones de variables reales. Realmente, para una variable aleatoria X , no negativa, uno puede definir la transformada de Laplace como

$$E[e^{-sX}], \quad \forall s \geq 0,$$

lo cual entonces siempre existe. Análogamente, uno puede ver la función generatriz de momentos como una transformada de Laplace a dos lados.

Nota 5.2. Notemos que para variables aleatorias con valores enteros no negativos tenemos $\psi_X(t) = g_X(e^t), \forall |t| < h$, supuesto que la función generatriz de momentos exista (para $|t| < h$)

Ejemplo 2

Dada una variable aleatoria con densidad uniforme en el intervalo $[a, b]$, es decir,

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{si } x \in [a, b] \\ 0 & \text{en otro caso} \end{cases}$$

1. Calcular la función generatriz de momentos.
2. Haciendo uso del apartado anterior, demostrar que si Y es una variable aleatoria con distribución G , la variable aleatoria $Z = G(Y)$ sigue la distribución uniforme en el intervalo $[0,1]$.

Teorema 5.2. Sean X e Y variables aleatorias.

$$\text{Si } \exists h > 0, \text{ tal que } \psi_X(t) = \psi_Y(t) \quad \forall |t| < h \Rightarrow X \stackrel{d}{=} Y$$

Si existe función generatriz de momentos, existe correspondencia biunívoca con la función de distribución. Las derivadas en 0 de la función generatriz de momentos producen los momentos, de ahí el nombre de la transformación. Asimismo, las derivadas de la función generatriz de probabilidad en el 0 producen las probabilidades y de ahí su nombre.

Teorema 5.3. Sea X una v.a. cuya función generatriz de momentos $\psi_X(t)$ existe para $|t| < h$ para algún $h > 0$. Entonces

- a) Existen momentos de todo orden, esto es, $E[|X|^r] < \infty$ para todo $r > 0$; y
- b) $E[X^n] = \psi_X^{(n)}(0) \quad \forall n = 1, 2, \dots$

Demostración:

Demostremos el teorema en el caso continuo.

Por la suposición, tenemos

$$\int_{-\infty}^{\infty} e^{tx} f_X(x) dx < \infty, \forall |t| < h.$$

Sea t , $0 < t < h$ dado. El supuesto implica que para cada $x_1 > 0$,

$$\int_{x_1}^{\infty} e^{tx} f_X(x) dx < \infty \text{ y } \int_{-\infty}^{-x_1} e^{-tx} f_X(x) dx < \infty. \quad (5.3)$$

Puesto que

$$\begin{array}{rcl} \frac{|x|^r}{e^{|tx|}} & \mapsto & 0 \\ x & \mapsto & \infty \end{array}$$

para $r > 0$ tenemos que existe x_2 tal que

$$|x|^r \leq e^{|tx|} \quad \forall |x| > x_2 \quad (5.4)$$

Ahora, sea $x_0 > x_2$. Se tiene de 5.3 y 5.4 que

$$\begin{aligned} \int_{-\infty}^{\infty} |x|^r f_X(x) dx &= \int_{-\infty}^{-x_0} |x|^r f_X(x) dx + \int_{-x_0}^{x_0} |x|^r f_X(x) dx + \int_{x_0}^{\infty} |x|^r f_X(x) dx \\ &\leq \int_{-\infty}^{-x_0} e^{tx} f_X(x) dx + |x_0|^r \cdot P(|X| \leq x_0) + \int_{x_0}^{\infty} e^{tx} f_X(x) dx < \infty \end{aligned}$$

Esto demuestra a), y de ahí se deduce b) por diferenciación:

$$\psi_X^{(n)}(t) = \int_{-\infty}^{\infty} x^n e^{tx} f_X(x) dx$$

y por lo tanto

$$\psi_X^{(n)}(0) = \int_{-\infty}^{\infty} x^n f_X(x) dx = E[X^n]$$

Nota 5.3. La idea en a) es que la función exponencial crece más rápidamente que toda función polinomial. Como consecuencia, $|x|^r \leq e^{|tx|}$ tan pronto como $|x| > x_2$. Por otro lado para $|x| < x_2$, tenemos trivialmente $|x|^r \leq Ce^{tx}$, para alguna constante C . Sigue de esto que, para todo x ,

$$|x|^r \leq (C+1)e^{|tx|}$$

y por tanto

$$E[|X|^r] \leq (C+1)E[e^{|tX|}] < \infty, \quad \forall |t| < h.$$

Esto demuestra a) en el caso continuo y también se puede demostrar en el caso discreto.

