

Tema 1. Introducción al Cálculo de Probabilidades

Índice

- 1.1 Introducción. Experimentos deterministas vs. aleatorios
- 1.2 Espacio muestral y sucesos aleatorios
- 1.3 Recuento de sucesos elementales. Espacio muestral finito
- Ejercicios

1.1 Introducción. Experimentos deterministas vs. aleatorios

Existen fenómenos donde la concurrencia de unas circunstancias fijas no permite anticipar cuál será el efecto producido. Por ejemplo, si una moneda cae al suelo, no es posible conocer por anticipado el punto exacto donde irá a parar; cuando se colocan bolas idénticas numeradas en una bolsa y se extrae una bola a ciegas, no es posible determinar con total certeza qué bola será elegida; si se realiza todos los días el mismo trayecto entre dos puntos alejados, entonces nunca se tardará exactamente el mismo tiempo; etc.

Estos experimentos son llamados **aleatorios**, puesto que el resultado del fenómeno en estudio es consecuencia del azar. Los experimentos no aleatorios son llamados **deterministas**.

En estas situaciones el carácter impredecible de las consecuencias del azar hace inútil cualquier intento de hallar reglas deterministas que rijan la aparición de resultados individuales. Sin embargo, es falso decir que el azar no está sometido a leyes, lo que ocurre es que no son leyes necesarias, que determinen unívocamente el resultado de cada experimento, sino que atañen a la frecuencia de los resultados que se obtienen cuando el fenómeno se repite un gran número de veces.

El Cálculo de Probabilidades se ocupa de estudiar fenómenos aleatorios; es decir, situaciones que, repetidas bajo condiciones idénticas, pueden dar lugar a diversos resultados A_1, A_2, \dots , de manera que no puede predecirse con certeza absoluta cuál de ellos ocurrirá. Se dice entonces que el resultado es consecuencia del azar o que se trata de un fenómeno aleatorio.

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; es decir, un valor numérico que informa de la frecuencia con que hay que esperar que se presente cada uno, después de numerosas observaciones del fenómeno. En concreto, la probabilidad de cada acontecimiento posible es un número de $[0, 1]$, que expresa la frecuencia teórica con que dicho acontecimiento se presentaría en una serie indefinidamente larga de repeticiones del experimento realizadas en condiciones idénticas.

En el estudio de un fenómeno en que interviene el azar hay dos facetas fundamentales:

- Los posibles acontecimientos que pueden producirse; es decir, el espacio muestral y sus correspondientes sucesos.
- La valoración de la probabilidad de los acontecimientos posibles.

A continuación dedicamos la sección 1.2 al estudio de las nociones de espacio muestral y sucesos aleatorios. De forma más especializada, la sección 1.3 trata los espacios muestrales finitos. El concepto de probabilidad es abordado en el tema 2.

1.2 Espacio muestral y sucesos aleatorios

La mejor descripción de las consecuencias que pueden presentarse en relación con un fenómeno aleatorio consiste en poner de manifiesto la lista de todos los resultados posibles.

Ejemplo 1.1 Si se mezcla una baraja española y se observa la primera carta del mazo, entonces la lista de los resultados posibles coincide con la lista de 40 naipes del mazo.

Ejemplo 1.2 Después de barajar conjuntamente dos barajas españolas de dorsos diferentes (en color), al extraer una carta podrían distinguirse 80 resultados distintos.

La lista de resultados posibles determina el **espacio muestral** asociado a un experimento aleatorio.

Definición 1.1 (Espacio muestral) *Los resultados posibles constituyen un conjunto denominado espacio muestral del fenómeno aleatorio y se denota por Ω . Se llaman sucesos a los distintos subconjuntos de Ω , que en el caso de contener a un único elemento son llamados sucesos elementales y en el caso de contener dos o más elementos son llamados compuestos (y son uniones de sucesos elementales).*

Es habitual decir que

- Ω es **discreto** si sus elementos pueden ponerse en correspondencia uno-a-uno con \mathbb{N} o un subconjunto finito de \mathbb{N} (en este caso, se dice que Ω es **finito**).
- Ω es **continuo** si su cardinal es el cardinal de $[0, 1]$.

Es interesante observar que un suceso tiene dos descripciones: una descripción verbal (referida a alguna característica de sus elementos) y una descripción extensiva (enumerando todos los elementos que lo componen).

Ejemplo 1.3 Se considera el lanzamiento de dos monedas. El espacio muestral es

$$\Omega = \{(C, C), (C, X), (X, C), (X, X)\},$$

donde C = "cara" y X = "cruz". El suceso "obtener dos caras" $\{(C, C)\}$ es elemental (análogo para el suceso "obtener dos cruces") y el suceso "obtener una cara" $\{(C, X), (X, C)\}$ es compuesto.

La identificación de los sucesos relativos a un experimento aleatorio implica disponer de operaciones para formar nuevos sucesos desde otros sucesos dados. Las operaciones básicas son

- Unión de sucesos: si A y B son subconjuntos de Ω , entonces $A \cup B$ se describe como "ocurre A u ocurre B "; es decir, el resultado pertenece o bien a A , o bien a B o bien a ambos simultáneamente.
- Intersección de sucesos: si A y B son subconjuntos de Ω , entonces $A \cap B$ se describe como "ocurren A y B simultáneamente". Si $A \cap B = \emptyset$, se dice que A y B son sucesos incompatibles.
- Complementario de un suceso: si $A \subset \Omega$, entonces $A^c = \Omega - A$ significa que el resultado no pertenece a A .

La familia de todos los sucesos asociados al espacio muestral Ω es denotada por $\mathcal{P}(\Omega)$ y llamada **partes de Ω** . A continuación estudiamos algunos elementos básicos de la Teoría de Conjuntos; contenidos adicionales aparecen recogidos en el tema 2.

Sea Ω un espacio muestral. Llamamos sucesión en $\mathcal{P}(\Omega)$ a toda aplicación de \mathbb{N} en $\mathcal{P}(\Omega)$ (es decir, $\mathbb{N} \rightarrow \mathcal{P}(\Omega)$) y es representada por $\{A_n : n \geq 1\} \subset \mathcal{P}(\Omega)$.

Definición 1.2 Sea $\{A_n : n \geq 1\} \subset \mathcal{P}(\Omega)$ una sucesión de subconjuntos.

(i) Se llama *límite inferior* de $\{A_n : n \geq 1\}$, denotado por $\liminf A_n$, al conjunto de puntos $\omega \in \Omega$ que pertenecen a todo A_n , excepto a lo sumo a un número finito de éstos.

(ii) Se llama *límite superior* de $\{A_n : n \geq 1\}$, denotado por $\limsup A_n$, al conjunto de puntos $\omega \in \Omega$ que pertenecen a un número infinito de conjuntos A_n .

Los conjuntos $\liminf A_n$ y $\limsup A_n$ no necesariamente coinciden; por ejemplo, dada la sucesión $\{A_n : n \geq 1\}$ tal que $A_{2n} = B$ y $A_{2n+1} = C$, se tienen $\liminf A_n = B \cap C$ y $\limsup A_n = B \cup C$.

La definición 1.2 no resulta sencilla de aplicar en la práctica. En su lugar, es más operativo disponer de una caracterización conjuntista.

Proposición 1.1 Sea $\{A_n : n \geq 1\} \subset \mathcal{P}(\Omega)$ una sucesión de subconjuntos. Entonces,

$$\liminf A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n, \quad (1)$$

$$\limsup A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n. \quad (2)$$

Demostración.- Sean los conjuntos $\delta_k = \bigcap_{n=k}^{\infty} A_n$ y $\sigma_k = \bigcup_{n=k}^{\infty} A_n$. Es claro que $\delta_n \subset \delta_{n+1}$ y $\sigma_{n+1} \subset \sigma_n$.

En primer lugar, demostramos la relación $\liminf A_n \subset \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$, en el caso $\liminf A_n \neq \emptyset$. Supongamos que $\omega \in \liminf A_n$. Desde la definición 1.2.(i), $\exists k = k_\omega \in \mathbb{N}$ tal que $\forall n \geq k$, se tiene $\omega \in A_n$; es decir, $\exists k \in \mathbb{N}$ tal que $\omega \in \delta_k \subset \bigcup_{k=1}^{\infty} \delta_k = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$. Por tanto, $\liminf A_n \subset \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$.

Por otra parte, si $\omega \in \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$, entonces $\exists k = k_\omega \in \mathbb{N}$ tal que $\omega \in \delta_k$; es decir, $\exists k \in \mathbb{N}$ tal que $\forall n \geq k$ se tiene $\omega \in A_n$, lo cual implica que $\omega \in \liminf A_n$. Por tanto, $\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n \subset \liminf A_n$.

Similarmente se demuestra (2) desde los conjuntos σ_k . \square

Proposición 1.2 Sea $\{A_n : n \geq 1\} \subset \mathcal{P}(\Omega)$. Entonces, $\liminf A_n \subset \limsup A_n$.

Demostración.- Si $\omega \in \liminf A_n$, la igualdad (1) implica que $\exists k \in \mathbb{N}$ tal que $\omega \in \bigcap_{n=k}^{\infty} A_n$. Por lo tanto, $\exists k \in \mathbb{N}$ tal que $\omega \in A_n$, $\forall n \geq k$, lo cual establece que $\omega \in \limsup A_n$. \square

Definición 1.3 Se dice que $\{A_n : n \geq 1\} \subset \mathcal{P}(\Omega)$ es una sucesión convergente si y sólo si $\liminf A_n = \limsup A_n$. En este caso, el límite de $\{A_n : n \geq 1\}$ viene dado por

$$\lim_{n \rightarrow \infty} A_n = \liminf A_n = \limsup A_n.$$

Definición 1.4 Una sucesión $\{A_n : n \geq 1\} \subset \mathcal{P}(\Omega)$ se dice monótona creciente (respectivamente, decreciente) y se denota por $A_n \uparrow$ (respectivamente, $A_n \downarrow$) si y sólo si $\forall n \in \mathbb{N}$ se tiene $A_n \subset A_{n+1}$ (respectivamente, $A_{n+1} \subset A_n$).

Debe notarse que toda sucesión monótona tiene límite. En concreto, si $A_n \uparrow$, entonces $A_n \subset A_{n+1}$, $\forall n \in \mathbb{N}$, de modo que $\delta_k = \bigcap_{n=k}^{\infty} A_n = A_k$ y $\sigma_k = \bigcup_{n=k}^{\infty} A_n = \bigcup_{n=1}^{\infty} A_n$. Como consecuencia,

$$\liminf A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n = \bigcup_{k=1}^{\infty} A_k,$$

$$\limsup A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} A_n.$$

Es decir, si $A_n \uparrow$, entonces $\exists \lim_{n \rightarrow \infty} A_n$ y $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$. Análogamente, si $A_n \downarrow$, entonces $\exists \lim_{n \rightarrow \infty} A_n$ y $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

La próxima sección estudia propiedades de la teoría de conjuntos en el caso de un espacio muestral finito.

1.3 Recuento de sucesos elementales. Espacio muestral finito

En el Cálculo de Probabilidades, a menudo se presentan conjuntos demasiado grandes como para poder enumerar exhaustivamente sus elementos aunque, por otra parte, obedecen a unas reglas de formación fijas que permiten construir procedimientos para conocer su cardinal, sin necesidad de elaborar una lista de elementos. La Combinatoria agrupa los procedimientos orientados a contabilizar el número de elementos de conjuntos de este tipo, sin necesidad de enumeraciones.

Método 1.- Consiste en establecer una correspondencia biyectiva entre el conjunto en cuestión y un conjunto de n elementos.

Ejemplo 1.4 Contabilizar el número de múltiplos de 7 comprendidos entre 39 y 292. La correspondencia

$$i \leftrightarrow 7(5 + i)$$

establece una aplicación biyectiva entre los elementos del conjunto $\{1, 2, \dots, 36\}$ y los múltiplos de 7, desde el 42 hasta el 287; es decir, desde $7(5+1)$ hasta $7(5+36)$. Por ello, hay 36 múltiplos de 7.

Con frecuencia la correspondencia no es tan evidente como en el ejemplo 1.4. En otras ocasiones, el conjunto no es biyectivo con ningún conjunto de estructura más simple que él mismo, ni está ordenado, ni puede ordenarse de manera que su recuento sea sencillo.

Método 2.- El procedimiento constructivo recorre mentalmente los pasos a seguir para formar todos los elementos del conjunto, anotando las alternativas que pueden elegirse en cada uno.

Ejemplo 1.5 Contabilizar el número de palabras distintas de cinco letras diferentes que puedan formarse con las letras $\{A, E, I, L, M, N, P\}$ de forma que haya una vocal en la primera posición.

Para ello, deben rellenarse cinco posibles posiciones

$$\parallel \parallel \parallel \parallel \parallel \parallel.$$

La primera posición debe ocuparse con una vocal y, para ello, hay 3 elecciones posibles $\{A, E, I\}$

$$\|3\| \quad \| \quad \| \quad \| \quad \|.$$

La segunda posición debe ocuparse por una letra diferente de la primera, de modo que, sea cual sea la primera, hay 6 posibilidades

$$\|3\|6\| \quad \| \quad \| \quad \|.$$

Elegidas la primera y la segunda letras, hay 5 elecciones posibles para la tercera posición

$$\|3\|6\|5\| \quad \| \quad \|.$$

Razonando de esta forma,

$$\|3\|6\|5\|4\|3\|,$$

es decir, el número pedido es $3 \cdot 6 \cdot 5 \cdot 4 \cdot 3 = 1080$.

El resultado final del ejemplo 1.5 obedece a una regla de conteo: si los conjuntos A_1, A_2, \dots, A_k tienen n_1, n_2, \dots, n_k elementos, respectivamente, entonces el producto cartesiano $A_1 \times A_2 \times \dots \times A_k$ tiene $n_1 n_2 \dots n_k$ elementos.

Existen métodos generales que permiten resolver infinidad de problemas de recuento. Entre ellos hay algunos que se presentan con frecuencia:

(A) Ordenaciones.- Consideremos la siguiente situación.

Ejemplo 1.5 (Continuación) Contabilizar el número de maneras distintas de ordenar en fila a 8 personas. Un argumento parecido al del ejemplo 1.5

$$\|8\|7\|6\|5\|4\|3\|2\|1\|$$

conduce al número $8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 40320$ ordenaciones distintas.

En general, $n! =$ "n factorial" $= n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1$ determina el número de ordenaciones en fila de n objetos diferentes. Por definición, $0! = 1$.

Cuando los objetos que se ordenan no son todos distintos, puede que no interese considerar como ordenaciones diferentes las que se obtienen al permutar los objetos iguales entre sí.

Ejemplo 1.6 En un bar, cinco amigos han pedido 3 cafés y 2 cervezas. Contabilizar de cuántas maneras pueden consumir las cinco bebidas:

(i) Si los cafés son "sólo", "cortado" y "con leche", y las cervezas son de marcas diferentes, se tendrían $5! = 120$ maneras de repartir

las cinco bebidas, tantas como formas de alinearlas en la barra ante ellos.

(ii) Una vez hecho el reparto, si los 3 cafés son idénticos, pueden ser permutados entre sí, sin que el resultado se vea afectado. Esto significa que los 120 resultados pueden agruparse en $3! = 6$ grupos que no presentan diferencias; es decir, hay $120/6=20$ resultados distintos.

Análogamente, si ambas cervezas son iguales, entonces hay $20/2=10$ maneras de distribuir los 3 cafés y las 2 cervezas.

En general, una colección de n objetos, clasificados en k grupos de objetos idénticos entre sí, el primero con n_1 objetos, el segundo con n_2 objetos, etc., puede ordenarse en fila de

$$\frac{n!}{n_1!n_2!\dots n_k!}$$

maneras distintas, si no se consideran distintas las ordenaciones donde dos objetos iguales han permutado su posición. En este caso, se habla de "permutaciones de n objetos de los que n_1 son iguales entre sí, n_2 objetos son iguales entre sí, etc.". .

(B) Subconjuntos ordenados.-

Ejemplo 1.7 En una competición entre n participantes se distribuyen r medallas (de valores consecutivos). Entonces, el número de atribuciones posibles es

$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-r+1) = \frac{n!}{(n-r)!}.$$

Es decir, hay $\frac{n!}{(n-r)!}$ subconjuntos ordenados posibles de r elementos que pueden extraerse de un conjunto de r elementos.

En este ejemplo, el término "subconjuntos ordenados" se refiere a "grupos de r elementos distintos que se diferencian en alguno de sus componentes o en el orden en que aparecen". Los subconjuntos ordenados de tamaño r de un conjunto con n elementos se denominan "variaciones de r elementos tomados entre n ". Se representa por

$$(n)_r = \frac{n!}{(n-r)!}.$$

Al formar una lista ordenada de longitud r , con elementos de un conjunto de cardinal n , puede ocurrir que un mismo elemento del conjunto se repita en varios lugares diferentes de la lista. Por ello, el "número de variaciones con repetición de r elementos tomados entre n " viene dado por n^r .

(C) Subconjuntos.- Aquí nos centramos en determinar el número de subconjuntos de cardinal r que pueden formarse dentro de un conjunto con un número finito de n elementos. De acuerdo con la interpretación de permutaciones con repetición de n objetos de los cuales r son iguales entre sí (es decir, forman parte del subconjunto) y $(n-r)$ son iguales entre sí (es decir, no forman parte del subconjunto), se tiene que el número de subconjuntos distintos, con r elementos, que pueden extraerse desde un conjunto de n elementos es

$$\binom{n}{r} = \frac{n!}{r!(n-r)!},$$

siendo llamado "combinaciones de r elementos tomados entre n ".

(D) Repartos.- Si hay que repartir r objetos iguales en n grupos, existen

$$\binom{n+r-1}{r}$$

repartos posibles, tantos como secuencias en orden creciente, de r números distintos, pueden formarse con números entre 1 y $n+r-1$.

Este valor puede razonarse pensando que hay que repartir r bolas idénticas en n urnas. Las n urnas son representadas por los espacios comprendidos entre $n+1$ barras verticales; por ejemplo, la secuencia

$$| | \bullet | \bullet \bullet | \bullet \bullet \bullet | \bullet |$$

corresponde al reparto de 7 bolas en 5 urnas, donde la urna 1 está vacía, las urnas 2 y 5 contienen 1 bola, la urna 3 contiene 2 bolas y la urna 4 contiene 3 bolas. En general, un reparto de r bolas en n urnas supone emplear r símbolos \bullet y $n+1$ símbolos $|$. Dos símbolos $|$ aparecen fijos (la primera y la última barra que indican el comienzo de la urna 1 y el final de la urna n , respectivamente) y los restantes $n+r-1$ (es decir, $r+(n+1)-2$) símbolos aparecen en orden arbitrario. Hay tantos repartos de r bolas en n urnas como formas distintas de ordenar $n+r-1$ elementos

$$\frac{(n+r-1)!}{(n-1)!r!} = \binom{n+r-1}{r}.$$

Ejemplo 2.8 Contabilizar cuántos resultados son posibles en el lanzamiento simultáneo de r dados iguales. Cada dado puede mostrar uno de los seis dígitos $\{1, 2, 3, 4, 5, 6\}$. Por ello, hay 6^r resultados posibles. Sin embargo, frecuentemente los diversos resultados en los que sea idéntico el reparto de las r puntuaciones obtenidas entre los seis dígitos posibles, no se consideran resultados diferentes. Existe entonces una biyección entre el conjunto de resultados distinguibles y los repartos de r bolas en 6 urnas; es decir, existen $\binom{r+5}{r} = \binom{r+5}{5}$ resultados distinguibles.

Para finalizar esta sección, proponemos algunas identidades combinatorias:

$$\begin{aligned}
 \binom{n}{0} &= 1, \\
 \binom{n}{k} &= \binom{n}{n-k}, \\
 2^n &= \sum_{k=0}^n \binom{n}{k}, \\
 \binom{2n}{n} &= \sum_{k=0}^n \binom{n}{k}^2, \\
 \sum_{k=0}^{m-1} \binom{m}{k} \binom{n-1}{m-k-1} &= \binom{n+m-1}{n}, \\
 \binom{n}{k} &= \binom{n-1}{k} + \binom{n-1}{k-1}.
 \end{aligned}$$

Notemos que $\binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n-1} + \binom{n}{n}$ representa el número total de subconjuntos de un conjunto con n elementos, sumando los de tamaño 0, tamaño 1, ... y tamaño n . Cuando n es grande, la evaluación directa de $n!$ puede resultar ineficaz. En tal caso, se usa la aproximación de Stirling

$$n! \approx \sqrt{2\pi n} n^n e^{-n}.$$

Tema 2. Probabilidad

Indice

- 2.1 Introducción
- 2.2 Espacio de probabilidad. Definición axiomática
- 2.3 Otras aproximaciones a la noción de probabilidad
- 2.4 Probabilidad condicionada. Independencia de sucesos
- 2.5 Espacio producto
- 2.6 Teorema de la probabilidad total. Teorema de Bayes
- Ejercicios

2.1 Introducción

El Cálculo de Probabilidades no se estableció por sí solo como una ciencia matemática hasta principios del siglo XX. En ese momento, el desarrollo de las Ciencias Naturales implicó fuertes demandas sobre esta disciplina. Se hizo necesario estudiar sistemáticamente los conceptos básicos de la Teoría de la Probabilidad y clarificar las condiciones bajo las cuales los resultados de la teoría pudieran ser empleados. Por este motivo, resultó esencial una construcción axiomática que introdujera una estructura lógico-formal en la Teoría de la Probabilidad.

El problema de una construcción axiomática de la Teoría de la Probabilidad como una ciencia lógica fue propuesto y resuelto en 1919 por el matemático S.N. Bernstein, quien procedió desde una comparación cualitativa de los sucesos sobre la base de su mayor o menor probabilidad. Una aproximación diferente fue propuesta por A.N. Kolmogorov, quien relaciona la Teoría de la Probabilidad con la *moderna* Teoría de la Medida y la Teoría de Conjuntos.

El presente tema introduce, de forma básica, al alumno en el Cálculo de Probabilidades desde la visión de la Teoría de la Medida y la Teoría de Conjuntos. En concreto, la sección 2.2 comienza estudiando dos estructuras de subconjuntos de $\mathcal{P}(\Omega)$, álgebra y σ -álgebra, que se encuentran relacionadas con la definición axiomática de espacio de probabilidad. Otras aproximaciones a la noción de probabilidad son brevemente revisadas en la sección 2.3. Los conceptos de probabilidad condicionada e independencia de sucesos, estudiados en la sección 2.4, tienen por objetivo reflejar cómo alguna información disponible

sobre un experimento aleatorio puede influir en la frecuencia de aparición de un suceso. La sección 2.5 estudia el producto de espacios de probabilidad y la independencia de experimentos aleatorios. Finalmente, la sección 2.6 se dedica a dos resultados fundamentales, el teorema de la probabilidad total y el teorema de Bayes.

2.2 Espacio de probabilidad. Definición axiomática

En el Cálculo de Probabilidades es conveniente asociar a cada experimento aleatorio un espacio muestral Ω , una familia de sucesos $\mathcal{A} \subset \mathcal{P}(\Omega)$ y una medida de probabilidad definida sobre el par (Ω, \mathcal{A}) .

La familia de sucesos $\mathcal{A} \subset \mathcal{P}(\Omega)$ debe ser capaz de contener a todos los posibles sucesos elementales (obtenidos de forma directa desde el experimento aleatorio) y todos aquellos sucesos que puedan deducirse desde éstos mediante operaciones entre conjuntos, en número finito o numerable.

Comenzamos revisando dos estructuras de subconjuntos de $\mathcal{P}(\Omega)$, álgebra y σ -álgebra.

Definición 2.1 (Álgebra) *Dado un espacio muestral Ω , se dice que una familia de subconjuntos $\mathcal{A} \subset \mathcal{P}(\Omega)$ tiene estructura de álgebra si y sólo si se verifican*

- (a) $\Omega \in \mathcal{A}$.
- (b) $\forall A \in \mathcal{A}, A^c \in \mathcal{A}$.
- (c) $\forall A, B \in \mathcal{A}, A \cup B \in \mathcal{A}$.

La proposición 2.1 recoge algunas de las propiedades básicas de la estructura de álgebra.

Proposición 2.1 *Sea $\mathcal{A} \subset \mathcal{P}(\Omega)$ una familia con estructura de álgebra.*

- (a) $\forall A, B \in \mathcal{A}, A \cap B \in \mathcal{A}$.
- (b) $\forall A, B \in \mathcal{A}, A - B \in \mathcal{A}$.
- (c) $\forall A, B \in \mathcal{A}, A \triangle B \in \mathcal{A}$, donde $A \triangle B = (A - B) \cup (B - A)$ denota la diferencia simétrica de conjuntos.
- (d) $\forall \{A_1, \dots, A_n\} \subset \mathcal{A}, \cup_{i=1}^n A_i \in \mathcal{A}$.

Demostración.- (a) Sean $A, B \in \mathcal{A}$. Por ser \mathcal{A} un álgebra, $A^c, B^c \in \mathcal{A}$, de modo que $A^c \cup B^c \in \mathcal{A}$ y $(A^c \cup B^c)^c \in \mathcal{A}$. Dado que $A \cap B = (A^c \cup B^c)^c$ por las leyes de Morgan, se ha demostrado que $A \cap B \in \mathcal{A}$.

(b) Sean $A, B \in \mathcal{A}$. Entonces, $A - B = A \cap B^c \in \mathcal{A}$, dado que $A, B^c \in \mathcal{A}$.

(c) Sean $A, B \in \mathcal{A}$. Entonces, $A - B, B - A \in \mathcal{A}$ por el apartado anterior y, como consecuencia, $(A - B) \cup (B - A) \in \mathcal{A}$.

(d) La propiedad se sigue directamente desde la definición 2.1.(c). \square

La estructura de álgebra tiene asociada una clara limitación: la unión numerable de conjuntos del álgebra no está necesariamente contenida en ésta.

Definición 2.2 (σ -Álgebra) *Dado un espacio muestral Ω , se dice que una familia de subconjuntos $\mathcal{A} \subset \mathcal{P}(\Omega)$ tiene estructura de σ -álgebra si y sólo si se verifican*

- (a) $\Omega \in \mathcal{A}$.
- (b) $\forall A \in \mathcal{A}, A^c \in \mathcal{A}$.
- (c) $\forall \{A_n : n \geq 1\} \subset \mathcal{A}, \cup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Es claro que si \mathcal{A} tiene estructura de σ -álgebra, entonces \mathcal{A} es también álgebra. Para ello es suficiente observar que si $\{A_1, \dots, A_n\} \subset \mathcal{A}$, entonces es posible definir la sucesión de conjuntos

$$\begin{aligned} B_k &= A_k, & 1 \leq k \leq n, \\ B_k &= A_n, & k \geq n. \end{aligned}$$

Entonces, $\cup_{i=1}^n A_i = \cup_{i=1}^{\infty} B_i \in \mathcal{A}$.

Proposición 2.2 *Sea $\mathcal{A} \subset \mathcal{P}(\Omega)$ una familia con estructura de σ -álgebra.*

- (a) $\phi \in \mathcal{A}$.
- (b) $\forall \{A_n : n \geq 1\} \subset \mathcal{A}, \cap_{n=1}^{\infty} A_n \in \mathcal{A}$. Además, para cada n , $\cap_{i=1}^n A_i \in \mathcal{A}$.
- (c) *Todo álgebra con un número finito de elementos es σ -álgebra.*
- (d) *Toda familia \mathcal{A} con estructura de σ -álgebra es cerrada respecto a la operación paso al límite; es decir, $\forall \{A_n : n \geq 1\} \subset \mathcal{A}$, sucesión convergente, $\lim_{n \rightarrow \infty} A_n \in \mathcal{A}$.*
- (e) *La intersección de σ -álgebras, definidas sobre Ω , es una σ -álgebra.*
- (f) *Dada una familia $\mathcal{B} \subset \mathcal{P}(\Omega)$, existe una mínima σ -álgebra que la contiene. Ésta será la intersección de todas las σ -álgebras que contengan a \mathcal{B} , siendo denotada por $\sigma(\mathcal{B})$.*
- (g) *Dado un espacio muestral Ω , $A \subset \Omega$ y una σ -álgebra $\mathcal{A} \subset \mathcal{P}(\Omega)$, la familia de subconjuntos*

$$\mathcal{A}_A = \{B \subset \Omega : B = A \cap C \text{ con } C \in \mathcal{A}\}$$

es una σ -álgebra sobre $\mathcal{P}(A)$. (Notación: $\mathcal{A}_A = A \cap \mathcal{A}$.)

Demostración.- (a) Es suficiente observar que si $\Omega \in \mathcal{A}$, entonces $\phi = \Omega^c \in \mathcal{A}$.

(b) Dado $A_n \in \mathcal{A}$, se tiene que $A_n^c \in \mathcal{A}, \forall n \geq 1$. Entonces, $\cup_{n=1}^{\infty} A_n^c \in \mathcal{A}$ y $(\cup_{n=1}^{\infty} A_n^c)^c \in \mathcal{A}$. Por ello, $\cap_{n=1}^{\infty} A_n = (\cup_{n=1}^{\infty} A_n^c)^c \in \mathcal{A}$. Análogo para $\cap_{i=1}^n A_i \in \mathcal{A}$.

(c) Se sigue de observar que si \mathcal{A} sólo contiene a un número finito de conjuntos, entonces toda sucesión de conjuntos de \mathcal{A} es realmente una secuencia finita de conjuntos.

(d) Observemos que, para cualquier sucesión $\{A_n : n \geq 1\} \subset \mathcal{A}$ convergente, la proposición 1.1 implica

$$\lim_{n \rightarrow \infty} A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n.$$

Por tanto, $\lim_{n \rightarrow \infty} A_n \in \mathcal{A}$ desde la definición 2.2.(c) y la proposición 2.2.(b).

(e) Sea \mathcal{T} la familia de σ -álgebras definida por

$$\mathcal{T} = \{\mathcal{A} \subset \mathcal{P}(\Omega) : \mathcal{A} \text{ es } \sigma\text{-álgebra}\}.$$

Demostramos que $\mathcal{H} = \bigcap_{\mathcal{A} \in \mathcal{T}} \mathcal{A}$ es una σ -álgebra sobre $\mathcal{P}(\Omega)$ verificando las condiciones que caracterizan a las σ -álgebras:

- $\Omega \in \mathcal{H}$, puesto que $\Omega \in \mathcal{A}$, $\forall \mathcal{A} \in \mathcal{T}$.
- Si $A \in \mathcal{H}$, entonces $A \in \mathcal{A}$, $\forall \mathcal{A} \in \mathcal{T}$. Por ello, $A^c \in \mathcal{A}$, $\forall \mathcal{A} \in \mathcal{T}$, lo que implica $A^c \in \mathcal{H}$.
- Si $A_n \in \mathcal{H}$, entonces $A_n \in \mathcal{A}$, $\forall n \geq 1$ y $\forall \mathcal{A} \in \mathcal{T}$. Por ello, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, $\forall \mathcal{A} \in \mathcal{T}$, y $\bigcup_{n=1}^{\infty} A_n \in \mathcal{H}$.

(El apartado (g) es propuesto en la Hoja 2.) \square

Definición 2.3 (Espacio medible) *Al par (Ω, \mathcal{A}) , donde $\mathcal{A} \subset \mathcal{P}(\Omega)$ es una σ -álgebra de subconjuntos de Ω , se le denomina espacio medible (o probabilizable). Los elementos de \mathcal{A} son llamados conjuntos \mathcal{A} -medibles.*

Ejemplo 2.1 En el lanzamiento de una moneda el espacio muestral es $\Omega = \{C, X\}$. Entonces, la familia $\mathcal{A} = \{\phi, \{C\}, \{X\}, \{C, X\}\} \subset \mathcal{P}(\Omega)$ tiene estructura de σ -álgebra.

Si la moneda es lanzada dos veces, entonces

$$\begin{aligned} \Omega' &= \{(C, C), (C, X), (X, C), (X, X)\}, \\ \mathcal{A}' &= \{\phi, \\ &\quad \{(C, C)\}, \{(C, X)\}, \{(X, C)\}, \{(X, X)\}, \\ &\quad \{(C, C), (C, X)\}, \{(C, C), (X, C)\}, \{(C, C), (X, X)\}, \\ &\quad \{(C, X), (X, C)\}, \{(C, X), (X, X)\}, \{(X, C), (X, X)\}, \\ &\quad \{(C, C), (C, X), (X, C)\}, \{(C, C), (C, X), (X, X)\}, \\ &\quad \{(C, C), (X, C), (X, X)\}, \{(C, X), (X, C), (X, X)\}, \\ &\quad \{(C, C), (C, X), (X, C), (X, X)\}\}. \end{aligned}$$

Si la moneda es lanzada hasta que aparezca la primera cruz, entonces

$$\Omega'' = \{X, (C, X), (C, C, X), (C, C, C, X), \dots\}$$

y \mathcal{A}'' es la familia de todos los subconjuntos de Ω . Es claro que una forma equivalente de describir Ω'' sería tomando el número de caras requeridas antes de obtener la primera cruz; es decir, $\Omega'' = \{0\} \cup \mathbb{N}$ y $\mathcal{A} = \mathcal{P}(\{0\} \cup \mathbb{N})$.

La elección de una σ -álgebra $\mathcal{A} \subset \mathcal{P}(\Omega)$ es importante y son necesarias algunas observaciones. Si Ω es discreto, entonces siempre podemos tomar $\mathcal{A} = \mathcal{P}(\Omega)$. Entonces, cada conjunto unipuntual de \mathcal{A} es el objeto fundamental. Si Ω es continuo, entonces $\mathcal{A} = \mathcal{P}(\Omega)$ es también σ -álgebra, pero contiene más conjuntos de aquéllos que consideramos de interés.

La σ -álgebra de Borel sobre \mathbb{R} . Uno de los espacios muestrales más comunes es $\Omega = \mathbb{R}$ o algún intervalo $I \subset \mathbb{R}$. En este caso, la σ -álgebra usada se denomina σ -álgebra de Borel y es denotada por $\beta = \beta(\mathbb{R})$, o por $\beta(I)$ si $I \subset \mathbb{R}$ es un intervalo.

La σ -álgebra de Borel $\beta = \beta(\mathbb{R})$ es la mínima σ -álgebra que contiene a cada una de las siguientes colecciones de conjuntos:

- (a) La colección de intervalos cerrados de \mathbb{R}

$$\mathcal{Q}_a = \{[a, b] \subset \mathbb{R} : a \leq b\}.$$

- (b) La colección de intervalos no-acotados semi-cerrados de \mathbb{R}

$$\mathcal{Q}_b = \{(-\infty, b] \subset \mathbb{R} : b \in \mathbb{R}\}.$$

- (c) La colección de intervalos semi-abiertos de \mathbb{R}

$$\mathcal{Q}_c = \{(a, b] \subset \mathbb{R} : a < b\}.$$

- (d) La colección de intervalos abiertos de \mathbb{R}

$$\mathcal{Q}_d = \{(a, b) \subset \mathbb{R} : a < b\}.$$

Es decir, $\beta = \sigma(\mathcal{Q}_a) = \sigma(\mathcal{Q}_b) = \sigma(\mathcal{Q}_c) = \sigma(\mathcal{Q}_d)$. Claramente, la σ -álgebra de Borel β de \mathbb{R} contiene a todos los subconjuntos unipuntuales y a todos los intervalos (acotados y no-acotados). No obstante, existen subconjuntos de \mathbb{R} que no están en β (no lo mostraremos).

Con frecuencia emplearemos el hecho de que un conjunto de Borel $B \in \beta$ puede ser obtenido mediante una cantidad numerable de operaciones (uniones, intersecciones, diferencias y complementaciones) entre intervalos. En este caso, las siguientes relaciones serán de utilidad:

$$\begin{aligned} \{x\} &= \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x\right], \\ (x, y) &= (x, y] - \{y\}, \\ [x, y] &= (x, y] \cup \{x\}, \\ [x, y) &= \{x\} \cup (x, y] - \{y\}, \\ (x, \infty) &= (-\infty, x]^c. \end{aligned}$$

Los comentarios anteriores son aplicables de forma análoga al caso $\Omega = \mathbb{R}^n$, para $n \geq 2$. En concreto, sea

$$\Omega = \mathbb{R}^n = \{(x_1, \dots, x_n) : x_i \in \mathbb{R}, 1 \leq i \leq n\}.$$

La σ -álgebra de Borel sobre \mathbb{R}^n , $\beta^n = \beta(\mathbb{R}^n)$, está generada por cualquiera de las siguientes familias de subconjuntos:

- (a) Los subconjuntos abiertos de \mathbb{R}^n .
- (b) Los subconjuntos cerrados de \mathbb{R}^n .
- (c) Los semiespacios cerrados de la forma $\{(x_1, \dots, x_n) : x_i \leq b, \text{ para algún } i\}$, donde $b \in \mathbb{R}$.
- (d) Los rectángulos de \mathbb{R}^n , es decir, subconjuntos de la forma $\{(x_1, \dots, x_n) : a_i < x_i \leq b_i, 1 \leq i \leq n\}$, donde $a_i < b_i \in \mathbb{R}$.

El tercero de los elementos a considerar en relación a un experimento aleatorio es la medida de probabilidad de los sucesos contenidos en una σ -álgebra $\mathcal{A} \subset \mathcal{P}(\Omega)$.

Definición 2.4 (Medida de probabilidad) *Una función de conjunto $P : \mathcal{A} \rightarrow [0, 1]$ es llamada medida de probabilidad (o probabilidad) si y sólo si satisface*

- (a) $P(A) \geq 0, \forall A \in \mathcal{A}$.
- (b) $P(\Omega) = 1$.
- (c) (Aditividad contable) $\forall \{A_n : n \geq 1\} \subset \mathcal{A}$ tal que $A_i \cap A_j = \emptyset, \forall i \neq j$,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Llamamos probabilidad del suceso A al valor $P(A)$. La terna (Ω, \mathcal{A}, P) se denomina espacio de probabilidad.

La definición 2.4 corresponde a la definición axiomática dada por A.N. Kolmogorov (1933, *Foundations of the Theory of Probability*).

Ejemplo 2.2 Si Ω es finito (es decir, $\text{Card}(\Omega) = N < \infty$), entonces cada conjunto unipuntual $\{\omega_j\}$, para $1 \leq j \leq N$, es un suceso elemental y es suficiente asignar probabilidades a cada $\{\omega_j\}$. Entonces, si $A \in \mathcal{A} = \mathcal{P}(\Omega)$,

$$P(A) = \sum_{\omega \in A} P(\{\omega\}).$$

Una posible asignación que refleja la idea de "igualdad de verosimilitudes" determina que

$$P(\{\omega_j\}) = \frac{1}{N}, \quad 1 \leq j \leq N.$$

Entonces, se obtiene la **regla de Laplace** que asigna probabilidades en términos de

$$P(A) = \frac{\text{Card}(A)}{N} = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

Ejemplo 2.3 Si Ω es discreto, pero no-finito, no es posible asignar probabilidades mediante el principio de igualdad de verosimilitudes del ejemplo 2.2. No obstante, es suficiente la asignación de probabilidades a los sucesos elementales; es decir,

$$P(A) = \sum_{\omega \in A} P(\{\omega\}).$$

Ejemplo 2.4 Si Ω es continuo, cada subconjunto unipuntual es un suceso elemental. No es posible asignar probabilidades desde el principio de igualdad de verosimilitudes. De hecho, no es posible asignar probabilidades positivas a cada suceso elemental sin violar el axioma $P(\Omega) = 1$. Por contra, sí es posible asignar probabilidades a sucesos que consisten en intervalos. Por ejemplo, en el caso $\Omega = [0, 1]$ y $\mathcal{A} = \beta[0, 1]$, la función $P(I)$ = "longitud de I ", siendo $I \subset [0, 1]$ un intervalo, define una medida de probabilidad.