Nota 5.4. El desarrollo de Taylor de la función exponencial mantiene que

$$e^{tX} = 1 + \sum_{n=1}^{\infty} \frac{t^n X^n}{n!} \quad \forall |t| < h$$

Aplicando la esperanza término a término tenemos

$$\psi_X(t) = E[e^{tX}] = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} E[X^n] \quad \forall |t| < h.$$

Derivando término a término tenemos b).

Estudiemos ahora algunas distribuciones conocidas:

Distribución de Bernoulli. Sea $X \sim Be(p)$. Entonces $\psi_X(t) = q + pe^t$. Derivando obtenemos $E[X] = p$ y $Var(X) = pq$. La serie de Taylor de e^t permite

$$\psi_X(t) = q + p \sum_{n=0}^{\infty} \frac{t^n}{n!} = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \cdot p,$$

desde lo cual se tiene que $E[X^n] = p, n = 1, 2, \dots$. En particular, $E[X] = p$ y $Var(X) = p - p^2 = pq$.

Distribución Normal: Supongamos que $X \in N(\mu, \sigma^2)$. Entonces

$$\begin{aligned}\psi_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \\ &= e^{\mu t + \frac{\sigma^2 t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}\right\} dx = e^{\mu t + \frac{\sigma^2 t^2}{2}}, -\infty < t < \infty\end{aligned}$$

Distribución de Cauchy: La función generatriz de momentos no existe para la distribución de Cauchy, ya que $\int [e^{tx}/(1+x^2)]dx$ es divergente para todo $t \neq 0$.

El teorema anterior, sólo va en un sentido, y es que puede haber distribuciones con momentos de todo orden para las cuales no existe la función generatriz de momentos. De hecho la distribución log-Normal es un ejemplo de esto. Para ver esto, notemos que si $X \in LN(\mu, \sigma^2)$, entonces $X = e^Y$, donde $Y \in N(\mu, \sigma^2)$ lo cual implica que

$$f_X(x) = \begin{cases} \frac{1}{\sigma x \sqrt{2\pi}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, & \forall x > 0 \\ 0 & \text{resto} \end{cases}$$

Se sigue que

$$E[X^r] = E[e^{rY}] = \psi_Y(r) = \exp\left\{r\mu + \frac{1}{2}\sigma^2 r^2\right\}, \forall r > 0$$

esto es, todos los momentos existen.

Sin embargo, ya que $e^x \geq \frac{x^n}{n!} \forall n$, sigue que para todo $t > 0$

$$\begin{aligned}E[e^{tX}] &= E[e^{te^Y}] \geq E\left[\frac{(te^Y)^n}{n!}\right] = \frac{t^n}{n!} E[e^{nY}] \\ &= \frac{t^n}{n!} \psi_Y(n) = \frac{t^n}{n!} \exp\left\{n\mu + \frac{1}{2}\sigma^2 n^2\right\} = \frac{1}{n!} \exp\left\{n(\log t + \mu + \frac{1}{2}\sigma^2 n)\right\},\end{aligned}$$

lo cual puede ser muy grande eligiendo n suficientemente grande, ya que $\log t + \mu + \frac{1}{2}\sigma^2 n \mapsto \infty$ para cualquier t fijado $t > 0$ cuando $n \mapsto \infty$ y $\exp\{cn\}/n! \mapsto \infty$ cuando $n \mapsto \infty$ para toda constante positiva c . La función generatriz de momentos no existe para cualquier positivo t .

5.2. Introducción a la función característica.

Hemos introducido dos transformaciones: la función generatriz de momentos y la función generatriz de probabilidad. La ventaja de la función generatriz de momentos sobre la función generatriz de probabilidad es que esta puede definirse para todas las variables. Sin embargo la función generatriz de momentos no existe para algunas distribuciones, por ejemplo para la distribución de Cauchy y la log-Normal. En esta sección se introduce una tercera transformación, la función característica, la cual existe para todas las distribuciones. Una complicación técnica menor es que esta transformación es de valor complejo y por lo tanto requiere algún soporte matemático más sofisticado para poder tratarla con rigurosidad.