A continuación proponemos algunos ejemplos de interés sobre espacios de probabilidad.

Ejemplo 2.5 Lanzamos una moneda perfectamente equilibrada. Entonces, el espacio de probabilidad está formado por el espacio muestral $\Omega = \{C, X\}$, la σ -álgebra $\mathcal{A} = \{\phi, \{C\}, \{X\}, \{C, X\}\}$ y la medida de probabilidad

$$\begin{array}{lll} P : & \mathcal{A} & \rightarrow [0, 1] \\ & \phi & \rightarrow 0 \\ & \{C\} & \rightarrow 1/2 \\ & \{X\} & \rightarrow 1/2 \\ & \{C, X\} & \rightarrow 1. \end{array}$$

Ejemplo 2.6 Sean $\Omega = \mathbb{N}$ y $\mathcal{A} = \mathcal{P}(\mathbb{N})$. Definamos $P : \mathcal{A} \rightarrow [0, 1]$ desde la relación $P(\{i\}) = 2^{-i}$, para cada $i \in \mathbb{N}$. Entonces, P es una medida de probabilidad sobre (Ω, \mathcal{A}) .

Ejemplo 2.7 Sean $\Omega = (0, \infty)$ y $\mathcal{A} = \beta(0, \infty)$. Definamos $P : \mathcal{A} \rightarrow [0, 1]$ a través de la relación

$$P(I) = \int_I e^{-x} dx, \quad \forall I \subset (0, \infty) \text{ intervalo.}$$

Entonces, $P(\Omega) = 1$, $P(I) \geq 0$ y P es contablemente aditiva (por las propiedades de la integral de Riemann).

El teorema 2.1 propone las principales consecuencias de los axiomas que definen un espacio de probabilidad.

Teorema 2.1 Sea (Ω, \mathcal{A}, P) un espacio de probabilidad.

(a) $P(\emptyset) = 0$.

(b) (Aditividad finita) $\forall \{A_1, \dots, A_n\} \subset \mathcal{A}$ tales que $A_i \cap A_j = \emptyset, \forall i \neq j$,

$$P\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n P(A_j).$$

(c) $\forall A \in \mathcal{A}, P(A^c) = 1 - P(A)$.

(d) $\forall A, B \in \mathcal{A}$ tales que $A \subset B, P(A) \leq P(B)$.

(e) $\forall A \in \mathcal{A}, P(A) \leq 1$.

(f) (Regla de la adición) $\forall A, B \in \mathcal{A}, P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(g) (Principio de inclusión-exclusión) $\forall \{A_1, \dots, A_n\} \subset \mathcal{A}$,

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) \\ &\quad - \sum_{i_1=1}^n \sum_{i_1 < i_2}^n P(A_{i_1} \cap A_{i_2}) \\ &\quad + \sum_{i_1=1}^n \sum_{i_1 < i_2}^n \sum_{i_2 < i_3}^n P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\ &\quad - \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n). \end{aligned}$$

(h) (Subaditividad) $\forall A, B \in \mathcal{A}, P(A \cup B) \leq P(A) + P(B)$.

(i) (Subaditividad finita) $\forall \{A_1, \dots, A_n\} \subset \mathcal{A}$,

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

(j) (Subaditividad contable o desigualdad de Boole) $\forall \{A_n : n \geq 1\} \subset \mathcal{A}$,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n).$$

(k) $\forall \{A_n : n \geq 1\} \subset \mathcal{A}$,

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) \geq 1 - \sum_{n=1}^{\infty} P(A_n^c).$$

Demostración.- (a) Sea la sucesión $A_1 = A$ y $A_n = \phi$, $\forall n \geq 2$. Entonces, $\bigcup_{n=1}^{\infty} A_n = A$ y $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$; es decir, $P(A) = P(A) + \sum_{n=2}^{\infty} P(\phi)$, lo cual sólo es posible si $P(\phi) = 0$.

(b) Es suficiente tomar

$$\begin{aligned} B_k &= A_k, & 1 \leq k \leq n, \\ B_k &= \phi, & k \geq n+1. \end{aligned}$$

Entonces,

$$P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n),$$

donde $P(\bigcup_{n=1}^{\infty} B_n) = P(\bigcup_{k=1}^n A_k)$ y $\sum_{n=1}^{\infty} P(B_n) = \sum_{k=1}^n P(A_k)$.

(c) Tomamos la igualdad $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$ que equivale a $P(A^c) = 1 - P(A)$.

(d) Es suficiente observar que $B = A \cup (B - A)$ implica $P(B) = P(A) + P(B - A)$. Dado que $P(B - A) \geq 0$, se tiene que $P(A) \leq P(B)$.

(e) Se hace uso del apartado anterior (d) con $B = \Omega$.

(f) Sean $A, B \in \mathcal{A}$. Es posible expresar A y B como uniones de conjuntos disjuntos

$$\begin{aligned} A &= (A \cap B) \cup (A \cap B^c), \\ B &= (A \cap B) \cup (B \cap A^c). \end{aligned}$$

Entonces, las igualdades

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c), \\ P(B) &= P(A \cap B) + P(B \cap A^c), \end{aligned}$$

equivalen a $P(A \cap B^c) = P(A) - P(A \cap B)$ y $P(B \cap A^c) = P(B) - P(A \cap B)$. Por otra parte, desde la identidad $A \cup B = (A \cap B) \cup (A \cap B^c) \cup (B \cap A^c)$, es posible establecer

$$\begin{aligned} P(A \cup B) &= P(A \cap B) + P(A \cap B^c) + P(B \cap A^c) \\ &= P(A \cap B) + P(A) - P(A \cap B) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

- (g) Se demuestra de forma inductiva desde el apartado (f).
- (h) Se deduce desde la desigualdad $P(A \cap B) \geq 0$ y el apartado (f).
- (i) Se demuestra mediante un argumento inductivo desde el apartado (h).
- (j) Sea $\{A_n : n \geq 1\} \subset \mathcal{A}$. Definamos

$$\begin{aligned} B_1 &= A_1, \\ B_n &= A_n \cap \left(\bigcap_{i=1}^{n-1} A_i^c \right), \quad n \geq 2. \end{aligned}$$

Es claro que $\{B_n : n \geq 1\} \subset \mathcal{A}$ es una sucesión de conjuntos disjuntos tal que $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. Entonces,

$$P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n).$$

Por otra parte, $B_n \subset A_n$ implica $P(B_n) \leq P(A_n)$. Como resultado,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} P(A_n).$$

- (k) Dada la sucesión $\{A_n : n \geq 1\} \subset \mathcal{A}$,

$$\begin{aligned} P\left(\bigcap_{n=1}^{\infty} A_n\right) &= 1 - P\left(\left(\bigcap_{n=1}^{\infty} A_n\right)^c\right) \\ &= 1 - P\left(\bigcup_{n=1}^{\infty} A_n^c\right). \end{aligned}$$

Además,

$$P\left(\bigcup_{n=1}^{\infty} A_n^c\right) \leq \sum_{n=1}^{\infty} P(A_n^c).$$

Por ello, $P(\bigcap_{n=1}^{\infty} A_n) \geq 1 - \sum_{n=1}^{\infty} P(A_n^c)$. \square

Teorema 2.2 Sea $\{A_n : n \geq 1\} \subset \mathcal{A}$ una sucesión monótona. Entonces, existe $\lim_{n \rightarrow \infty} P(A_n)$ y $\lim_{n \rightarrow \infty} P(A_n) = P(\lim_{n \rightarrow \infty} A_n)$.

Demostración.- Supongamos que $A_n \uparrow A = \bigcup_{n=1}^{\infty} A_n$. Entonces, la expresión de A como unión de conjuntos disjuntos

$$A = A_1 \cup \bigcup_{n=1}^{\infty} (A_{n+1} - A_n)$$

permite escribir

$$\begin{aligned}
P(A) &= P(A_1) + \lim_{n \rightarrow \infty} \sum_{k=1}^n P(A_{k+1} - A_k) \\
&= P(A_1) + \lim_{n \rightarrow \infty} \sum_{k=1}^n (P(A_{k+1}) - P(A_k)) \\
&= P(A_1) + \lim_{n \rightarrow \infty} (P(A_{n+1}) - P(A_n) + \dots + P(A_2) - P(A_1)) \\
&= P(A_1) + \lim_{n \rightarrow \infty} (P(A_{n+1}) - P(A_1)) \\
&= \lim_{n \rightarrow \infty} P(A_n).
\end{aligned}$$

La demostración es análoga en el caso $A_n \downarrow A = \cap_{n=1}^{\infty} A_n$. \square

Un resultado más general que el teorema 2.2 puede demostrarse en el caso de una sucesión convergente.

Teorema 2.3 Sea $\{A_n : n \geq 1\} \subset \mathcal{A}$ una sucesión convergente, donde $A = \lim_{n \rightarrow \infty} A_n$. Entonces, existe $\lim_{n \rightarrow \infty} P(A_n)$ y $P(A) = \lim_{n \rightarrow \infty} P(A_n)$.

2.3 Otras aproximaciones a la noción de probabilidad

La definición 2.4 parece indicar que un problema de naturaleza matemática sobre un experimento aleatorio comienza cuando el espacio de probabilidad (Ω, \mathcal{A}, P) está dado. Sin embargo, en la resolución de un problema hay un paso previo: el ajuste del modelo. Este paso consiste en extraer desde los mecanismos de azar que rigen el experimento aleatorio un espacio muestral Ω , una familia adecuada de sucesos \mathcal{A} y una asignación de probabilidades a los sucesos. Para realizar esta operación no hay una regla general. La práctica en la resolución de casos concretos permite desarrollar una intuición suficiente para enfrentarse con cualquier otro problema.

Para atribuir probabilidades a los sucesos relativos a un experimento aleatorio con espacio muestral Ω finito, existe una norma de utilidad cuando todos los sucesos elementales son igualmente probables; véase el ejemplo 2.2 en el caso de igualdad de verosimilitudes. La **regla de Laplace** aludida en el ejemplo 2.2 fue propuesta por P.S. Laplace (1749-1827) y representa el primer antecedente explícito del concepto de probabilidad.

La aplicación de la regla de Laplace puede presentar dificultades a la hora de comprobar la equiprobabilidad de los sucesos elementales. No obstante, en situaciones sencillas, la simetría de los resultados del experimento permite garantizar su equiprobabilidad.

Otra definición aceptada en la práctica (que puede formalizarse desde la definición 2.4 y resultados de convergencia del tema 5) es la definición **frecuentista**.

Ante un experimento aleatorio regido por el azar, la tendencia implica medir el grado de verosimilitud de los 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 experimento. Por ejemplo, supongamos una urna con 3 bolas idénticas (2 rojas y 1 blanca). Entonces,

$$\begin{aligned}P(\text{"Obtener bola roja"}) &= \frac{2}{3} = 0.6666\dots, \\P(\text{"Obtener bola blanca"}) &= \frac{1}{3} = 0.3333\dots,\end{aligned}$$

cuando se extrae al azar una bola. La frecuencia real se conoce con exactitud cuando, por ejemplo,

- (i) Repitiendo el experimento 12 veces, se obtienen 7 rojas y 5 blancas; es decir,

$$\begin{aligned}\text{frecuencia de "obtener bola roja"} &= \frac{7}{12} = 0.5833\dots \\ \text{frecuencia de "obtener bola blanca"} &= \frac{5}{12} = 0.4166\dots\end{aligned}$$

- (ii) Repitiendo el experimento 120 veces, se obtienen 69 rojas y 51 blancas; es decir,

$$\begin{aligned}\text{frecuencia de "obtener bola roja"} &= \frac{69}{120} = 0.575, \\ \text{frecuencia de "obtener bola blanca"} &= \frac{51}{120} = 0.425.\end{aligned}$$

- (iii) Repitiendo el experimento 1200 veces, se obtienen 822 rojas y 378 blancas; es decir,

$$\begin{aligned}\text{frecuencia de "obtener bola roja"} &= \frac{822}{1200} = 0.685, \\ \text{frecuencia de "obtener bola blanca"} &= \frac{378}{1200} = 0.715.\end{aligned}$$

Análogamente, el experimento podría repetirse 10000, 100000, ... veces, concluyendo que en una sucesión ilimitada de repeticiones, en idénticas condiciones, de un experimento aleatorio, las frecuencias tras cada repetición tienden a aproximarse hacia ciertos valores límites. El valor límite de la frecuencia de un suceso es lo que se desea expresar mediante su probabilidad frecuentista.

2.4 Probabilidad condicionada. Independencia de sucesos

En la sección 2.2, las probabilidades de los sucesos de \mathcal{A} han tenido un carácter estático; es decir, antes de realizar el experimento, es posible emitir un juicio

sobre su resultado, indicando la frecuencia de aparición de cada uno de los sucesos que pueden ocurrir.

No obstante, es frecuente considerar situaciones intermedias donde el experimento no ha concluido o, en el caso de experimentos que se desarrollan en varias etapas, nos interesamos por el resultado de una etapa inicial ya conociendo el resultado en una etapa posterior.

Ejemplo 2.8 Al lanzar dos veces un dado, la suma de las puntuaciones ha sido 8. Nos preguntamos por la probabilidad de que, en el primer lanzamiento, el resultado haya sido 2.

El espacio muestral es $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. No obstante, es más conveniente observar el árbol de resultados

1	---	1	Suma: 2	4	---	1	Suma: 5
	---	2	Suma: 3		---	2	Suma: 6
	---	3	Suma: 4		---	3	Suma: 7
	---	4	Suma: 5		---	4	Suma: 8
	---	5	Suma: 6		---	5	Suma: 9
	---	6	Suma: 7		---	6	Suma: 10
2	---	1	Suma: 3	5	---	1	Suma: 6
	---	2	Suma: 4		---	2	Suma: 7
	---	3	Suma: 5		---	3	Suma: 8
	---	4	Suma: 6		---	4	Suma: 9
	---	5	Suma: 7		---	5	Suma: 10
	---	6	Suma: 8		---	6	Suma: 11
3	---	1	Suma: 4	6	---	1	Suma: 7
	---	2	Suma: 5		---	2	Suma: 8
	---	3	Suma: 6		---	3	Suma: 9
	---	4	Suma: 7		---	4	Suma: 10
	---	5	Suma: 8		---	5	Suma: 11
	---	6	Suma: 9		---	6	Suma: 12

y tomar como espacio muestral

$$\Omega'' = \{(1, 1), \dots, (1, 6), (2, 1), \dots, (2, 6), \dots, (6, 1), \dots, (6, 6)\}.$$

El suceso "la suma de las puntuaciones es 8" viene dado por

$$S_8 = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$$

y el suceso "obtener 2 en el primer lanzamiento" es

$$A_2 = \{(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6)\}.$$

Sin tener en cuenta el resultado obtenido como suma de puntuaciones, la probabilidad del suceso A_2 es $P(A_2) = 1/6$.

Para calcular $P(A_2|S_8)$ ="probabilidad condicionada del suceso A_2 , dado el suceso S_8 " se requiere del concepto de **probabilidad condicionada** que se introduce a continuación.

Definición 2.5 (Probabilidad condicionada) *Dados el espacio de probabilidad (Ω, \mathcal{A}, P) y el suceso $A \in \mathcal{A}$ con $P(A) > 0$, se llama probabilidad condicionada del suceso $B \in \mathcal{A}$, dado el suceso A , al valor*

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

Ejemplo 2.8 (continuación) Tenemos que

$$P(A_2|S_8) = \frac{P(\{(2,6)\})}{P(S_8)} = \frac{1/36}{5/36} = \frac{1}{5} \neq \frac{1}{6} = P(A_2).$$

Teorema 2.4 Sean (Ω, \mathcal{A}, P) un espacio de probabilidad y $A \in \mathcal{A}$ tal que $P(A) > 0$. Entonces, $(\Omega, \mathcal{A}, P_A)$ es un espacio de probabilidad, siendo

$$P_A(B) = P(B|A), \quad B \in \mathcal{A}.$$

Demostración.— Es claro que $P_A(B) \geq 0$, $\forall B \in \mathcal{A}$, y que $P_A(\Omega) = P(\Omega \cap A)/P(A) = P(A)/P(A) = 1$. Además, si $\{A_n : n \geq 1\} \subset \mathcal{A}$ es tal que $A_i \cap A_j = \emptyset$, $\forall i \neq j$, entonces

$$\begin{aligned} P_A\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} A_n \middle| A\right) = \frac{P(\bigcup_{n=1}^{\infty} (A_n \cap A))}{P(A)} \\ &= \frac{\sum_{n=1}^{\infty} P(A_n \cap A)}{P(A)} = \sum_{n=1}^{\infty} P(A_n|A), \end{aligned}$$

donde $P(A_n|A) = P_A(A_n)$. \square

Similarmente, es posible tomar un nuevo espacio de probabilidad (A, \mathcal{A}_A, P_A) asociado a la medida de probabilidad P_A definida por

$$P_A(B) = P(B|A), \quad \forall B \in \mathcal{A}_A.$$

Desde la definición 2.5, observemos que para sucesos $A, B \in \mathcal{A}$ con $P(A) > 0$ y $P(B) > 0$, se tienen las igualdades

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A), \\ P(A \cap B) &= P(B)P(A|B). \end{aligned}$$

Estas fórmulas pueden ser extendidas a un número finito de conjuntos. Sean $\{A_1, \dots, A_n\} \subset \mathcal{A}$ tales que $P(\bigcap_{i=1}^{n-1} A_i) > 0$. Dado que

$$\bigcap_{i=1}^{n-1} A_i \subset \bigcap_{i=1}^{n-2} A_i \subset \dots \subset A_1 \cap A_2 \subset A_1,$$

se sigue que $P(\cap_{i=1}^{n-2} A_i) > 0$, ..., $P(A_1 \cap A_2) > 0$ y $P(A_1) > 0$, de modo que $P(A_k | \cap_{i=1}^{k-1} A_i)$ está bien definida para cada $2 \leq k \leq n$.

Teorema 2.5 (Regla de la multiplicación) Sean (Ω, \mathcal{A}, P) un espacio de probabilidad y $\{A_1, \dots, A_n\} \subset \mathcal{A}$ verificando $P(A_1 \cap \dots \cap A_{n-1}) > 0$. Entonces,

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1}). \quad (3)$$

Demostración.— El resultado se demuestra de forma inductiva. En primer lugar, notamos que el caso $n = 1$ es trivial.

Cuando $n = 2$, la definición 2.5 permite escribir

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)},$$

que equivale a $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$.

Asumamos como hipótesis de inducción que (3) es válida para $n - 1$; es decir,

$$P\left(\bigcap_{i=1}^{n-1} A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_{n-1}|A_1 \cap \dots \cap A_{n-2}). \quad (4)$$

Desde la definición de probabilidad condicionada de A_n , dado el suceso $\bigcap_{i=1}^{n-1} A_i$, se tiene

$$P(A_n|A_1 \cap \dots \cap A_{n-1}) = \frac{P(A_1 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})},$$

que conduce a

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1 \cap \dots \cap A_{n-1})P(A_n|A_1 \cap \dots \cap A_{n-1}). \quad (5)$$

Entonces, (3) se deduce sustituyendo (4) sobre (5). \square

Sean (Ω, \mathcal{A}, P) un espacio de probabilidad y $A, B \in \mathcal{A}$ sucesos con $P(B) > 0$. El teorema 2.5 establece que

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

Existen experimentos aleatorios donde la información que suministra el suceso B no afecta a la probabilidad de otro suceso A ; es decir, $P(A|B) = P(A)$. Esta relación refleja la idea de **independencia de sucesos**.