5.2.1. Variables aleatorias complejas.

Definición 5.3. *Dado un espacio probabilístico (Ω, \mathcal{F}, P) una función Y sobre Ω cuyos valores son números complejos $U + iV$, se dice que es una variable aleatoria compleja valorada si y sólo si U y V son variables aleatorias real valoradas.*

Definición 5.4. *Si Y es una variable aleatoria compleja valorada se define la esperanza de Y como $E[Y] = E[U] + iE[V]$ supuesto que existen $E[U]$ y $E[V]$.*

Con estas definiciones, todos los resultados para variables aleatorias reales sobre integración, independencia y distribuciones (en lo referente a su distribución Y se considera como un vector (U, V)) subsisten para variables aleatorias que toman valores en el conjunto \mathbb{C} de los números complejos.

Esto se debe esencialmente al hecho de que los números complejos se suman componente a componente y a que la distancia entre dos números complejos está mayorada por la suma de las distancias de las partes real e imaginaria. En particular la distancia entre dos sucesiones de números complejos tiende a cero si y sólo si tienden a cero las distancias entre las sucesiones de partes reales y las sucesiones de partes imaginarias.

Proposición 5.1. *Si X es una variable aleatoria compleja tal que $E[|X|]$ es finita, se verifica $|E[X]| \leq E[|X|]$.*

El concepto de función característica está basado en la función exponencial compleja valorada de números imaginarios puros; es decir, a cada variable aleatoria real valorada X se le asocia la variable aleatoria compleja valorada

$$Y = e^{itX} = \cos tX + i \sin tX$$

siendo t un número real

5.2.2. Función característica.

Si X es una variable aleatoria unidimensional, $e^{itX} = \cos tX + i \operatorname{sen} tX$ es una variable aleatoria con valores complejos y de módulo 1, para cada $t \in \mathbb{R}$.

Definición 5.5. Se denomina función característica de una variable aleatoria X , con función de distribución $F(x)$, como la aplicación $\varphi : \mathbb{R} \mapsto \mathbb{C}$

$$\varphi(t) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} F(dx) = \int_{-\infty}^{\infty} \cos tx F(dx) + i \int_{-\infty}^{\infty} \operatorname{sen} tx F(dx) \quad (5.5)$$

En el caso de una variable aleatoria discreta, concentrada en el conjunto $D = \{x_n\}_{n=1}^{\infty}$ y con función de probabilidad $p(x_n)$

$$\varphi(t) = \sum_{n=1}^{\infty} e^{itx_n} p(x_n) = \sum_{n=1}^{\infty} \cos tx_n p(x_n) + i \sum_{n=1}^{\infty} \operatorname{sen} tx_n p(x_n) \quad (5.6)$$

Mientras que si X es absolutamente continua, con densidad $f(x)$

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx = \int_{-\infty}^{\infty} \cos tx f(x) dx + i \int_{-\infty}^{\infty} \operatorname{sen} tx f(x) dx \quad (5.7)$$

Como $\cos tx$ y $\operatorname{sen} tx$ tienen valor absoluto menor o igual que 1, las series incluídas en 5.6 son convergentes. También son funciones continuas, de manera que las integrales de 5.7 existen y son finitas para cualquier densidad $f(x)$. En el caso de una distribución mixta, la función característica incluye términos de ambos tipos, todos finitos, de manera que $\varphi(t)$ tiene valor bien definido para dada $t \in \mathbb{R}$ y constituye una función $\varphi : \mathbb{R} \mapsto \mathbb{C}$, de la forma $\varphi(t) = u(t) + iv(t)$, donde $u(t) = E[\cos tX]$ y $v(t) = E[\operatorname{sen} tX]$.