Definición 2.6 (Independencia de sucesos) Se dice que dos sucesos $A, B \in \mathcal{A}$ son independientes si y sólo si

$$P(A \cap B) = P(A)P(B). \quad (6)$$

La definición de independencia de los sucesos A y B es válida incluso si $P(A) = 0$ y/o $P(B) = 0$; en este caso, ambos miembros de (6) son nulos. Debemos observar que si A y B son sucesos independientes, entonces A y B^c también son independientes dado que

$$\begin{aligned} P(A \cap B^c) &= P(A \cap (\Omega - B)) = P(A - (A \cap B)) = P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) = P(A)P(B^c). \end{aligned}$$

Similarmente, se tiene que A^c y B son sucesos independientes, y que A^c y B^c son sucesos independientes.

Definición 2.7 Sean (Ω, \mathcal{A}, P) un espacio de probabilidad y $\mathcal{C} \subset \mathcal{A}$ una familia no vacía de subconjuntos.

(a) Se dice que los sucesos de \mathcal{C} son independientes dos-a-dos si y sólo si

$$\forall A, B \in \mathcal{C} \text{ con } A \neq B, \quad P(A \cap B) = P(A)P(B).$$

(b) Se dice que la familia \mathcal{C} es (mutuamente) independiente si y sólo si

$$\forall \{A_{i_1}, \dots, A_{i_n}\} \subset \mathcal{C} \text{ con } A_{i_j} \neq A_{i_k}, \quad P(A_{i_1} \cap \dots \cap A_{i_n}) = \prod_{k=1}^n P(A_{i_k}).$$

Notemos que la relación de independencia entre sucesos puede ser consecuencia lógica de las características de un experimento aleatorio o, simplemente, se debe a una coincidencia numérica.

Ejemplo 2.9 Se lanzan una moneda y un dado perfectamente equilibrados. El espacio muestral viene dado por $\Omega = \{C1, C2, \dots, C6, X1, X2, \dots, X6\}$ y es inmediato que

$$\begin{aligned} P(\text{"Obtener cara"}) &= \frac{6}{12} = \frac{1}{2}, \\ P(\text{"Obtener al menos 3"}) &= \frac{8}{12} = \frac{2}{3}, \\ P(\text{"Obtener cara y al menos 3"}) &= \frac{4}{12} = \frac{1}{3}. \end{aligned}$$

La igualdad $P(\text{"Obtener cara y al menos 3"}) = P(\text{"Obtener cara"})P(\text{"Obtener al menos 3"})$ establece que ambos sucesos son independientes.

La independencia *obvia* que existe entre cualquier suceso relativo a la moneda y cualquier suceso relativo al dado se convierte en una justificación alternativa de la atribución de probabilidad $1/12$ a cada suceso elemental de Ω ; por ejemplo,

$$P(\{X4\}) = P(\{X\})P(\{4\}) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}.$$

El procedimiento seguido al final del ejemplo 2.9 da lugar a un método usual de construcción de modelos probabilísticos. Puede ser empleado cuando el experimento aleatorio consta de varias componentes, sin aparente relación entre ellas, es decir, de resultados a priori independientes y con único vínculo entre ellas, el hecho de que son partes de un mismo experimento aleatorio.

2.5 Espacio producto

Comenzamos introduciendo algunos elementos de interés.

Definición 2.8

- (a) *(Producto cartesiano)* Dados dos espacios muestrales Ω_1 y Ω_2 , el producto cartesiano $\Omega_1 \times \Omega_2$ viene dado por $\{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$.
- (b) *(Rectángulo)* Cualquier conjunto en $\Omega_1 \times \Omega_2$ de la forma $A \times B$, donde $A \subset \Omega_1$ y $B \subset \Omega_2$, se denomina rectángulo.
- (c) *(Producto de clases)* Sean \mathcal{A}_1 y \mathcal{A}_2 dos clases de $\mathcal{P}(\Omega_1)$ y $\mathcal{P}(\Omega_2)$, respectivamente. El producto de las clases \mathcal{A}_1 y \mathcal{A}_2 es la clase de rectángulos

$$\mathcal{A}_1 \times \mathcal{A}_2 = \{A \times B : A \in \mathcal{A}_1, B \in \mathcal{A}_2\}.$$

Una cuestión de interés es relativa al hecho de que, supuesto que \mathcal{A}_1 y \mathcal{A}_2 son cerradas bajo la operación unión, el producto de clases $\mathcal{A}_1 \times \mathcal{A}_2$ no es necesariamente cerrado bajo uniones. Como resultado, en el caso de que $\mathcal{A}_1 \subset \mathcal{P}(\Omega_1)$ y $\mathcal{A}_2 \subset \mathcal{P}(\Omega_2)$ sean dos σ -álgebras, no es posible asegurar que $\mathcal{A}_1 \times \mathcal{A}_2 \subset \mathcal{P}(\Omega_1 \times \Omega_2)$ tenga estructura de σ -álgebra. En lo sucesivo, denotamos por

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\mathcal{A}_1 \times \mathcal{A}_2)$$

la σ -álgebra producto de $\mathcal{A}_1 \times \mathcal{A}_2$; es decir, la σ -álgebra engendrada por $\mathcal{A}_1 \times \mathcal{A}_2$.

Definición 2.9 Sea $E \subset \Omega_1 \times \Omega_2$.

- (a) *(Sección)* Se llama sección de E en $\omega_1 \in \Omega_1$ al conjunto $E_{\omega_1} = \{\omega_2 : (\omega_1, \omega_2) \in E\}$. Análogamente, la sección de E en $\omega_2 \in \Omega_2$ viene definida como el conjunto $E^{\omega_2} = \{\omega_1 : (\omega_1, \omega_2) \in E\}$.
- (b) Se llaman proyecciones de E sobre los espacios muestrales Ω_1 y Ω_2 , respectivamente, a los conjuntos

$$\begin{aligned} \{\omega_1 \in \Omega_1 : \exists \omega_2 \in \Omega_2 \text{ tal que } (\omega_1, \omega_2) \in E\}, \\ \{\omega_2 \in \Omega_2 : \exists \omega_1 \in \Omega_1 \text{ tal que } (\omega_1, \omega_2) \in E\}. \end{aligned}$$

Es interesante observar que las secciones de conjuntos medibles son conjuntos medibles.

Teorema 2.6 Sean $\mathcal{A}_1 \subset \mathcal{P}(\Omega_1)$ y $\mathcal{A}_2 \subset \mathcal{P}(\Omega_2)$ σ -álgebras y $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Entonces, $E_{\omega_1} \in \mathcal{A}_2$, para cada $\omega_1 \in \Omega_1$, y $E^{\omega_2} \in \mathcal{A}_1$, para cada $\omega_2 \in \Omega_2$.

En el ejemplo 2.9, se asumen dos espacios de probabilidad $(\Omega_1, \mathcal{A}_1, P_1)$ y $(\Omega_2, \mathcal{A}_2, P_2)$. Entonces, mediante la realización de los experimentos aleatorios asociados, hemos construido el espacio muestral producto $\Omega = \Omega_1 \times \Omega_2$, la σ -álgebra producto $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ y una medida de probabilidad sobre (Ω, \mathcal{A}) que sobre los rectángulos medibles satisface

$$P(A \times B) = P_1(A)P_2(B), \quad A \in \mathcal{A}_1, B \in \mathcal{A}_2.$$

El espacio de probabilidad (Ω, \mathcal{A}, P) es denominado espacio de probabilidad producto de $(\Omega_1, \mathcal{A}_1, P_1)$ y $(\Omega_2, \mathcal{A}_2, P_2)$. La construcción de un espacio producto se extiende sin dificultad al caso de experimentos compuestos por más de dos subexperimentos, entre los que la única conexión es su observación conjunta.

Una cuestión que subyace en la construcción del espacio producto es la existencia de la medida producto P , no sólo sobre los rectángulos.

Sean $(\Omega_1, \mathcal{A}_1, P_1)$ y $(\Omega_2, \mathcal{A}_2, P_2)$ dos espacios de probabilidad y

$$\mathcal{F} = \left\{ \bigcup_{i=1}^n (A_1^i \times A_2^i) : A_1^i \times A_2^i \text{ rectángulos de } \mathcal{A}_1 \times \mathcal{A}_2 \text{ disjuntos dos-a-dos} \right\}.$$

Entonces, es sencillo demostrar que la clase $\mathcal{F} \subset \mathcal{P}(\Omega_1 \times \Omega_2)$ tiene estructura de álgebra. Además, la función de conjunto

$$P_{\mathcal{F}} : \begin{array}{ll} \mathcal{F} & \longrightarrow [0, 1] \\ \bigcup_{i=1}^n (A_1^i \times A_2^i) & \longrightarrow \sum_{i=1}^n P_1(A_1^i)P_2(A_2^i) \end{array}$$

es una medida de probabilidad sobre $(\Omega_1 \times \Omega_2, \mathcal{F})$, con las salvedades necesarias asociadas a que \mathcal{F} no es necesariamente una σ -álgebra.

Teorema 2.7 (De extensión) En las condiciones anteriores, existe una única medida de probabilidad $P = P_1 \otimes P_2$ sobre $\sigma(\mathcal{F}) = \mathcal{A}_1 \otimes \mathcal{A}_2$ que satisface

$$P(A_1 \times A_2) = P_1(A_1)P_2(A_2), \quad A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2.$$

Desde el teorema 2.7 queda establecida formalmente la noción de **espacio de probabilidad producto**.

Definición 2.10 (Espacio de probabilidad producto) Se llama espacio de probabilidad producto de $(\Omega_1, \mathcal{A}_1, P_1)$ y $(\Omega_2, \mathcal{A}_2, P_2)$ a la terna (Ω, \mathcal{A}, P) , donde $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ y P es la medida de probabilidad del teorema 2.7.

Con los elementos anteriores estamos en disposición de introducir la noción de experimentos aleatorios independientes.

Sea un espacio de probabilidad $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P)$, donde no necesariamente se asume $P = P_1 \otimes P_2$. Definimos $P_1 : \mathcal{A}_1 \rightarrow [0, 1]$ y $P_2 : \mathcal{A}_2 \rightarrow [0, 1]$ en términos de

$$\begin{aligned} P_1(A_1) &= P(A_1 \times \Omega_2), \\ P_2(A_2) &= P(\Omega_1 \times A_2). \end{aligned}$$

Entonces, es sencillo comprobar que P_1 y P_2 son medidas de probabilidad sobre $(\Omega_1, \mathcal{A}_1)$ y $(\Omega_2, \mathcal{A}_2)$, respectivamente.

Definición 2.11 *En las condiciones anteriores, los espacios de probabilidad $(\Omega_1, \mathcal{A}_1, P_1)$ y $(\Omega_2, \mathcal{A}_2, P_2)$ se dicen independientes si y sólo si $P = P_1 \otimes P_2$; es decir, para cualesquiera $A_1 \in \mathcal{A}_1$ y $A_2 \in \mathcal{A}_2$, se tiene $P(A_1 \times A_2) = P_1(A_1)P_2(A_2)$.*

Podemos definir, para cada $A_1 \in \mathcal{A}_1$ fijo con $P_1(A_1) > 0$, la función de conjunto

$$P_2(A_2|A_1) = P(\Omega_1 \times A_2 | A_1 \times \Omega_2)$$

que resulta ser una medida de probabilidad sobre $(\Omega_2, \mathcal{A}_2)$. Además, se tiene

$$\begin{aligned} P(A_1 \times A_2) &= P(A_1 \times \Omega_2)P(\Omega_1 \times A_2 | A_1 \times \Omega_2) \\ &= P_1(A_1)P_2(A_2|A_1), \end{aligned}$$

que no necesariamente coincide con $P_1(A_1)P_2(A_2)$.

Definición 2.12 *En las condiciones anteriores, si para algún rectángulo $A_1 \times A_2 \in \mathcal{A}_1 \times \mathcal{A}_2$ se tiene $P(A_1 \times A_2) \neq P_1(A_1)P_2(A_2)$, entonces se dice que $(\Omega_1, \mathcal{A}_1, P_1)$ y $(\Omega_2, \mathcal{A}_2, P_2)$ son espacios de probabilidad dependientes.*

Aunque supera los objetivos de la asignatura, debemos notar que en ocasiones es necesario considerar un espacio de probabilidad producto construido desde una cantidad numerable de espacios de probabilidad.

2.6 Teorema de la probabilidad total. Teorema de Bayes

En la sección 2.4 hemos mostrado que, supuesto dado un suceso $A \in \mathcal{A}$ con $P(A) > 0$, es posible reasignar probabilidades a los sucesos de \mathcal{A} desde la nueva medida de probabilidad $P_A(\cdot) = P(\cdot|A)$; es decir, nos referimos a reemplazar el espacio de probabilidad original (Ω, \mathcal{A}, P) por $(\Omega, \mathcal{A}, P_A)$ o alternativamente por (A, \mathcal{A}_A, P_A) .

Las propiedades de mayor interés de la noción de probabilidad condicionada surgen ante la posibilidad de condicionar por múltiples sucesos.

Teorema 2.8 (De la probabilidad total) Sean (Ω, \mathcal{A}, P) un espacio de probabilidad y $\{B_n : n \geq 1\} \subset \mathcal{A}$ tal que $B_i \cap B_j = \phi, \forall i \neq j$, $\Omega = \cup_{n=1}^{\infty} B_n$ y $P(B_n) > 0, \forall n \geq 1$. Entonces,

$$P(A) = \sum_{n=1}^{\infty} P(B_n)P(A|B_n), \quad A \in \mathcal{A}.$$

Demostración. – Es posible representar $A \in \mathcal{A}$ de la forma

$$A = A \cap \Omega = \bigcup_{n=1}^{\infty} (A \cap B_n),$$

siendo $(A \cap B_i) \cap (A \cap B_j) = \phi, \forall i \neq j$. Por este motivo, $P(A) = \sum_{n=1}^{\infty} P(A \cap B_n) = \sum_{n=1}^{\infty} P(B_n)P(A|B_n)$. La última igualdad tiene sentido debido a que $P(B_n) > 0, \forall n \geq 1$. \square

Podemos interpretar la fórmula $P(A) = \sum_{n=1}^{\infty} P(B_n)P(A|B_n)$ diciendo que los sucesos B_n son las *circunstancias* que hacen más o menos posible la aparición del suceso A . Entonces, los factores $P(B_n)$ valoran la probabilidad de que se presente cada circunstancia B_n y los factores $P(A|B_n)$ miden la probabilidad del efecto A , cuando se tienen las circunstancias representadas por B_n .

Ejemplo 2.10 Se emplea un análisis clínico en el diagnóstico de tres enfermedades B_1, B_2 y B_3 . En la población, la proporción de enfermos de cada una es 3%, 2% y 1%, respectivamente. Por otra parte, el análisis genera un resultado positivo en el 85% de los enfermos de B_1 , el 92% de los enfermos de B_2 y el 78% de los enfermos de B_3 , mientras que sólo genera positivo en un 0.5% de las personas no afectadas (B_0).

Si A representa el suceso "el análisis ofrece un resultado positivo", entonces

$$\begin{aligned} P(A) &= P(B_0)P(A|B_0) + P(B_1)P(A|B_1) + P(B_2)P(A|B_2) \\ &\quad + P(B_3)P(A|B_3) \\ &= 0.94 \cdot 0.005 + 0.03 \cdot 0.85 + 0.02 \cdot 0.92 + 0.01 \cdot 0.78 \\ &= 0.0564. \end{aligned}$$

Debemos notar que $P(B_0) > 0, P(B_1) > 0, P(B_2) > 0, P(B_3) > 0$ y $P(B_0) + P(B_1) + P(B_2) + P(B_3) = 1$.

En el ejemplo 2.10 hemos centrado la atención en evaluar la probabilidad del suceso A desde la valoración de la probabilidad condicionada de A , dado cada suceso B_n (circunstancia). En una situación dual estaríamos interesados

en determinar las probabilidades de las circunstancias B_n cuando se sabe que ha ocurrido el suceso A .

Teorema 2.9 (Bayes) Sean (Ω, \mathcal{A}, P) un espacio de probabilidad, $\{B_n : n \geq 1\} \subset \mathcal{A}$ tal que $B_i \cap B_j = \emptyset, \forall i \neq j$, con $P(B_n) > 0, \forall n \geq 1$, y $\Omega = \bigcup_{n=1}^{\infty} B_n$. Para cada suceso $A \in \mathcal{A}$ con $P(A) > 0$, se verifica

$$P(B_n|A) = \frac{P(B_n)P(A|B_n)}{\sum_{k=1}^{\infty} P(B_k)P(A|B_k)}.$$

Es decir, $P(B_n|A) = \frac{P(B_n)P(A|B_n)}{P(A)}$.

Demostración.– En primer lugar, notemos que $P(A) > 0$ permite escribir

$$P(B_n|A) = \frac{P(A \cap B_n)}{P(A)},$$

donde $P(A \cap B_n) = P(B_n)P(A|B_n)$ dado que $P(B_n) > 0$. Ahora, desde el teorema 2.8, se tiene $P(A) = \sum_{k=1}^{\infty} P(B_k)P(A|B_k)$. \square

Ejemplo 2.10 (continuación) Si se asume que el análisis clínico ha dado positivo, entonces las probabilidades de que el paciente no esté afectado y de que padezca cada una de las enfermedades son

$$\begin{aligned} P(B_0|A) &= \frac{P(B_0)P(A|B_0)}{P(A)} = \frac{0.94 \cdot 0.005}{0.0564} = 0.083, \\ P(B_1|A) &= \frac{P(B_1)P(A|B_1)}{P(A)} = 0.452, \\ P(B_2|A) &= \frac{P(B_2)P(A|B_2)}{P(A)} = 0.326, \\ P(B_3|A) &= \frac{P(B_3)P(A|B_3)}{P(A)} = 0.138. \end{aligned}$$

En este esquema, el suceso A transforma el conocimiento de los sucesos B_n . En concreto,

Prob. a priori	Prob. a posteriori
$P(B_0) = 0.94$	$P(B_0 A) = 0.083$
$P(B_1) = 0.03$	$P(B_1 A) = 0.452$
$P(B_2) = 0.02$	$P(B_2 A) = 0.326$
$P(B_3) = 0.01$	$P(B_3 A) = 0.138$

En particular, la proporción de individuos sanos es el 94% en el conjunto de la población, pero sólo es el 8.3% entre aquéllos que dan positivo en el análisis clínico.

Tema 3. Variables aleatorias unidimensionales

Indice

- 3.1 Introducción
- 3.2 Funciones medibles y variables aleatorias. Resultados básicos
- 3.3 Función de distribución
- 3.4 Variables aleatorias discretas y continuas
- 3.5 Transformaciones medibles
- 3.6 Momentos y cuantiles
- 3.7 Función característica y función generatriz de momentos
- 3.8 Distribuciones discretas
- 3.9 Distribuciones absolutamente continuas
- Ejercicios

3.1 Introducción

Con frecuencia es conveniente sustituir el espacio muestral Ω_1 asociado a un experimento aleatorio por otro espacio muestral Ω_2 . Por ejemplo, en el lanzamiento de dos monedas equilibradas, el espacio muestral viene dado por

$$\Omega_1 = \{CC, CX, XC, XX\},$$

la σ -álgebra es $\mathcal{A}_1 = \mathcal{P}(\Omega_1)$ y la medida de probabilidad P_1 es tal que $P_1(\{\omega\}) = 1/4$, para cada $\omega \in \Omega_1$. Ahora, consideremos el espacio muestral $\Omega_2 = \{0, 1, 2\}$ y la σ -álgebra $\mathcal{A}_2 = \mathcal{P}(\Omega_2)$. Entonces, podemos definir la aplicación $Y : \Omega_1 \rightarrow \Omega_2$ mediante $Y(\omega) = \text{'número de caras en el lanzamiento con el resultado } \omega\text{'}$; es decir,

$$\begin{aligned} Y(CC) &= 2, & Y(CX) &= 1, \\ Y(XC) &= 1, & Y(XX) &= 0. \end{aligned}$$

Debemos notar que

- (i) Y no es una aplicación biyectiva.
- (ii) Si definimos $Y^{-1}(B) = \{\omega \in \Omega_1 : Y(\omega) \in B\}$, entonces $Y^{-1}(\{0\}) = \{XX\}$, $Y^{-1}(\{1\}) = \{CX, XC\}$ e $Y^{-1}(\{2\}) = \{CC\}$.
- (iii) Para $B \in \mathcal{A}_2$, $Y^{-1}(B) \in \mathcal{A}_1$.

La aplicación Y permite "probabilizar" el espacio medible $(\Omega_2, \mathcal{A}_2)$. En particular, nos referimos a

$$Y : (\Omega_1, \mathcal{A}_1, P_1) \rightarrow (\Omega_2, \mathcal{A}_2, P_2),$$

donde $P_2(B) = P_1(Y^{-1}(B))$, $\forall B \in \mathcal{A}_2$. Si aplicamos esta definición, entonces

$$\begin{aligned} P_2(\{0\}) &= P_1(Y^{-1}(\{0\})) = P_1(\{XX\}) = \frac{1}{4}, \\ P_2(\{1\}) &= P_1(Y^{-1}(\{1\})) = P_1(\{CX, XC\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \\ P_2(\{2\}) &= P_1(Y^{-1}(\{2\})) = P_1(\{CC\}) = \frac{1}{4}. \end{aligned}$$

La función Y es llamada "variable aleatoria" y es un caso particular de la noción de "aplicación medible". La familia de las aplicaciones medibles permite transformar espacios de probabilidad y, en el caso de las funciones medibles, constituye una clase de funciones estable respecto a las operaciones aritméticas, la formación de ínfimos y supremos, y la toma de límites de sucesiones.

El presente tema aborda el estudio de las variables aleatorias y sus principales propiedades. La sección 3.2 define las aplicaciones medibles, las funciones medibles y las variables aleatorias, exponiendo las propiedades de mayor interés. La sección 3.3 introduce la noción de función de distribución, siendo ésta la base para distinguir, en la sección 3.4, entre variables aleatorias discretas y continuas. La transformación medible de variables aleatorias es estudiada en la sección 3.5. Las principales características o parámetros de las variables aleatorias, tales como los momentos y los cuantiles, son propuestos en la sección 3.6. La sección 3.7 está dedicada a caracterizar la distribución de una variable aleatoria en términos de la función característica y la función generatriz de momentos. Para concluir, las secciones 3.8 y 3.9 recopilan las distribuciones discretas y absolutamente continuas más notables.

3.2 Funciones medibles y variables aleatorias. Resultados básicos

Sean Ω_1 y Ω_2 espacios muestrales arbitrarios y $f : \Omega_1 \rightarrow \Omega_2$ una aplicación. Desde f es posible definir la función de conjunto $f^{-1} : \mathcal{P}(\Omega_2) \rightarrow \mathcal{P}(\Omega_1)$ de forma que $\forall B \in \mathcal{P}(\Omega_2)$

$$f^{-1}(B) = \{\omega \in \Omega_1 : f(\omega) \in B\}.$$

Entonces, es sencillo comprobar que

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

y, como consecuencia, establecer que si $\mathcal{C} \subset \mathcal{P}(\Omega_2)$ es álgebra o σ -álgebra, entonces $f^{-1}(\mathcal{C}) \subset \mathcal{P}(\Omega_1)$ es también álgebra o σ -álgebra, respectivamente.

Definición 3.1 Sean $(\Omega_1, \mathcal{A}_1)$ y $(\Omega_2, \mathcal{A}_2)$ espacios medibles.

- (a) (Aplicación medible) Se dice que $f : \Omega_1 \rightarrow \Omega_2$ es una aplicación medible si y sólo si $f^{-1}(B) \in \mathcal{A}_1, \forall B \in \mathcal{A}_2$.
- (b) (Función medible) Se llama función medible a toda aplicación medible $f : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ donde $(\Omega_2, \mathcal{A}_2) = (\mathbb{R}^n, \beta^n)$ para algún $n \geq 1$.

Definición 3.2 (Variable aleatoria) Una variable aleatoria es una función medible $f : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \beta)$; es decir, donde el espacio inicial es un espacio de probabilidad.

Ejemplo 3.1 Dos jugadores A y B lanzan un dado cada uno, de manera que se obtiene un resultado del espacio muestral

$$\Omega = \{(1, 1), \dots, (1, 6), \dots, (6, 1), \dots, (6, 6)\}.$$

La primera componente representa el resultado obtenido por A y la segunda el obtenido por B . Entonces, la variable aleatoria

$$\begin{aligned} X_A : \quad \Omega &\rightarrow \{1, 2, 3, 4, 5, 6\}, \\ (i, j) &\rightarrow X_A(i, j) = i, \end{aligned}$$

describe "la puntuación obtenida por A ". La función

$$Y(i, j) = \begin{cases} i, & \text{si } i \geq j, \\ j, & \text{si } i < j, \end{cases}$$

es la variable aleatoria que representa "la puntuación mayor". También es una variable aleatoria

$$Z(i, j) = \begin{cases} 1, & \text{si } i > j, \\ 0, & \text{si } i = j, \\ -1, & \text{si } i < j, \end{cases}$$

y representa "la ganancia del jugador A ", cuando el juego consiste en ganar una moneda cuando se obtiene el mayor resultado, perder una moneda cuando se obtiene el menor resultado y mantener la riqueza en caso de empate.

En el ejemplo 3.1, la comprobación asociada a que X_A es variable aleatoria implica tomar X_A como

$$X_A : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \beta),$$

donde $\mathcal{A} = \mathcal{P}(\Omega)$ y

$$\begin{aligned} X_A^{-1}(\{i\}) &= \{(i, 1), (i, 2), (i, 3), (i, 4), (i, 5), (i, 6)\} \in \mathcal{A}, \quad \forall i \in \{1, \dots, 6\}, \\ X_A^{-1}(B) &= \phi \in \mathcal{A}, \quad \text{si } i \notin B, \quad \forall i \in \{1, \dots, 6\}, \\ X_A^{-1}(B) &= \bigcup_{i \in B} X_A^{-1}(\{i\}) \in \mathcal{A}, \quad \text{si existe } i_0 \in \{1, \dots, 6\} \text{ tal que } i_0 \in B. \end{aligned}$$

Con generalidad, es necesario disponer de herramientas que permitan desarrollar esta comprobación con sencillez para una variable aleatoria arbitraria.

Definición 3.3 Sean $\mathcal{D} \subset \mathcal{P}(\Omega_2)$, $(\Omega_1, \mathcal{A}_1)$ un espacio medible y $f : \Omega_1 \rightarrow \Omega_2$ una aplicación. Se dice que f es \mathcal{D} -medible si y sólo si $f^{-1}(D) \in \mathcal{A}_1$, $\forall D \in \mathcal{D}$.

Proposición 3.1 Sean $f : \Omega_1 \rightarrow \Omega_2$ una aplicación y \mathcal{C} una clase de subconjuntos de Ω_2 . Entonces,

$$\sigma(f^{-1}(\mathcal{C})) = f^{-1}(\sigma(\mathcal{C})).$$

Demostración.— La demostración es propuesta en la Hoja 3. \square

Teorema 3.1 Una condición necesaria y suficiente para que $f : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ sea medible es que exista una clase $\mathcal{D} \subset \mathcal{P}(\Omega_2)$ tal que $\mathcal{A}_2 = \sigma(\mathcal{D})$ y f sea \mathcal{D} -medible.

Demostración.— Si f es medible, entonces es suficiente con tomar $\mathcal{D} = \mathcal{A}_2$, siendo el resultado trivial. Para demostrar el recíproco, asumamos que existe una clase $\mathcal{D} \subset \mathcal{P}(\Omega_2)$ verificando $\mathcal{A}_2 = \sigma(\mathcal{D})$ y tal que f es \mathcal{D} -medible. Entonces, $f^{-1}(\mathcal{D}) \subset \mathcal{A}_1$ y, desde la proposición 3.1, se sigue que $f^{-1}(\sigma(\mathcal{D})) = \sigma(f^{-1}(\mathcal{D})) \subset \mathcal{A}_1$; es decir, $f^{-1}(\mathcal{A}_2) \subset \mathcal{A}_1$ y, por lo tanto, f es medible. \square

El teorema 3.1 aplicado al caso de variables aleatorias permite comprobar que una función f es una variable aleatoria, demostrando que cualquiera de las siguientes clases está contenida en \mathcal{A}_1 :

$$\begin{aligned} \mathcal{Q}_1 &= \{f^{-1}(-\infty, c] = \{\omega \in \Omega : f(\omega) \leq c\}, \forall c \in \mathbb{R}\}, \\ \mathcal{Q}_2 &= \{f^{-1}(-\infty, c) = \{\omega \in \Omega : f(\omega) < c\}, \forall c \in \mathbb{R}\}, \\ \mathcal{Q}_3 &= \{f^{-1}(c, \infty) = \{\omega \in \Omega : f(\omega) > c\}, \forall c \in \mathbb{R}\}. \end{aligned}$$

A continuación proponemos algunos resultados relativos a las operaciones con funciones medibles.

Teorema 3.2 Sean (Ω, \mathcal{A}) , $(\Omega_1, \mathcal{A}_1)$ y $(\Omega_2, \mathcal{A}_2)$ espacios medibles.

- (a) Si $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \beta)$ es continua, entonces f es medible.
- (b) Una función $f = (f_1, \dots, f_n) : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^n, \beta^n)$ es medible si y sólo si f_i es medible, $\forall 1 \leq i \leq n$.
- (c) Sean las aplicaciones medibles $f : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ y $g : (\Omega_2, \mathcal{A}_2) \rightarrow (\Omega_3, \mathcal{A}_3)$. Entonces $g \circ f : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_3, \mathcal{A}_3)$ es medible.
- (d) Sean f y g funciones medibles definidas desde (Ω, \mathcal{A}) en (\mathbb{R}, β) . Entonces, las funciones $f + g$, αf (para $\alpha \in \mathbb{R}$) y fg son medibles. Además, f/g es medible sobre su dominio de definición.
- (e) Sean f y g funciones medibles definidas desde (Ω, \mathcal{A}) en (\mathbb{R}, β) . Entonces, $\max(f, g)$, $\min(f, g)$, $f^+ = \max(f, 0)$ y $f^- = \max(-f, 0)$ son funciones medibles.
- (f) Si $\{f_n : n \geq 1\}$ es una sucesión de funciones medibles definidas desde (Ω, \mathcal{A}) en (\mathbb{R}, β) , entonces las funciones $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$ y $\liminf_n f_n$ son medibles. Si existe $\lim_{n \rightarrow \infty} f_n$, entonces $\lim_{n \rightarrow \infty} f_n$ es medible.

Demostración.— (Por conveniencia, sólo probamos algunos de los apartados del teorema.)

- (b) Si f es medible, entonces

$$f^{-1}((-\infty, x_1] \times \dots \times (-\infty, x_n]) \in \mathcal{A}, \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Por tanto,

$$f_i^{-1}((-\infty, x_i]) = \bigcup_{k=1}^{\infty} \{\omega \in \Omega : f_i(\omega) \leq x_i \text{ y } f_j(\omega) \leq k, \forall j \neq i\} \in \mathcal{A}$$

y, como resultado, f_i es medible, para $1 \leq i \leq n$. Recíprocamente, si f_i es medible, para cada $1 \leq i \leq n$, entonces

$$(f_1, \dots, f_n)^{-1}((-\infty, x_1] \times \dots \times (-\infty, x_n]) = \bigcap_{i=1}^n \{\omega \in \Omega : f_i(\omega) \leq x_i\} \in \mathcal{A},$$

y, desde el teorema 3.1, $f = (f_1, \dots, f_n)$ es medible.

(c) Sea $A \in \mathcal{A}_3$. Por ser g medible, $g^{-1}(A) \in \mathcal{A}_2$ y, por ser f medible, tenemos $f^{-1}(g^{-1}(A)) = (g \circ f)^{-1}(A) \in \mathcal{A}_1$. Es decir, $g \circ f$ es medible.

- (d) Si f y g son medibles, la función

$$\begin{aligned} F : (\Omega, \mathcal{A}) &\longrightarrow (\mathbb{R}^2, \beta^2) \\ \omega &\longrightarrow F(\omega) = (f(\omega), g(\omega)) \end{aligned}$$

es medible, desde (b). Entonces, las funciones $G_1, G_2, G_3 : \mathbb{R}^2 \rightarrow \mathbb{R}$ definidas por $G_1(x, y) = \alpha x$, $G_2(x, y) = x + y$ y $G_3(x, y) = xy$ son continuas y, por tanto, medibles. En definitiva, $\alpha f = G_1 \circ F$, $f + g = G_2 \circ F$ y $f \cdot g = G_3 \circ F$ son

medibles. Por otra parte, el conjunto definido como el dominio de definición de f/g viene dado por

$$A = \{\omega \in \Omega : g(\omega) > 0\} \cup \{\omega \in \Omega : g(\omega) < 0\} \in \mathcal{A}.$$

Entonces, es posible establecer que

$$\begin{aligned} \left\{ \omega \in \Omega : \left(\frac{f}{g} \right) (\omega) < t \right\} &= (\{\omega \in \Omega : g(\omega) > 0\} \cap \{\omega \in \Omega : f(\omega) < tg(\omega)\}) \\ &\cup (\{\omega \in \Omega : g(\omega) < 0\} \cap \{\omega \in \Omega : f(\omega) > tg(\omega)\}) \in \mathcal{A}, \end{aligned}$$

puesto que la función tg es medible y $f - tg$ es medible.

(e) Se demuestra observando que

$$\begin{aligned} \max(f, g) &= \frac{f + g + |f - g|}{2}, \\ \min(f, g) &= \frac{f + g - |f - g|}{2}. \end{aligned}$$

(f) Sea $a \in \mathbb{R}$. Entonces,

$$\begin{aligned} \left(\sup_n f_n \right)^{-1} (-\infty, a] &= \{\omega \in \Omega : (\sup_n f_n)(\omega) \leq a\} \\ &= \{\omega \in \Omega : \sup_n f_n(\omega) \leq a\} \\ &= \{\omega \in \Omega : f_n(\omega) \leq a, \forall n \geq 1\} \\ &= \bigcap_{n=1}^{\infty} \{\omega \in \Omega : f_n(\omega) \leq a\} \\ &= \bigcap_{n=1}^{\infty} f_n^{-1}(-\infty, a] \in \mathcal{A}. \end{aligned}$$

Es decir, $\sup_n f_n$ es medible. Dado que $\inf_n f_n = -\sup_n(-f_n)$, se tiene que $\inf_n f_n$ es medible. Finalmente, $\limsup f_n = \inf_n \sup_{k \geq n} f_k$ y $\liminf f_n = \sup_n \inf_{k \geq n} f_k$ son medibles. \square

El teorema 3.3 caracteriza a las funciones medibles como límites de sucesiones de funciones simples.

Definición 3.4 Una función medible $f : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \beta)$ se dice simple si $\exists \{A_1, \dots, A_n\} \subset \mathcal{A}$, con $A_i \cap A_j = \emptyset$, $\forall i \neq j$, y $\Omega = \cup_{i=1}^n A_i$, tal que

$$f(\omega) = \sum_{i=1}^n x_i I_{A_i}(\omega), \quad \forall \omega \in \Omega,$$

con $x_i \in \mathbb{R}$, donde $I_A(\cdot)$ es la función indicadora¹ del conjunto $A \in \mathcal{A}$.

¹La función indicadora de un conjunto $A \in \mathcal{A}$, $I_A : \Omega \longrightarrow \mathbb{R}$, se define como $I_A(\omega) = 1$ si $\omega \in A$, y 0 si $\omega \notin A$.

Es sencillo demostrar que si $f, g : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \beta)$ son funciones simples, entonces $f + g$, αf , fg , $\max(f, g)$ y $\min(f, g)$ son también funciones simples.

Teorema 3.3 (Caracterización)

- (a) Una función no-negativa $f : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \beta)$ es medible si y sólo si $\exists \{f_n : n \geq 1\}$ sucesión no-decreciente de funciones simples no-negativas finito-valoradas tal que $\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega)$, $\forall \omega \in \Omega$.
- (b) Una función $f : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \beta)$ es medible si y sólo si $\exists \{f_n : n \geq 1\}$ sucesión de funciones simples finito-valoradas con $|f_n| \leq |f|$, $\forall n \geq 1$, tal que $\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega)$, $\forall \omega \in \Omega$.

Demostración.– (a) La demostración hace uso de la sucesión

$$f_n(\omega) = \begin{cases} \frac{k-1}{2^n}, & \text{si } \frac{k-1}{2^n} \leq f(\omega) < \frac{k}{2^n}, \quad 1 \leq k \leq n2^n, \\ n, & \text{si } f(\omega) \geq n. \end{cases}$$

(b) Se sigue desde el apartado (a); es decir, las sucesiones $\{s_n : n \geq 1\}$ y $\{t_n : n \geq 1\}$ no-decrecientes de funciones simples no-negativas finito-valoradas tales que $\lim_{n \rightarrow \infty} s_n(\omega) = f^+(\omega)$ y $\lim_{n \rightarrow \infty} t_n(\omega) = f^-(\omega)$, $\forall \omega \in \Omega$. \square

3.3 Función de distribución

Sea (\mathbb{R}, β, P) un espacio de probabilidad. Sobre (\mathbb{R}, β, P) juegan un papel destacado las denominadas funciones de distribución.

Definición 3.5 (Función de distribución) Se llama función de distribución sobre \mathbb{R} a toda aplicación $F : \mathbb{R} \rightarrow \mathbb{R}$ tal que

- (i) F es monótona no-decreciente.
- (ii) F es continua por la derecha.
- (iii) $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$ y $F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$.

La importancia de las funciones de distribución radica en la posibilidad de caracterizar a la medida de probabilidad P a través de ellas. Esta circunstancia se basa en los teoremas 3.4 y 3.5.

Teorema 3.4 La función $F(x) = P(-\infty, x]$ define una función de distribución sobre \mathbb{R} .

Demostración.– Probamos las condiciones (a)-(c) de la definición 3.5. En primer lugar, si $x < y$, entonces $(-\infty, x] \subset (-\infty, y]$, lo cual implica $F(x) = P(-\infty, x] \leq P(-\infty, y] = F(y)$; es decir, se verifica (a).

Para demostrar la continuidad por la derecha² de $F(\cdot)$, establecemos la igual-

²Dado que \mathbb{R} es separable, si $x \rightarrow \pm\infty$ entonces $\exists \{a_n\}$ sucesión numerable tal que $a_n \rightarrow \pm\infty$ de modo que toda aplicación F verifica $\lim_{x \rightarrow \pm\infty} F(x) = \lim_{n \rightarrow \infty} F(a_n)$.

dad $\lim_{n \rightarrow \infty} F(x + 1/n) = F(x)$ desde

$$\begin{aligned} \lim_{n \rightarrow \infty} F\left(x + \frac{1}{n}\right) &= \lim_{n \rightarrow \infty} P\left(-\infty, x + \frac{1}{n}\right] = P\left(\lim_{n \rightarrow \infty} \left(-\infty, x + \frac{1}{n}\right]\right) \\ &= P\left(\bigcap_{n=1}^{\infty} \left(-\infty, x + \frac{1}{n}\right]\right) = P(-\infty, x] = F(x). \end{aligned}$$

Además, $F(-\infty) = 0$ y $F(\infty) = 1$, puesto que

$$\begin{aligned} F(-\infty) &= \lim_{n \rightarrow -\infty} F(n) = \lim_{n \rightarrow -\infty} P(-\infty, n] = P\left(\lim_{n \rightarrow -\infty} (-\infty, n]\right) \\ &= P\left(\bigcap_{n=1}^{\infty} (-\infty, -n]\right) = P(\emptyset) = 0, \\ F(\infty) &= \lim_{n \rightarrow \infty} F(n) = \lim_{n \rightarrow \infty} P(-\infty, n] = P\left(\lim_{n \rightarrow \infty} (-\infty, n]\right) \\ &= P\left(\bigcup_{n=1}^{\infty} (-\infty, n]\right) = P(\mathbb{R}) = 1, \end{aligned}$$

lo cual concluye la demostración. \square

Teorema 3.5 Sea F una función de distribución sobre \mathbb{R} . Entonces, F induce en (\mathbb{R}, β) una medida de probabilidad P_F tal que la función de distribución asociada a esta medida de probabilidad es F .