Ejemplo 3:

- La distribución binomial $B(n, p)$, que asigna a los enteros $k = 0, 1, 2, \dots, n$, probabilidades $p_k = \binom{n}{k} p^k q^{n-k}$ ($q = 1 - p$), tiene función característica

$$\varphi(t) = \sum_{k=0}^n e^{itk} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^n \binom{n}{k} (pe^{it})^k q^{n-k} = (pe^{it} + q)^n.$$

- La distribución geométrica, con función de probabilidad $p(x) = pq^{x-1}$ para $x = 1, 2, \dots$, tiene función característica

$$\varphi(t) = \sum_{x=1}^{\infty} e^{itx} pq^{x-1} = \frac{p}{q} \frac{qe^{it}}{1 - qe^{it}} = \frac{pe^{it}}{1 - qe^{it}}.$$

Ejemplo 4:

- Para la distribución exponencial, de densidad $f(x) = \lambda e^{-\lambda x}$ para $x > 0$, es

$$\varphi(t) = \int_0^\infty e^{itx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{-(\lambda - it)x} dx = \frac{\lambda}{\lambda - it} = \left(1 - \frac{it}{\lambda}\right)^{-1}.$$

Más detalladamente, integrando por partes dos veces, se obtiene

$$\int_0^\infty \cos tx \lambda e^{-\lambda x} dx = \frac{\lambda^2}{\lambda^2 + t^2} \text{ e } \int_0^\infty \sin tx \lambda e^{-\lambda x} dx = \frac{\lambda t}{\lambda^2 + t^2}$$

luego $\varphi(t) = \lambda(\lambda + it)/(\lambda^2 + t^2) = \lambda/(\lambda - it)$.

5.2.3. Propiedades.

La siguiente proposición enumera ciertas propiedades comunes a cualquier función característica.

Proposición 5.2. *La función característica φ de cualquier variable aleatoria X verifica*

- a) $|\varphi(t)| \leq \varphi(0) = 1$ para cualquier $t \in \mathbb{R}$.
- b) φ es una función uniformemente continua.
- c) $\varphi(-t) = \overline{\varphi(t)}$; luego $u(t) = \text{parte real } \varphi(t)$ es par y $v(t) = \text{parte imaginaria } \varphi(t)$ impar.
- d) $\varphi(at)e^{itb}$ es la función característica de la variable $aX + b$ ($a, b \in \mathbb{R}$). En particular $-X$ tiene función característica $\varphi(-t) = \overline{\varphi(t)}$.
- e) Propiedad de inyectividad: Dos funciones de distribución F_1 y F_2 , con la misma función característica, coinciden.
- f) La función característica de X es real si y sólo si la distribución de X es simétrica (es decir, X y $-X$ tienen la misma distribución).
- g) Si X_1, \dots, X_n son variables aleatorias independientes

$$\varphi_{X_1+X_2+\dots+X_n}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t)\dots\varphi_{X_n}(t)$$

Demostración:

a) Según la proposición 5.1, $|\varphi(t)| \leq E[|e^{itX}|] = 1 = \varphi(0)$.

b) Para cualquier $t \in \mathbb{R}$ es

$$|\varphi(t+h) - \varphi(t)| = |E[e^{itX}(e^{ihX} - 1)]| \leq E[|e^{ihX} - 1|].$$

Ahora bien, $|e^{ihX} - 1| = \sqrt{2 - 2\cos hX} \leq 2$ y $|e^{ihX} - 1| \rightarrow 0$ cuando $h \rightarrow 0$; luego en virtud del teorema de convergencia dominada, $E[|e^{ihX} - 1|] \rightarrow 0$.

c) $\varphi(-t) = E[e^{-itX}] = E[\cos tX] - iE[\sin tX]$ es el conjugado de $\varphi(t)$

d) $E[e^{it(aX+b)}] = E[e^{itaX}e^{itb}] = e^{itb}\varphi(at)$.

e) Es consecuencia del teorema de inversión

f) De hecho, si X es simétrica, $\varphi(t) = E[e^{itX}] = E[e^{-itX}] = \varphi(-t) = \overline{\varphi(t)}$; o sea que φ es real. Y, si φ es real, $E[e^{itX}] = \varphi(t) = \overline{\varphi(t)} = E[e^{-itX}]$; luego X y $-X$ tienen la misma función característica y la misma distribución.

g) Por la independencia de las variables aleatorias X_1, X_2, \dots, X_n ,

$$\varphi_{X_1+X_2+\dots+X_n}(t) = E[e^{it(X_1+X_2+\dots+X_n)}] = E[e^{itX_1}]E[e^{itX_2}] \dots E[e^{itX_n}]$$