Cuando F es una función no-decreciente sobre \mathbb{R} ,

$$F(x^-) = \lim_{t \rightarrow x-0} F(t) \quad \text{y} \quad F(x^+) = \lim_{t \rightarrow x+0} F(t)$$

existen y son finitos. En general, $F(x^-) \leq F(x) \leq F(x^+)$, de modo que x es un punto de salto de F si y sólo si $F(x^-)$ y $F(x^+)$ existen y son distintos. Por ello, una función no-decreciente F sólo tiene discontinuidades de salto.

Teorema 3.6 El conjunto de puntos de discontinuidad de una función de distribución F es, a lo sumo, numerable.

Demostración.— Sea $I = (a, b]$ un intervalo acotado con, al menos, n puntos de discontinuidad $a < x_1 < x_2 < \dots < x_n \leq b$. Entonces,

$$F(a) \leq F(x_1^-) < F(x_1) \leq \dots \leq F(x_n^-) < F(x_n) \leq F(b).$$

Sean las amplitudes de salto $p_k = F(x_k) - F(x_k^-)$, para $1 \leq k \leq n$. Es claro que

$$\sum_{k=1}^n p_k = F(x_1) - F(x_1^-) + \dots + F(x_n) - F(x_n^-) \leq F(b) - F(a).$$

Por lo tanto, el número de puntos $x \in (a, b]$ con salto $p(x) = F(x) - F(x^-) > \varepsilon > 0$ es, a lo sumo,

$$\frac{F(b) - F(a)}{\varepsilon}.$$

Entonces, para cada $N \in \mathbb{N}$, el número de puntos de discontinuidad con salto mayor que $1/N$ es finito; es decir, no existe más que una cantidad numerable de puntos de discontinuidad en cada intervalo $(a, b]$. Este argumento es suficiente, dado que $\mathbb{R} = \cup_{n=1}^{\infty} (-n, n]$. \square

En la práctica debemos tener en cuenta las siguientes equivalencias:

$$\begin{aligned} P_F(x, \infty) &= 1 - F(x), \\ P_F(x_1, x_2] &= F(x_2) - F(x_1), \\ P_F(-\infty, x) &= F(x^-), \\ P_F(\{x\}) &= F(x) - F(x^-), \\ P_F(x_1, x_2) &= F(x_2^-) - F(x_1), \\ P_F[x_1, x_2] &= F(x_2) - F(x_1^-), \\ P_F[x_1, x_2) &= F(x_2^-) - F(x_1^-). \end{aligned}$$

Ejemplo 3.2 Sea la función de distribución

$$F(x) = \begin{cases} 0, & \text{si } x < -2, \\ 1/8, & \text{si } -2 \leq x < 0, \\ 1/4 + x/8, & \text{si } 0 \leq x < 1, \\ 1/2 + (x-1)/8, & \text{si } 1 \leq x < 2, \\ 3/4 + (x-2)/8, & \text{si } 2 \leq x < 3, \\ 1, & \text{si } 3 \leq x. \end{cases}$$

Entonces,

$$\begin{aligned} P(\mathbb{Q}) &= P(\{-2\}) + P(\{0\}) + P(\{1\}) + P(\{2\}) + P(\{3\}) \\ &= (F(-2) - F(-2^-)) + (F(0) - F(0^-)) + (F(1) - F(1^-)) \\ &\quad + (F(2) - F(2^-)) + (F(3) - F(3^-)) \\ &= \left(\frac{1}{8} - 0\right) + \left(\frac{1}{4} - \frac{1}{8}\right) + \left(\frac{1}{2} - \frac{3}{8}\right) + \left(\frac{3}{4} - \frac{5}{8}\right) + \left(1 - \frac{7}{8}\right) \\ &= \frac{5}{8}. \end{aligned}$$

Distribuciones discretas

Definición 3.6 (Función de distribución discreta) Una función de distribución F se dice discreta si y sólo si existe $\{a_1, \dots, a_m, \dots\} \subset \mathbb{R}$ (finito o numerable) con $P_F(\{a_i\}) > 0$, $\forall i \geq 1$, y $\sum_{i=1}^{\infty} P_F(\{a_i\}) = 1$. En este caso, la colección

$$\{P_F(\{a_i\}) : i \geq 1\}$$

recibe el nombre de función de masa (de probabilidad) o de cuantía.

Podemos calcular

$$F(x) = P_F(-\infty, x] = \sum_{a_i \leq x} P_F(\{a_i\}),$$

de modo que F es una función escalonada, donde

$$P_F(\{a_n\}) = F(a_n) - F(a_n^-) = \text{amplitud del salto de } F \text{ en } a_n.$$

Con generalidad, es posible evaluar $P_F(A) = \sum_{a_i \in A} P_F(\{a_i\})$, $\forall A \in \beta$.

Distribuciones absolutamente continuas

Definición 3.7 (Función de distribución absolutamente continua) Una función de distribución F es absolutamente continua si y sólo si existe una función $f(x)$ no-negativa e integrable tal que

$$\forall x \in \mathbb{R}, \quad F(x) = \int_{(-\infty, x]} f(u) du,$$

en el sentido de la integral de Lebesgue³. La función $f(x)$ es denominada función de densidad.

El teorema 3.7 resume algunas propiedades de interés. (La demostración de (a) es propuesta en la Hoja 3.)

Teorema 3.7

(a) Sea $f : \mathbb{R} \rightarrow \mathbb{R}^+ = [0, \infty)$ una función Riemann-integrable tal que

$$\int_{-\infty}^{\infty} f(u) du = 1.$$

Entonces, $F(x) = \int_{-\infty}^x f(u) du$ define una función de distribución absolutamente continua.

(b) Sea F una función de distribución absolutamente continua. Entonces,

(b.1) F es una función continua.

³La integral de Lebesgue es una integral definida respecto a una medida; es decir, una función de conjunto que verifica las condiciones (a) y (c) de la definición 2.4. La medida de Lebesgue definida sobre (\mathbb{R}, β) equivale a la noción de "longitud" y asigna valores desde $\mu(a, b] = b - a$ y $\mu(\mathbb{R}) = \infty$. En este contexto, si f es una función Riemann-integrable, entonces f es Lebesgue-integrable y ambas integrales coinciden; es decir,

$$\int_{(a, b]} f(t) d\mu(t) = \int_a^b f(t) dt.$$

(b.2) Si la función de densidad f de F es continua en $x = x_0$, entonces F es derivable en $x = x_0$ y $F'(x_0) = f(x_0)$.

(b.3) $P_F(\{x\}) = 0, \forall x \in \mathbb{R}$.

(b.4) Para cualquier intervalo I de la forma (a, b) , $(a, b]$, $[a, b)$ y $[a, b]$,

$$P_F(I) = \int_a^b f(u)du.$$

(b.5) Para todo $A \in \beta$, $P_F(A) = \int_A f(u)du$.

La propiedad (b.1) pone de manifiesto que no todas las funciones de distribución continuas son absolutamente continuas. Existe un tercer tipo de funciones de distribución, denominadas "singulares", que también son continuas. (Su consideración supera los objetivos de la asignatura.)

Ejemplo 3.3 Dada la función $f(x) = 3x^2/2$, si $x \in [-1, 1]$, y 0, en caso contrario, es claro que $f(x) \geq 0, \forall x \in \mathbb{R}$. Además, f es Riemann-integrable e $\int_{-\infty}^{\infty} f(t)dt = 1$. Para determinar F distinguimos:

- Si $x < -1$, entonces $F(x) = 0$.
- Si $-1 \leq x \leq 1$, entonces $F(x) = \int_{-1}^x 3t^2/2dt = (x^3 + 1)/2$.
- Si $1 \leq x$, entonces $F(x) = 1$.

Distribuciones mixtas

Es posible combinar las distribuciones discretas y absolutamente continuas y generar una nueva función de distribución.

Definición 3.8 (Función de distribución mixta) Una función de distribución F es de tipo mixto si y sólo si existen F_1 y F_2 funciones de distribución discreta y absolutamente continua, respectivamente, y $\lambda \in [0, 1]$ tales que $F(x) = \lambda F_1(x) + (1 - \lambda)F_2(x), \forall x \in \mathbb{R}$.

La definición 3.8 asegura que F es una función de distribución. De hecho, esta propiedad es válida incluso si F_1 y F_2 son funciones de distribución (mixtas).

Ejemplo 3.4 Consideremos la función de distribución

$$F(x) = \begin{cases} 0, & \text{si } x < -1, \\ 1/4, & \text{si } -1 \leq x < 1/4, \\ x, & \text{si } 1/4 \leq x < 1/2, \\ 1/2, & \text{si } 1/2 \leq x < 4, \\ (x-1)/x, & \text{si } 4 \leq x. \end{cases}$$

Tomamos

$$\begin{aligned} P(\{-1\}) &= F(-1) - F(-1^-) = \frac{1}{4}, \\ P(\{4\}) &= F(4) - F(4^-) = \frac{1}{4}, \\ f(x) &= I_{(1/4, 1/2)}(x) + \frac{1}{x^2} I_{(4, \infty)}(x). \end{aligned}$$

Entonces, $\lambda = 1/2$, puesto que $F(\infty) = F_1(\infty) = F_2(\infty) = 1$ y

$$F(\infty) = \lambda F_1(\infty) + (1 - \lambda) F_2(\infty),$$

con $\sum_{i=1}^2 P(\{a_i\}) = P(\{-1\}) + P(\{4\}) = 1/4 + 1/4 = 1/2$ e $\int_{-\infty}^{\infty} f(u) du = 1/2$. Por lo tanto, tomamos F_1 y F_2 como las funciones de distribución con función de masa

$$\{P_1(\{-1\}) = 1/2, P_1(\{4\}) = 1/2\}$$

y función de densidad

$$f_2(x) = 2I_{(1/4, 1/2)}(x) + \frac{2}{x^2} I_{(4, \infty)}(x).$$

Para concluir, destacamos el siguiente resultado:

Teorema 3.8 *Toda función de distribución F admite una única descomposición de la forma*

$$F(x) = \alpha F_d(x) + (1 - \alpha) F_c(x), \quad \forall x \in \mathbb{R}, \quad \alpha \in [0, 1],$$

donde F_d y F_c son funciones de distribución escalonada y continua, respectivamente.

3.4 Variables aleatorias discretas y continuas

Sea $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \beta)$ una variable aleatoria. Entonces, X induce una medida de probabilidad P_X sobre (\mathbb{R}, β) mediante la aplicación

$$\begin{aligned} P_X &: \beta \rightarrow [0, 1], \\ B &\rightarrow P_X(B) = (P \circ X)^{-1}(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\}). \end{aligned}$$

Entonces, (\mathbb{R}, β, P_X) es un espacio de probabilidad, donde P_X tiene asociada la función de distribución $F_X : \mathbb{R} \rightarrow \mathbb{R}$ definida por

$$F_X(x) = P_X(-\infty, x] = P(X^{-1}(-\infty, x]) = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

Definición 3.9 (Variable aleatoria discreta) *Sea $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \beta)$ una variable aleatoria. Definamos la función de masa de X mediante $\{p_X(x) = P_X(\{x\}), x \in \mathbb{R}\}$ y sea*

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

Si $D_X \neq \emptyset$ y $\sum_{x \in D_X} p_X(x) = 1$, entonces se dice que X es una variable aleatoria discreta con soporte D_X .

En la práctica, el soporte D_X es el conjunto de valores observables de X y tiene cardinal, a lo sumo, numerable por el teorema 3.6. Además, $D_X \in \beta$ y $\forall B \in \beta$ se tiene

$$P_X(B) = \begin{cases} 0, & \text{si } B \cap D_X = \emptyset, \\ \sum_{x \in B \cap D_X} p_X(x), & \text{si } B \cap D_X \neq \emptyset. \end{cases}$$

A partir de este resultado, la función de distribución de X viene dada por

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

Es decir, F_X es una función escalonada (o definida a saltos).

Definición 3.10 (Variable aleatoria continua) Una variable aleatoria $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \beta)$ se dice continua si y sólo si su función de distribución F_X admite ser representada como

$$F_X(x) = \int_{-\infty}^x f_X(u) du, \quad \forall x \in \mathbb{R},$$

donde $f_X(x)$ es una función de densidad sobre \mathbb{R} ; es decir, cuando F_X es una función de distribución absolutamente continua.

En este contexto, f_X es la función de densidad de X y $C_X = \{x \in \mathbb{R} : f_X(x) > 0\}$ es el soporte de X . El soporte C_X corresponde al conjunto de valores observables de X y tiene cardinal infinito no-numerable.

Supuesta dada la función de densidad f_X de una variable aleatoria X continua, la función de distribución F_X es el punto x_0 viene dada por

$$F_X(x_0) = \int_{-\infty}^{x_0} f_X(u) du,$$

de modo que el valor $F_X(x_0)$ representa el área de la región comprendida entre el eje de abscisas, la función de densidad y la recta (vertical) $x = x_0$.

Teorema 3.9 Si X es una variable aleatoria continua con función de densidad f_X y función de distribución F_X , entonces

- (i) F_X es continua.
- (ii) Si f_X es continua en el punto $x = x_0$, entonces F_X es derivable en $x = x_0$ y $F_X'(x_0) = f_X(x_0)$.
- (iii) $D_X = \{x \in \mathbb{R} : p_X(x) > 0\} = \emptyset$.

(iv) $\forall I \in \{(a, b), (a, b], [a, b), [a, b]\}$ se tiene $P_X(I) = P(X \in I) = \int_a^b f_X(u)du$.

Demostración.— Es consecuencia del teorema 3.7 y la definición 3.10. \square

En cada punto de continuidad de f_X , la función de densidad coincide con la derivada de F_X . Por ello, para comprobar que X es una variable aleatoria continua es suficiente con demostrar que existe F'_X , salvo en un conjunto finito de puntos sobre cada intervalo finito, y se satisfacen:

- (i) $F'_X \geq 0$.
- (ii) F'_X es continua en los intervalos donde está definida.
- (iii) $\int_{-\infty}^{\infty} F'_X(u)du = 1$.

Podemos prolongar F'_X de manera arbitraria en una función no-negativa sobre \mathbb{R} , obteniendo una función de densidad para la variable aleatoria X continua, puesto que el hecho de modificar el valor de una función sobre un conjunto numerable (no denso) en \mathbb{R} no altera el valor de su integral.

Para concluir, proponemos una interpretación de la función de densidad f_X relativa a la variable aleatoria X . Sea x un punto de continuidad de f_X . Entonces,

$$\begin{aligned} P_X[x, x + \Delta x] &= P(\{\omega \in \Omega : x \leq X(\omega) \leq x + \Delta x\}) \\ &= \int_x^{x+\Delta x} f_X(t)dt = f_X(\xi)\Delta x, \end{aligned}$$

donde $\xi \in [x, x + \Delta x]$. Como $P(x \leq X \leq x + \Delta x)$ representa una masa, al dividir por una longitud Δx , es obtenida una "densidad":

$$f_X(\xi) = \frac{P(x \leq X \leq x + \Delta x)}{\Delta x}.$$

3.5 Transformaciones medibles

En ocasiones, dada la ley de probabilidad de una variable aleatoria X , estamos interesados en conocer la ley de probabilidad de $Y = g(X)$; es decir, de una transformación medible de X . Es necesario determinar entonces el tipo de distribución de X e Y .

Caso discreto

Teorema 3.10 *Sea X una variable aleatoria discreta con soporte D_X y función de masa p_X . Sean $g : \mathbb{R} \rightarrow \mathbb{R}$ una función medible e $Y = g(X)$ la variable aleatoria transformada. Entonces, Y es una variable aleatoria discreta con soporte $D_Y = g(D_X)$ y función de masa dada por*

$$p_Y(y) = \sum_{\{x \in \mathbb{R} : g(x)=y\} \cap D_X} p_X(x), \quad \text{si } y \in D_Y.$$

Caso continuo

Si X es una variable aleatoria continua, entonces la variable aleatoria $Y = g(X)$ puede resultar ser discreta o continua, según que $g(C_X)$ sea un conjunto discreto o continuo, respectivamente.

Teorema 3.11 Sea X una variable aleatoria continua con soporte C_X y función de densidad $f_X(x)$. Sea $Y = g(X)$ una variable aleatoria tal que el conjunto $g(C_X)$ es un conjunto discreto. Entonces, Y es discreta con soporte $D_Y \subset g(C_X)$ y función de masa

$$p_Y(y) = \int_{\{x \in \mathbb{R}: g(x)=y\}} f_X(u) du, \quad \text{si } y \in g(C_X).$$

Teorema 3.12 Sea X una variable aleatoria continua con soporte C_X y función de densidad $f_X(x)$. Supongamos que C_X es un intervalo (o unión finita o numerable de intervalos). Sea $g \in \mathbb{R} \rightarrow \mathbb{R}$ una función continua, estrictamente creciente o decreciente sobre C_X , tal que la función inversa g^{-1} sobre $g(C_X)$ admite una derivada continua. Entonces, $Y = g(X)$ es una variable aleatoria continua con soporte $C_Y = g(C_X)$ y función de densidad $f_Y(y)$ dada por

$$f_Y(y) = f_X(g^{-1}(y)) |(g^{-1})'(y)|, \quad \text{si } y \in C_Y.$$

Teorema 3.13 Sea X una variable aleatoria continua con soporte C_X y función de densidad $f_X(x)$. Sea $g \in \mathbb{R} \rightarrow \mathbb{R}$ una función derivable $\forall x \in C_X$ tal que g' es continua y $g'(x) \neq 0$, salvo para, a lo sumo, un número finito de puntos. Supongamos que, para cada $y \in \mathbb{R}$, se cumple una de las dos siguientes condiciones:

- (i) Existen exactamente $m(y) \geq 1$ puntos $\{x_1(y), \dots, x_{m(y)}(y)\} \subset C_X$ tales que

$$g(x_k(y)) = y, \quad g'(x_k(y)) \neq 0, \quad \forall 1 \leq k \leq m(y).$$

- (ii) No existe ningún punto $x \in C_X$ tal que $g(x) = y$ y $g'(x) \neq 0$. Entonces, $m(y) = 0$.

Entonces, $Y = g(X)$ es una variable aleatoria continua con función de densidad

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

Ejemplo 3.5 Sea X una variable aleatoria continua definida sobre $C_X = (-a, a)$, con función de densidad $f_X(x)$. Sea $Y = g(X) = X^2$.

Notemos que $g(x) = x^2$ satisface que $g'(x) = 2x$ es continua y no nula, salvo en $x = 0$, y $g(x) = x^2$ es estrictamente decreciente sobre $(-\infty, 0)$ y estrictamente creciente sobre $(0, \infty)$. Además, para $y \in$

$(0, a^2)$, se tiene que $y = x^2$ si y sólo si $x \in \{-\sqrt{y}, \sqrt{y}\}$, de modo que $m(y) = 2$ con $x_1(y) = -\sqrt{y}$ y $x_2(y) = \sqrt{y}$. Para $y \notin (0, a^2)$ tomamos $m(y) = 0$.

Entonces, para $y \in (0, a^2)$, tenemos que

$$\begin{aligned} f_Y(y) &= f_X(x_1(y)) \frac{1}{|g'(x_1(y))|} + f_X(x_2(y)) \frac{1}{|g'(x_2(y))|} \\ &= f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} \\ &= \frac{f_X(-\sqrt{y}) + f_X(\sqrt{y})}{2\sqrt{y}}. \end{aligned}$$

Por otra parte, si $y \notin (0, a^2)$, entonces $f_Y(y) = 0$.

3.6 Momentos y cuantiles

El estudio de la distribución de probabilidad de una variable aleatoria es también el estudio de algunas características numéricas asociadas, denominadas "parámetros". Es habitual clasificar los parámetros en momentos y parámetros de orden.

Sea X una variable aleatoria discreta con función de masa $\{p_k = P(X = x_k) : k \geq 1\}$. Si $\sum_{k=1}^{\infty} |x_k| p_k < \infty$, entonces se dice que el valor esperado (media o esperanza matemática) de X existe y viene dado por

$$\mu = E[X] = \sum_{k=1}^{\infty} x_k p_k.$$

Debe notarse que la serie $\sum_{k=1}^{\infty} x_k p_k$ puede converger, pero la serie $\sum_{k=1}^{\infty} |x_k| p_k$ ser divergente. En ese caso, se dice que $E[X]$ no existe.

Ejemplo 3.6 Sea X una variable aleatoria discreta con función de masa

$$p_j = P\left(X = (-1)^{j+1} \frac{3^j}{j}\right) = \frac{2}{3^j}, \quad j \geq 1.$$

Entonces, $\sum_{j=1}^{\infty} |x_j| p_j = \sum_{j=1}^{\infty} \frac{2}{3^j} \frac{3^j}{j} = \infty$ y $E[X]$ no existe, aunque la serie $\sum_{j=1}^{\infty} x_j p_j = \sum_{j=1}^{\infty} (-1)^{j+1} \frac{2}{j}$ converge.

Si X es una variable aleatoria continua con función de densidad $f_X(x)$, entonces se dice que $E[X]$ existe cuando

$$\int_{-\infty}^{\infty} |u| f_X(u) du < \infty,$$

en cuyo caso $\mu = E[X] = \int_{-\infty}^{\infty} u f_X(u) du$.

A continuación listamos algunas propiedades de interés de la esperanza de una variable aleatoria.

Teorema 3.14

- (a) Si $X(\omega) = I_A(\omega)$, para algún subconjunto $A \in \mathcal{A}$, entonces $E[X] = P(A)$.
- (b) $E[X]$ existe si y sólo si $E[|X|]$ existe.
- (c) Sea X una variable aleatoria simétrica⁴ respecto a α . Entonces, en el caso de que $E[|X|] < \infty$, se tiene $E[X] = \alpha$.
- (d) Si $a, b \in \mathbb{R}$ y X es una variable aleatoria con $E[|X|] < \infty$, entonces $E[|aX + b|] < \infty$ y $E[aX + b] = aE[X] + b$.
- (e) Si X es acotada⁵, entonces $E[X]$ existe.
- (f) Si $P(X \geq 0) = 1$ y existe $E[X]$, entonces $E[X] \geq 0$.

La aplicación de los teoremas 3.10-13 para transformaciones medibles de variables aleatorias conduce al siguiente resultado:

Teorema 3.15 Sean X una variable aleatoria discreta y g una función medible sobre \mathbb{R} . Tomamos $Y = g(X)$. Entonces,

$$E[Y] = \sum_{k=1}^{\infty} y_k P(Y = y_k) = \sum_{j=1}^{\infty} g(x_j) P(X = x_j),$$

en el sentido de que si alguna de las dos series converge absolutamente, entonces converge la otra y sus sumas son iguales. Similarmente, si X es una variable aleatoria continua con función de densidad $f_X(x)$ e $Y = g(X)$ tiene función de densidad $h_Y(y)$, entonces

$$E[Y] = \int_{-\infty}^{\infty} v h_Y(v) dv = \int_{-\infty}^{\infty} g(u) f_X(u) du,$$

cuando se asume que $E[|g(X)|] < \infty$.

Las funciones $g(x) = x^n$, donde $n \in \mathbb{N}$, y $g(x) = |x|^\alpha$, donde $\alpha > 0$, son de especial importancia. Si $E[X^n]$ existe, entonces llamamos **momento de X con respecto al origen** al valor $E[X^n]$. Si $E[|X|^\alpha] < \infty$, para $\alpha > 0$, se dice que $E[|X|^\alpha]$ es el **α -ésimo momento absoluto de X** . En lo sucesivo, se adopta la notación

$$\begin{aligned} m_n &= E[X^n], \\ \beta_\alpha &= E[|X|^\alpha], \end{aligned}$$

cuando tales esperanzas existan.

⁴ X es simétrica respecto a α si $P(X \geq \alpha + x) = P(X \leq \alpha - x)$, $\forall x \in \mathbb{R}$.

⁵ X es acotada si $P(|X| < M) = 1$, para alguna constante $0 < M < \infty$.

Ejemplo 3.7 Sea X una variable aleatoria que toma valores sobre $\{1, 2, \dots, N\}$ con probabilidades $1/N$. Entonces, todos los momentos (de todos los órdenes) de X existen. Por ejemplo,

$$\begin{aligned} E[X] &= \sum_{k=1}^N k \frac{1}{N} = \frac{1}{N} \frac{N(N+1)}{2} = \frac{N+1}{2}, \\ E[X^2] &= \sum_{k=1}^N k^2 \frac{1}{N} = \frac{1}{N} \frac{N(N+1)(2N+1)}{6} = \frac{(N+1)(2N+1)}{6}. \end{aligned}$$

Teorema 3.16 Si existe el momento de orden α de X , entonces existen los momentos de orden $\gamma \in (0, \alpha)$.

Demostración.— En el caso de una variable aleatoria X continua, es posible escribir

$$\begin{aligned} E[|X|^\gamma] &= \int_{-\infty}^{\infty} |u|^\gamma f_X(u) du \\ &= \int_{\{|x|^\gamma \leq 1\}} |u|^\gamma f_X(u) du + \int_{\{|x|^\gamma > 1\}} |u|^\gamma f_X(u) du \\ &\leq \int_{\{|x|^\gamma \leq 1\}} f_X(u) du + \int_{\{|x|^\gamma > 1\}} |u|^\alpha f_X(u) du \\ &\leq P(\{|X|^\gamma \leq 1\}) + \int_{-\infty}^{\infty} |u|^\alpha f_X(u) du \\ &= P(\{|X|^\gamma \leq 1\}) + E[|X|^\alpha] < \infty. \end{aligned}$$

La demostración es análoga cuando X es discreta. \square

Teorema 3.17 Si X es una variable aleatoria no-negativa con función de distribución F_X , entonces

$$E[X] = \int_0^{\infty} (1 - F_X(u)) du,$$

en el sentido de que si una de ellas existe, entonces existe la otra y son iguales. Para cualquier variable aleatoria X , $E[|X|] < \infty$ si y sólo si las integrales

$$\int_{-\infty}^0 P(X \leq u) du \quad \text{e} \quad \int_0^{\infty} P(X > u) du$$

convergen. En este caso,

$$E[X] = \int_0^{\infty} P(X > u) du - \int_{-\infty}^0 P(X \leq u) du.$$

Demostración.— Sea X una variable aleatoria no-negativa continua tal que $E[X] < \infty$. Entonces,

$$E[X] = \int_0^{\infty} u f_X(u) du = \lim_{n \rightarrow \infty} \int_0^n u f_X(u) du.$$

Integrando por partes,

$$\begin{aligned}\int_0^n u f_X(u) du &= n F_X(n) - \int_0^n F_X(u) du \\ &= -n(1 - F_X(n)) + \int_0^n (1 - F_X(u)) du.\end{aligned}$$

Además, $n(1 - F_X(n)) = n \int_n^\infty f_X(u) du \leq \int_n^\infty u f_X(u) du$. Dado que $E[X] < \infty$, se sigue que la desigualdad

$$0 \leq n(1 - F_X(n)) \leq \int_n^\infty u f_X(u) du$$

implica

$$0 \leq \lim_{n \rightarrow \infty} n(1 - F_X(n)) \leq \lim_{n \rightarrow \infty} \int_n^\infty u f_X(u) du = 0.$$

Es decir, $E[X] = \int_0^\infty (1 - F_X(u)) du$. En el caso discreto, la demostración es análoga. (Omitimos la demostración del resto del teorema.) \square

Sean $k \in \mathbb{N}$ y $c \in \mathbb{R}$. Si existe $E[(X - c)^k]$, entonces $E[(X - c)^k]$ se denomina **momento de orden k con respecto a c** . Si tomamos $c = \mu = E[X]$, que existe dado que $E[|X|^k] < \infty$, llamamos a $E[(X - \mu)^k]$ **momento central de orden k o momento de orden k respecto a la media**. Entonces, escribimos

$$\mu_k = E[(X - \mu)^k].$$

Tengamos en cuenta el siguiente resultado:

Teorema 3.18 Si h_1, h_2, \dots, h_n son funciones medibles de una variable aleatoria X y existen $E[h_i(X)]$, $\forall 1 \leq i \leq n$, entonces $E[\sum_{i=1}^n h_i(X)]$ existe y

$$E\left[\sum_{i=1}^n h_i(X)\right] = \sum_{i=1}^n E[h_i(X)].$$

Entonces, es inmediato establecer

$$\begin{aligned}\mu_k &= E[(X - \mu)^k] = \sum_{i=0}^k \binom{k}{i} E[X^i] (-1)^{k-i} \mu^{k-i} \\ &= \sum_{i=0}^k \binom{k}{i} (-1)^{k-i} m_i \mu^{k-i}, \\ m_k &= E[X^k] = E[((X - \mu) + \mu)^k] = \sum_{i=0}^k \binom{k}{i} E[(X - \mu)^i] \mu^{k-i} \\ &= \sum_{i=0}^k \binom{k}{i} \mu_i \mu^{k-i}.\end{aligned}$$

El caso $k = 2$ es de especial importancia. Si $E[X^2]$ existe, entonces llamamos **varianza de X** al valor $E[(X - \mu)^2]$ y lo denotamos por σ^2 . La cantidad σ es llamada **desviación típica de X** .

Es posible escribir

$$\sigma^2 = \mu_2 = E[X^2] - (E[X])^2 = m_2 - \mu^2.$$

La varianza tiene algunas propiedades de interés.

Teorema 3.19 (i) $Var(X) = 0$ si y sólo si X es degenerada (es decir, X toma un único valor).

(ii) $Var(X) < E[(X - c)^2]$, para cualquier $c \neq E[X]$.

Para concluir el estudio de los momentos de una variable aleatoria proponemos algunas desigualdades de interés.

Teorema 3.20 Sea $h(X)$ una función medible no-negativa de una variable aleatoria X . Si $E[h(X)]$ existe, entonces

$$P(h(X) \geq \varepsilon) \leq \frac{E[h(X)]}{\varepsilon}, \quad \text{para cada } \varepsilon > 0.$$

Demostración.— Probamos el resultado en el caso de una variable aleatoria X discreta con $p_k = P(X = x_k)$, $k \geq 1$. En primer lugar, expresamos

$$E[h(X)] = \sum_{k=1}^{\infty} h(x_k)p_k = \sum_{x_k \in A} h(x_k)p_k + \sum_{x_k \in A^c} h(x_k)p_k,$$

donde $A = \{x : h(x) \geq \varepsilon\}$. Entonces,

$$E[h(X)] \geq \sum_{x_k \in A} h(x_k)p_k \geq \varepsilon \sum_{x_k \in A} p_k = \varepsilon P(h(X) \geq \varepsilon),$$

lo cual concluye la demostración. \square

Si $h(x) = |x|^r$ y $\varepsilon = K^r$, con $r, K > 0$, entonces el teorema 3.20 conduce a la **desigualdad de Markov**

$$P(|X| \geq K) \leq \frac{E[|X|^r]}{K^r}.$$

Cuando $h(x) = (x - \mu)^2$ y $\varepsilon = K^2$, se deduce la **desigualdad de Chebyshev**

$$P(|X - \mu| \geq K) \leq \frac{\sigma^2}{K^2},$$

donde $\mu = E[X]$ y $\sigma^2 = Var(X)$.

Es posible mostrar que para algunas distribuciones incluso la media no existe. A continuación consideramos algunos parámetros, llamados parámetros de orden, que siempre existen.

En concreto, un número x que satisface

$$P(X \leq x) \geq p \quad \text{y} \quad P(X \geq x) \geq 1 - p,$$

para $p \in (0, 1)$, es llamado **cuantil de orden p** (o $100 \times p$ -ésimo percentil) de la variable aleatoria X . Escribimos $\xi_p(X)$ para el cuantil de orden p de X .

Si x es un cuantil de orden p de X , entonces

$$p \leq P(X \leq x) = F_X(x) = P(X < x) + P(X = x) \leq p + P(X = x).$$

Por ello, si $P(X = x) = 0$, entonces el cuantil de orden p es una solución de la ecuación

$$F_X(x) = p.$$

Si F_X es estrictamente creciente, entonces $F_X(x) = p$ tiene una única solución. En caso contrario, puede existir más de una solución, y cada una de ellas es llamada **cuantil de orden p** .

Si $p = 1/2$, el cuantil resultante es denominado **mediana**, siendo una constante de centralización de importancia, especialmente en casos donde la media de la distribución no existe.

Ejemplo 3.8 Sea X una variable aleatoria con función de densidad

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

Entonces,

$$\begin{aligned} \int_{-\infty}^{\infty} |u| f_X(u) du &= \int_{-\infty}^{\infty} \frac{|u|}{\pi} \frac{1}{1+u^2} du = \frac{2}{\pi} \int_0^{\infty} \frac{u}{1+u^2} du \\ &= \left. \frac{1}{\pi} \ln(1+u^2) \right|_0^{\infty} = \infty. \end{aligned}$$

Por ello, $E[X]$ no existe. No obstante, como X es continua, la ecuación $F(x) = 1/2$ tiene solución $x = 0$. Es decir, $x = 0$ es la mediana de X .

3.7 Función característica y función generatriz de momentos

La distribución de probabilidad de una variable aleatoria puede determinarse a través de su función de distribución o, equivalentemente, desde la función de densidad o de cuantía, según el caso. A continuación definimos nuevas funciones

(la función generatriz de momentos y la función característica) que permiten caracterizar, de forma única, la ley de probabilidad de la variable aleatoria en estudio.

La función generatriz de una sucesión $\{x_n : n \geq 0\} \subset \mathbb{R}$ es definida por

$$G(z) = \sum_{n=0}^{\infty} x_n z^n, \quad z \in \mathbb{C},$$

supuesto que la serie es convergente.

Definición 3.11 Dada una distribución de probabilidad discreta $\{p_n : n \geq 0\}$, se llama función generatriz a

$$G(z) = \sum_{n=0}^{\infty} p_n z^n,$$

que converge en $|z| \leq 1$. (Asumimos que $p_n = P(X = n)$, para $n \geq 0$.)

La definición 3.11 puede ser ampliamente extendida.

Dada una variable aleatoria X con función de distribución F_X , se define la función generatriz (de probabilidad) de X por

$$G_X(z) = E[z^X] = \int_{\mathbb{R}} z^x dF_X(x),$$

donde la integral $\int_{\mathbb{R}} z^x dF_X(x)$ es una integral de Lebesgue-Stieltjes dada por:

- (i) Si X es discreta con soporte $\mathcal{C}_X = \{x_i : i \geq 0\}$ y función de masa $\{p_i : i \geq 0\}$, entonces

$$\int_{\mathbb{R}} z^x dF_X(x) = \sum_{i=0}^{\infty} z^{x_i} p_i,$$

- (ii) Si X es continua con función de densidad f_X , entonces

$$\int_{\mathbb{R}} z^x dF_X(x) = \int_{-\infty}^{\infty} z^x f_X(x) dx,$$

supuesto que $\sum_{i=0}^{\infty} |z^{x_i}| p_i < \infty$ e $\int_{-\infty}^{\infty} |z^x| f_X(x) dx < \infty$, según el caso.

Debido a la convergencia uniforme dentro del radio de convergencia, la derivada de la serie (en el caso discreto) es la serie de derivadas, teniendo el mismo radio de convergencia. Si X toma valores en \mathbb{Z}_+ , desde $G_X(z) = \sum_{i=0}^{\infty} p_i z^i$, es posible determinar

$$\begin{aligned} G'_X(z) &= \sum_{i=0}^{\infty} i p_i z^{i-1}, \\ G_X^{(r)}(z) &= \sum_{i=0}^{\infty} i(i-1) \cdots (i-r+1) p_i z^{i-r}, \quad r \geq 2. \end{aligned}$$

Supuesto que las anteriores series convergen en el punto $z = 1$ de la frontera de convergencia, es posible trabajar con los valores $G_X(1) = 1$,

$$\begin{aligned} G'_X(1) &= \sum_{i=0}^{\infty} i p_i, \\ G_X^{(r)}(1) &= \sum_{i=0}^{\infty} i(i-1) \cdots (i-r+1) p_i, \quad r \geq 2. \end{aligned}$$

En el caso de una variable aleatoria X continua, las anteriores series son reemplazadas por $G_X(z) = \int_{-\infty}^{\infty} z^x f_X(x) dx$,

$$\begin{aligned} G'_X(z) &= \int_{-\infty}^{\infty} x z^{x-1} f_X(x) dx, \\ G_X^{(r)}(z) &= \int_{-\infty}^{\infty} x(x-1) \cdots (x-r+1) z^{x-r} f_X(x) dx, \quad r \geq 2. \end{aligned}$$

Cuando $z = 1$, suponiendo la convergencia absoluta de las integrales en $z = 1$, tenemos $G_X(1) = 1$,

$$\begin{aligned} G'_X(1) &= \int_{-\infty}^{\infty} x f_X(x) dx, \\ G_X^{(r)}(1) &= \int_{-\infty}^{\infty} x(x-1) \cdots (x-r+1) f_X(x) dx, \quad r \geq 2. \end{aligned}$$

Estas expresiones corresponden a los momentos factoriales de una variable aleatoria.

Definición 3.12 Dada la función de distribución F_X de una variable aleatoria X , se denomina momento factorial de orden r a la integral de Lebesgue-Stieltjes

$$\alpha_{(r)} = \int_{\mathbb{R}} x(x-1) \cdots (x-r+1) dF_X(x),$$

es decir $\alpha_{(r)} = G_X^{(r)}(1)$.

Con frecuencia, la computación de momentos de órdenes r elevados es laboriosa. Entonces, resulta aconsejable el uso de la función generatriz de momentos.

Definición 3.13 La función generatriz de momentos de una variable aleatoria X viene dada⁶ por

$$M_X(\theta) = E[e^{\theta X}] = \int_{\mathbb{R}} e^{\theta x} dF_X(x).$$

⁶Es decir, en el caso discreto $M_X(\theta) = \sum_{i=0}^{\infty} e^{\theta x_i} p_i$, supuesto que $\sum_{i=0}^{\infty} |e^{\theta x_i}| p_i < \infty$; en el caso continuo, $M_X(\theta) = \int_{-\infty}^{\infty} e^{\theta x} f_X(x) dx$, supuesto que $\int_{-\infty}^{\infty} |e^{\theta x}| f_X(x) dx < \infty$.

La función generatriz de momentos $M_X(\theta)$ no existe necesariamente para cualquier valor de θ . Por ejemplo, $M_X(\theta)$ existe si $X \geq 0$ y $\theta \leq 0$ (en cuyo caso, $e^{\theta X} \leq 1$), cuando X es acotada o para $\theta = 0$. No obstante, no es posible asegurar su existencia en un entorno del origen.