Ejemplo 5:

Hallar cuales de las siguientes funciones son características:

1. $\varphi(t) = e^{it^2}$
2. $\varphi(t) = (\cos t)^n$
3. $\varphi(t) = \frac{2e^{it}}{3 - e^{it}}$

5.3. Teoremas.

El siguiente objetivo consiste en establecer que la función característica de una variable aleatoria identifica su distribución. En este sentido:

Proposición 5.3. Formula de inversión de Lévy Si φ es la función característica de la variable aleatoria X con distribución F y $a < b$ son puntos en los que F es continua:

$$F(b) - F(a) = \lim_{n \rightarrow \infty} \frac{1}{2\pi} \int_{-n}^n \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt \quad (5.8)$$

Es decir, que el límite anterior caracteriza una cierta función de distribución en sus puntos de continuidad, lo cual prueba la correspondencia biunívoca entre las funciones características y las funciones de distribución.

Corolario 5.2. *Si $\varphi(t)$ es una función característica de módulo integrable, la distribución asociada F es absolutamente continua, con función de densidad acotada y uniformemente continua, dada por*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt \quad (5.9)$$

En efecto, para cualesquiera $a < b$ puntos de continuidad de F , se tiene

$$\int_a^b \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_a^b e^{-itx} dx \varphi(t) dt = F(b) - F(a)$$

pues, como el integrando tiene módulo inferior a $|\varphi(t)|$, $\int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt$ converge uniformemente en $x \in \mathbb{R}$ y se puede permutar el orden de integración; después basta efectuar la integral interior y aplicar la fórmula de inversión.

Ejemplo 6:

La distribución de Laplace, de densidad $f(x) = e^{-|x|}/2$ para $x \in \mathbb{R}$, es simétrica luego su función característica es real y vale

$$\varphi(t) = \int_0^{\infty} \cos txe^{-x} dx = 1 - t \int_0^{\infty} \sin txe^{-x} dx = 1 - t^2 \int_0^{\infty} \cos txe^{-x} dx = 1 - t^2 \varphi(t)$$

Es decir, $\varphi(t) = (1 + t^2)^{-1}$. Tal función característica es integrable, de manera que, en virtud del último corolario, se cumple

$$e^{-|x|} = \frac{1}{\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{dt}{1 + t^2} = \int_{-\infty}^{\infty} e^{ixy} \frac{dy}{\pi(1 + y^2)};$$

luego $\varphi(t) = e^{-|t|}$ es la función característica de la distribución de Cauchy $C(0, 1)$.

Para $a > 0$ y $b \in \mathbb{R}$, la distribución de Cauchy $C(b, a)$ corresponde a $aX + b$ donde X tiene distribución $C(0, 1)$. Su función característica es pues $\varphi(t) = e^{itb - a|t|}$.

Ejemplo 7:

Utilizar el teorema de inversión para encontrar la función de densidad correspondiente a la función característica $\varphi(t) = \frac{\sin t}{t}$.

5.3.1. La función característica y el problema de los momentos.

Supuesto que los momentos de la variable aleatoria existen, éstos se pueden hallar mediante las sucesivas derivadas de la función característica.

Proposición 5.4. *Si una variable aleatoria X tiene momento finito de orden $n \in \mathbb{N}$, su función característica φ tiene derivada uniformemente continua de orden $k \leq n$:*

$$\varphi^{(k)}(t) = i^k E[X^k e^{itX}] \text{ y, en particular, } \varphi^{(k)}(0) = i^k E[X^k] \quad (5.10)$$

Demostración: La haremos por inducción. Para $n = 1$, supongamos que

$$\exists \alpha_1 = \int_{\mathbb{R}} x dF(x)$$

y, por tanto, que

$$\int_{\mathbb{R}} |x| dF(x) < \infty$$

Sea $t \in \mathbb{R}$,

$$\begin{aligned} \varphi'(t) &= \lim_{h \rightarrow 0} \frac{\varphi(t+h) - \varphi(t)}{h} = \lim_{h \rightarrow 0} \int_{\mathbb{R}} \frac{e^{i(t+h)x} - e^{itx}}{h} dF(x) = \\ &= \int_{\mathbb{R}} \lim_{h \rightarrow 0} \frac{e^{i(t+h)x} - e^{itx}}{h} dF(x) = i \int_{\mathbb{R}} e^{itx} x dF(x) = iE[Xe^{itX}] \end{aligned}$$

en donde hemos utilizado la acotación

$$\left| \frac{e^{i(t+h)x} - e^{itx}}{h} \right| \leq |x|.$$

con lo que se puede aplicar el teorema de la convergencia dominada.

Tenemos que $\varphi'(0) = i\alpha_1$.

Supuesto cierto para $n - 1$, se tiene que

$$\begin{aligned} \varphi^{(n)}(t) &= \lim_{h \rightarrow 0} \frac{\varphi^{(n-1)}(t+h) - \varphi^{(n-1)}(t)}{h} = \lim_{h \rightarrow 0} i^{n-1} \int_{\mathbb{R}} e^{itx} \frac{e^{ihx} - 1}{h} x^{n-1} dF(x) = \\ &= i^{n-1} \int_{\mathbb{R}} \lim_{h \rightarrow 0} \frac{(e^{i(t+h)x} - e^{itx} - 1)}{h} x^{n-1} dF(x) = i^n \int_{\mathbb{R}} e^{itx} x^n dF(x) \end{aligned}$$

Em particular, haciendo $t = 0$ se tiene que

$$\varphi^{(n)}(0) = i^n \alpha_n \Rightarrow \alpha_n = \frac{\varphi^{(n)}(0)}{i^n}$$

Corolario 5.3. Si X tiene momento finito de orden $n \in \mathbb{N}$ su función característica se puede expresar

$$\varphi(t) = 1 + \sum_{k=1}^n \frac{i^k E[X^k]}{k!} t^k + \frac{\delta_n(t)}{n!} t^n \quad (5.11)$$

donde $\delta_n(t) \mapsto 0$ cuando $t \mapsto 0$.

En efecto, el desarrollo de Taylor de φ en el origen es

$$\varphi(t) = 1 + \sum_{k=1}^{n-1} \frac{i^k E[X^k]}{k!} t^k + \frac{\varphi^{(n)}(\lambda t)}{n!} t^n = 1 + \sum_{k=1}^n \frac{i^k E[X^k]}{k!} t^k + \frac{\delta_n(t)}{k!} t^n$$

donde $\lambda \in (0, 1)$ y $\delta_n(t) = \varphi^{(n)}(\lambda t) - i^n E[X^n]$. Puesto que $\varphi^{(n)}$ es continua, cuando $t \mapsto 0$, $\delta_n(t) \mapsto 0$.

Cuando F tiene momento de cualquier orden finito, la relación 5.11 es válida para cualquier $n \in \mathbb{N}$. Si $\delta_n(t)t^n/n! \mapsto 0$ cuando n crece, resulta

$$\varphi(t) = 1 + \sum_{k=1}^{\infty} \frac{i^k E[X^k]}{k!} t^k$$

así que la función característica, y por consiguiente la distribución, quedan determinadas por la sucesión de sus momentos.

En sentido inverso al de la proposición 5.4 se verifica:

Proposición 5.5. Si la función característica $\varphi(t)$ de una variable aleatoria X con función de distribución F , tiene derivada finita de orden $2n$ en el origen, entonces $E[X^{2n}]$ es finita.

Si existe $\varphi''(0)$, cuando $h \mapsto 0$ se tiene

$$\begin{aligned} E \left[2 \frac{\cos hX - 1}{h^2} \right] &= E \left[\frac{e^{ihX} + e^{-ihX} - 2}{h^2} \right] = \frac{\varphi(h) + \varphi(-h) - 2}{h^2} \\ &\sim \frac{\varphi'(h) - \varphi'(-h)}{2h} \mapsto \varphi''(0) \end{aligned}$$

según la regla de l'Hopital. Pero, si se toma $h = 1/2^n$,

$$Y_n = 2 \frac{1 - \cos hX}{h^2} = 2^{2n+1}(1 - \cos X/2^n)$$

es creciente hacia X^2 y no negativa. Luego, por el teorema de convergencia monótona, $E[X^2] = \lim_n E[Y^n] = -\varphi''(0)$.

Por inducción, supongamos que $E[X^{2n-2}]$ es finita cuando existe $\varphi^{(2n-2)}(0)$. Entonces

$$G(x) = \frac{1}{E[X^{2n-2}]} \int_{(-\infty, x]} y^{2n-2} F(dy)$$

es una función de distribución, con función característica

$$\psi(t) = \frac{1}{E[X^{2n-2}]} = \int_{-\infty}^{\infty} e^{itx} x^{2n-2} F(dx) = \frac{E[X^{2n-2} e^{itX}]}{E[X^{2n-2}]} = \frac{(-1)^{n-1} \varphi^{(2n-2)}(t)}{E[X^{2n-2}]}$$

de acuerdo con la proposición 5.4. La existencia de $\varphi^{2n}(0)$, equivale a la de $\psi''(0)$ y, según la primera parte del razonamiento,

$$-\psi''(0) = \int_{-\infty}^{\infty} x^2 G(dx) = \frac{1}{E[X^{2n-2}]} \int_{-\infty}^{\infty} x^{2n} F(dx) = \frac{E[X^{2n}]}{E[X^{2n-2}]}$$

con lo cual $E[X^{2n}]$ existe y vale $-\psi''(0)E[X^{2n-2}] = (-1)^n \varphi^{(2n)}(0)$.

Proposición 5.6. Sea X una variable aleatoria discreta definida en el conjunto de los números enteros \mathbb{Z} , con función de masa $p_k = p_X(k)$ y función característica

$$\varphi_X(t) = \sum_{k=-\infty}^{\infty} p_k e^{itk}.$$

Entonces

$$p_k = P\{X = k\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi_X(t) e^{-itk} dt.$$

En efecto:

Sea k' un entero arbitrario, pero fijo. Entonces,

$$e^{-ik't} \varphi_X(t) = \sum_{k=-\infty, k \neq k'}^{\infty} p_k e^{it(k-k')} + p_{k'}.$$

Sabemos que para todo número $x \neq 0$ se tiene que

$$\int_{-\pi}^{\pi} e^{itx} dt = 0 \Rightarrow \forall k \neq k'$$

integrando entre los límites $(-\pi, \pi)$ tendremos que

$$\begin{aligned} \int_{-\pi}^{\pi} e^{-itk'} \varphi_X(t) dt &= \int_{-\pi}^{\pi} \left[\sum_{k=-\infty, k \neq k'}^{\infty} p_k e^{it(k-k')} + p_{k'} \right] dt = \\ &= \sum_{k=-\infty, k \neq k'}^{\infty} p_k \int_{-\pi}^{\pi} e^{it(k-k')} dt + p_{k'} \int_{-\pi}^{\pi} dt = 0 + p_{k'} 2\pi = 2\pi p_{k'} \end{aligned}$$

con lo que, puesto que el k' era arbitrario,

$$p_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \varphi_X(t) dt$$

.

Ejemplo 8:

Sea X una variable aleatoria con función característica $\varphi_X(t) = (pe^{it} + q)^n$, siendo $p + q = 1$. Entonces, su función de masa será

$$\begin{aligned} p_k &= P\{X = k\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} (pe^{it} + q)^n dt = \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{r=0}^n \binom{n}{r} p^r e^{-it(k-r)} q^{n-r} dt \end{aligned}$$

integral que es cero $\forall k \neq r$ y si $k = r$

$$p_k = \frac{2\pi}{2\pi} \binom{n}{k} p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}.$$

Esta variable aleatoria, que más tarde estudiaremos, recibe el nombre de binomial.

Bibliografía

- [1] José M. Casas (2000) *Estadística I Probabilidad y Distribuciones*. Ed. Centro de Estudios Ramón Areces
- [2] P. Ibarrola (2005) *Apuntes sobre el mismo temario Curso 2005/06*
- [3] V. Quesada, A. García Pérez (2000) *Lecciones de Cálculo de Probabilidades*. Ed. Díaz de Santos
- [4] R. Vélez, V. Hernández (1995) *Cálculo de Probabilidades I*. UNED
- [5] R. Vélez.(2004) *Cálculo de Probabilidades II* Ed. Académicas