Cuando las funciones $G_X(\cdot)$ y $M_X(\cdot)$ existen simultáneamente, es claro que están relacionadas por

$$\begin{aligned} M_X(\ln z) &= G_X(z), \\ M_X(\theta) &= G_X(e^\theta). \end{aligned}$$

Si desarrollamos en serie la función exponencial $e^{\theta x}$ y asumimos que dicho desarrollo es uniformemente convergente y tal que podemos integrar término-a-término,

$$\begin{aligned} M_X(\theta) &= \int_{\mathbb{R}} e^{\theta x} dF_X(x) = \int_{\mathbb{R}} \left(1 + \frac{\theta}{1!} + \frac{\theta^2}{2!} + \dots \right) dF_X(x) \\ &= \int_{\mathbb{R}} dF_X(x) + \frac{\theta}{1!} \int_{\mathbb{R}} x dF_X(x) + \frac{\theta^2}{2!} \int_{\mathbb{R}} x^2 dF_X(x) + \dots \\ &= 1 + \frac{\theta}{1!} m_1 + \frac{\theta^2}{2!} m_2 + \dots \end{aligned}$$

Tomando derivadas en $\theta = 0$, se deduce que

$$\left. \frac{d^{(k)} M_X(\theta)}{d\theta^k} \right|_{\theta=0} = m_k, \quad k \geq 1,$$

y $M_X(0) = 1$.

Una alternativa para caracterizar la ley de probabilidad de una variable aleatoria es la función característica.

Definición 3.14 Sean (Ω, \mathcal{A}, P) un espacio de probabilidad y X una variable aleatoria con función de distribución F_X . Se denomina función característica⁷ de X a la transformada de Fourier-Stieltjes de F_X ; es decir,

$$\begin{aligned} \varphi_X(t) &= E[e^{itX}] = \int_{\mathbb{R}} e^{itx} dF_X(x) = \\ &= \int_{\mathbb{R}} \cos(tx) dF_X(x) + i \int_{\mathbb{R}} \sin(tx) dF_X(x), \quad \forall t \in \mathbb{R}, \end{aligned}$$

⁷Si X es discreta con soporte $\mathcal{D}_X = \{x_j : j \geq 0\}$ y función de masa $\{p_X(x_j) : j \geq 0\}$, entonces

$$\varphi_X(t) = \sum_{x_j \in \mathcal{D}_X} e^{itx_j} p_X(x_j).$$

Si X es continua, entonces

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx,$$

siendo f_X la función de densidad de X .

donde i denota la unidad compleja.

Teorema 3.21 Sea $\varphi(t)$ una función característica.

- (a) $\varphi(t)$ existe y $|\varphi(t)| \leq 1$, $\forall t \in \mathbb{R}$.
- (b) $\varphi(0) = 1$.
- (c) $\varphi(t)$ es uniformemente continua.
- (d) Si $\varphi_X(\cdot)$ es la función característica de X , entonces la función característica de $Y = a + bX$ viene dada por $\varphi_Y(t) = e^{ita}\varphi_X(bt)$.
- (e) Si existe $m_n = \int_{\mathbb{R}} x^n dF_X(x)$, entonces $\exists \varphi^{(n)}(0)$ y $\varphi^{(n)}(0) = i^n m_n$, y $\exists \varphi^{(n)}(t)$ y $\varphi^{(n)}(t) = i^n \int_{\mathbb{R}} e^{itx} x^n dF_X(x)$.
- (f) Si $\exists \varphi^{(2n)}(t)$, entonces $\exists m_{2n}$ y

$$m_{2n} = \frac{\varphi^{(2n)}(0)}{i^{2n}}.$$

- (g) Si $\exists \varphi^{(2n-1)}(t)$, entonces $\exists m_{2n-2}$ y

$$m_{2n-2} = \frac{\varphi^{(2n-2)}(0)}{i^{2n-2}}.$$

Dada una función de distribución F , la correspondiente función característica queda determinada por

$$\varphi(t) = \int_{\mathbb{R}} e^{itx} dF(x).$$

Un problema de interés consiste en determinar si dos funciones de distribución F_1 y F_2 distintas pueden tener la misma función característica $\varphi(t)$; es decir,

$$\int_{\mathbb{R}} e^{itx} dF_1(x) = \int_{\mathbb{R}} e^{itx} dF_2(x), \quad \forall t \in \mathbb{R}.$$

El teorema de inversión implica que este hecho no es posible. En concreto, establece que si x_1 y x_2 son puntos de continuidad (con $x_1 < x_2$) de una función de distribución F y $\varphi(t)$ es la función característica asociada, entonces

$$F(x_2) - F(x_1) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it} \varphi(t) dt.$$

Esta igualdad caracteriza a F en sus puntos de continuidad. En el caso de una función de distribución F absolutamente continua, se tiene que

$$\frac{F(x_2) - F(x_1)}{x_2 - x_1} = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it(x_2 - x_1)} \varphi(t) dt,$$

desde donde se sigue que

$$f(x_2) = \lim_{x_1 \rightarrow x_2} \frac{F(x_2) - F(x_1)}{x_2 - x_1}.$$

Por tanto, si asumimos que $|\varphi(t)|$ es integrable, entonces

$$\begin{aligned} \lim_{x_1 \rightarrow x_2} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx_1} - e^{-itx_2}}{it(x_2 - x_1)} \varphi(t) dt &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{x_1 \rightarrow x_2} \frac{e^{-itx_1} - e^{-itx_2}}{it(x_2 - x_1)} \varphi(t) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx_2} \varphi(t) dt. \end{aligned}$$

En definitiva, se tiene la relación

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t) dt.$$

Teorema 3.22 (Inversión) Sea $\varphi(t)$ la función característica con función de distribución $F(x)$. Si α y β son puntos de continuidad de F , con $\alpha < \beta$, se tiene

$$F(\beta) - F(\alpha) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-it\beta} - e^{-it\alpha}}{-it} \varphi(t) dt.$$

Teorema 3.23

- (a) Si la función característica $\varphi(t)$ es absolutamente integrable sobre \mathbb{R} , entonces la función de distribución F con función característica $\varphi(t)$ es absolutamente continua y tiene por función de densidad a

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t) dt.$$

- (b) Si X es una variable aleatoria discreta definida sobre \mathbb{Z} , con función de masa $\{p_k : k \in \mathbb{Z}\}$ y función característica

$$\varphi_X(t) = \sum_{k=-\infty}^{\infty} p_k e^{ikt}, \quad t \in \mathbb{R},$$

$$\text{entonces, } p_k = P(X = k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \varphi_X(t) dt.$$

3.8 Distribuciones discretas

Degenerada, causal o de Dirac. La distribución degenerada surge en experimentos deterministas. La variable aleatoria X con distribución degenerada (D) tiene soporte $D_X = \{D\}$ y función de masa

$$p_X(x) = \begin{cases} 1, & \text{si } x = D, \\ 0, & \text{si } x \neq D. \end{cases}$$

Entonces, $F_X(x) = I_{[D, \infty)}(x)$, $\varphi_X(t) = e^{iDt}$, $E[X] = D$ y $Var(X) = 0$.

Bernoulli. Surge en experimentos de dos únicos resultados (éxito o fracaso). Entonces, X contabiliza el número de éxitos obtenidos. Dada la probabilidad $p \in [0, 1]$, la variable aleatoria X tiene distribución Bernoulli (p) si su soporte es $D_X = \{0, 1\}$ y su función de masa viene dada por

$$p_X(x) = \begin{cases} p^x(1-p)^{1-x}, & \text{si } x \in D_X, \\ 0, & \text{si } x \notin D_X. \end{cases}$$

Entonces, $\varphi_X(t) = pe^{it} + (1-p)$, $E[X] = p$ y $Var(X) = p(1-p)$.

Binomial. Surge al repetir n veces un experimento de Bernoulli de forma independiente en cada repetición y se contabiliza el número de éxitos acumulados en las n repeticiones. Dados $n \in \mathbb{N}$ y $p \in [0, 1]$, la variable aleatoria X tiene distribución Binomial (n, p) si el soporte es $D_X = \{0, 1, \dots, n\}$ y la función de masa es

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{si } x \in D_X, \\ 0, & \text{si } x \notin D_X. \end{cases}$$

Entonces, $\varphi_X(t) = (pe^{it} + (1-p))^n$, $E[X] = np$ y $Var(X) = np(1-p)$. Debemos notar que, si X_1, \dots, X_m son variables aleatorias independientes⁸ Bernoulli de parámetros (n_1, p) , \dots , (n_m, p) , entonces $X = \sum_{i=1}^m X_i$ tiene distribución Bernoulli de parámetro (n, p) , donde $n = \sum_{i=1}^m n_i$.

Geométrica o de Pascal. Se presenta cuando se realiza un experimento de Bernoulli sucesiva e independientemente hasta obtener el primer éxito y se contabiliza el número de fracasos acumulados. Dado $p \in (0, 1]$, la variable aleatoria X con distribución geométrica (p) tiene soporte $D_X = \{0\} \cup \mathbb{N}$ y función de masa

$$p_X(x) = \begin{cases} p(1-p)^x, & \text{si } x \in D_X, \\ 0, & \text{si } x \notin D_X. \end{cases}$$

Entonces, $\varphi_X(t) = \frac{p}{1-(1-p)e^{it}}$, $E[X] = \frac{1-p}{p}$ y $Var(X) = \frac{1-p}{p^2}$.

Binomial negativa. Se genera cuando se realiza sucesiva e independientemente un experimento de Bernoulli hasta obtener r éxitos y se contabiliza el número de fracasos acumulados. La variable aleatoria X con distribución binomial negativa (r, p) tiene soporte $D_X = \{0\} \cup \mathbb{N}$ y función de masa

$$p_X(x) = \begin{cases} \binom{r+x-1}{x} p^r (1-p)^x, & \text{si } x \in D_X, \\ 0, & \text{si } x \notin D_X. \end{cases}$$

⁸Véase la definición de variables aleatorias independientes en el tema 4.

Entonces, $\varphi_X(t) = \left(\frac{p}{1-(1-p)e^{it}}\right)^r$, $E[X] = \frac{r(1-p)}{p}$ y $Var(X) = \frac{r(1-p)}{p^2}$. Además, si X_1, \dots, X_m son m variables aleatorias independientes binomial negativas de parámetros (r_1, p) , \dots , (r_m, p) , respectivamente, entonces $X = X_1 + \dots + X_m$ tiene distribución binomial negativa de parámetro (r, p) , donde $r = r_1 + \dots + r_m$.

Hipergeométrica. Corresponde al problema asociado a la extracción de n bolas al azar de una urna formada por N bolas, de las cuales D son blancas. Se contabiliza entonces el número de bolas blancas entre las n extraídas.

Dados $N \in \mathbb{N}$ y $p \in [0, 1]$, se toma el valor $D = Np$. Entonces, la variable aleatoria X con distribución hipergeométrica (N, D) (o (N, p)) tiene soporte

$$D_X = \{a, a+1, \dots, b\},$$

donde $a = \max\{0, n - N(1-p)\}$ y $b = \min\{n, Np\}$. La función de masa viene dada por

$$p_X(x) = \begin{cases} \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}}, & \text{si } x \in D_X, \\ 0, & \text{si } x \notin D_X. \end{cases}$$

Entonces, $\varphi_X(t) = \sum_{x=0}^n \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}} e^{itx}$, $E[X] = np$ y $Var(X) = np(1-p) \frac{N-n}{n-1}$.

Poisson. Surge cuando en un intervalo de longitud 1 (por ejemplo, $[0, 1]$) se reparten al azar n puntos. Sea I_n un subintervalo cualquiera de amplitud λ/n . Entonces, la distribución del número de puntos contenidos en I_n es Poisson.

Dado $\lambda > 0$, una variable aleatoria X con distribución Poisson (λ) tiene soporte $D_X = \{0\} \cup \mathbb{N}$ y función de masa

$$p_X(x) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!}, & \text{si } x \in D_X, \\ 0, & \text{si } x \notin D_X. \end{cases}$$

Entonces, $\varphi_X(t) = e^{\lambda(e^{it}-1)}$, $E[X] = \lambda$ y $Var(X) = \lambda$.

Para finalizar esta sección, destacamos dos resultados referidos a la convergencia de ciertas distribuciones:

1. Convergencia de la hipergeométrica a la binomial.

$$\lim_{N \rightarrow \infty} \frac{\binom{Np}{x} \binom{N(1-p)}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, \dots, n\}.$$

2. Convergencia de la binomial a la Poisson.

$$\lim_{n \rightarrow \infty} \text{con } np=\lambda \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x \in \{0\} \cup \mathbb{N}.$$

3.9 Distribuciones absolutamente continuas

Uniforme o rectangular. Para valores $a < b$, se define la distribución uniforme sobre (a, b) desde la función de densidad

$$f_X(x) = \frac{1}{b-a} I_{(a,b)}(x),$$

de modo que $\varphi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$, $E[X] = \frac{a+b}{2}$ y $Var(X) = \frac{(b-a)^2}{12}$.

Normal. Dados los valores $\mu \in \mathbb{R}$ y $\sigma > 0$, se define la distribución Normal de parámetro (μ, σ^2) como aquella con función de densidad

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}.$$

Es sencillo comprobar que $\varphi_X(t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}$, $E[X] = \mu$ y $Var(X) = \sigma^2$. El gráfico de $f_X(x)$ es simétrico con eje de simetría $x = \mu$ y es conocido como "campana de Gauss". Como consecuencia, todos los momentos centrales de orden impar son nulos. Para una distribución Normal $(0, 1)$, los coeficientes de asimetría y aplastamiento son nulos

$$\begin{aligned} A_F &= \frac{\mu_3}{\sigma^3} = \frac{E[(X-\mu)^3]}{\sigma^3} = 0, \\ b_2 - 3 &= \frac{\mu_4}{\sigma^4} - 3 = 0. \end{aligned}$$

Entre las propiedades de mayor interés, destacamos

- Si X es Normal (μ, σ^2) , entonces $Z = \frac{X-\mu}{\sigma}$ tiene distribución Normal $(0, 1)$.
- Si X es Normal (μ, σ^2) , entonces $Y = aX + b$ tiene distribución Normal $(a\mu + b, a^2\sigma^2)$.
- Si X_1, \dots, X_m son variables aleatorias independientes con distribuciones Normales de parámetros $(\mu_1, \sigma_1^2), \dots, (\mu_m, \sigma_m^2)$, respectivamente, entonces $a_1 X_1 + \dots + a_m X_m$ es Normal (α, β^2) , donde $\alpha = a_1 \mu_1 + \dots + a_m \mu_m$ y $\beta^2 = a_1^2 \sigma_1^2 + \dots + a_m^2 \sigma_m^2$.

El interés de la distribución Normal viene de su presencia en resultados límite como el siguiente:

- Si $\{X_n : n \geq 1\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con media $E[X_i] = \mu$ y varianza $\sigma^2 = Var(X_i)$ finitas, entonces

$$\lim_{n \rightarrow \infty} F_{\eta_n}(x) = F_Z(x),$$

donde

$$\eta_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}},$$

y Z tiene distribución Normal $(0, 1)$.

Exponencial. Para $\lambda > 0$, la distribución exponencial de tasa λ viene dada por la función de densidad

$$f_X(x) = \lambda e^{-\lambda x} I_{(0, \infty)}(x).$$

Entonces $\varphi_X(t) = \frac{1}{1+it/\lambda}$, $E[X] = \frac{1}{\lambda}$ y $Var(X) = \frac{1}{\lambda^2}$. Es la única distribución absolutamente continua con "ausencia de memoria"; es decir, tal que

$$P(X \geq b | X \geq a) = P(X \geq b - a), \quad b > a.$$

Gamma. La función Gamma se define para valores $p > 0$ como

$$\Gamma(p) = \int_0^\infty e^{-u} u^{p-1} du$$

y verifica un buen número de propiedades como $\Gamma(p) = (p-1)\Gamma(p-1)$, $\forall p > 1$, $\Gamma(n) = (n-1)!$, $\Gamma(1) = 1$, $\Gamma(1/2) = \sqrt{\pi}$ y

$$\int_0^\infty e^{-au} u^{p-1} du = \frac{\Gamma(p)}{a^p}, \quad \text{si } a > 0.$$

La distribución Gamma (a, p) se define desde la función de densidad

$$f_X(x) = \frac{a^p}{\Gamma(p)} e^{-ax} x^{p-1} I_{(0, \infty)}(x),$$

de modo que $\varphi_X(t) = \frac{1}{(1-it/a)^p}$, $E[X] = \frac{p}{a}$ y $Var(X) = \frac{p}{a^2}$. Es de interés observar que si X_1 y X_2 son variables aleatorias independientes Gamma de parámetros (a, p_1) y (a, p_2) , respectivamente, entonces $X_1 + X_2$ es Gamma de parámetro $(a, p_1 + p_2)$.

La distribución exponencial (λ) equivale a una Gamma ($a = \lambda, p = 1$). La distribución χ_n^2 (chi-cuadrado con n grados de libertad) se define como una distribución Gamma ($a = 1/2, p = n/2$) y puede ser vista como la distribución de la suma de los cuadrados de n variables aleatorias independientes Normales $(0, 1)$.

Beta. Dados los valores $p, q > 0$, la función Beta se define como

$$\beta(p, q) = \int_0^1 u^{p-1} (1-u)^{q-1} du.$$

Entonces, se verifican las propiedades

$$\beta(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}, \quad \beta\left(\frac{1}{2}, \frac{1}{2}\right) = \pi.$$

La función de densidad

$$f_X(x) = \frac{1}{\beta(p, q)} x^{p-1} (1-x)^{q-1} I_{(0,1)}(x)$$

define la distribución Beta (p, q) . Entonces,

$$E[X] = \frac{p}{p+q} \quad \text{y} \quad Var(X) = \frac{pq}{(p+q)^2(p+q+1)}.$$

Tema 4. Variables aleatorias multidimensionales

Indice

- 4.1 Introducción
- 4.2 Función de distribución
- 4.3 Vectores aleatorios discretos y continuos
- 4.4 Distribuciones marginales y condicionadas
- 4.5 Independencia de variables aleatorias
- 4.6 Transformaciones medibles de vectores aleatorios
- 4.7 Momentos y matriz de varianza-covarianzas. Recta de regresión
- 4.8 Función característica
- 4.9 Distribuciones multidimensionales notables
- Ejercicios

4.1 Introducción

Es frecuente tratar con experimentos aleatorios descritos desde una n -dupla (X_1, \dots, X_n) , donde cada componente X_i es una variable aleatoria 1-dimensional; por ejemplo, éste es el caso del ejemplo 2.1 cuando una moneda equilibrada es lanzada dos veces.

Desde la sección 3.2, sabemos que $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \beta^n)$ es una variable aleatoria o vector aleatorio n -dimensional si y sólo si cada componente X_i es una variable aleatoria 1-dimensional, $\forall 1 \leq i \leq n$; es decir, si y sólo si la imagen inversa

$$X^{-1}(-\infty, a] = \{\omega \in \Omega : -\infty < X_i(\omega) \leq a_i, 1 \leq i \leq n\}$$

de cada rectángulo $(-\infty, a] = (-\infty, a_1] \times \dots \times (-\infty, a_n]$ está en \mathcal{A} , $\forall a = (a_1, \dots, a_n) \in \mathbb{R}^n$.

Entonces, para $n \geq 2$, es posible desarrollar contenidos similares a aquéllos recogidos en el tema 3. A continuación, exponemos estos resultados haciendo mayor énfasis en el caso $n = 2$.

4.2 Función de distribución

Comenzamos con la definición de función de distribución de una variable aleatoria en \mathbb{R}^n , para $n = 2$.

Definición 4.1 (Función de distribución en \mathbb{R}^2) Una función $F(\cdot, \cdot)$ de dos variables es una función de distribución de algún vector aleatorio (2-dimensional) si y sólo si

(a) F es no-decreciente y continua por la derecha con respecto a ambas componentes.

(b)

$$\begin{aligned} F(-\infty, y) &= \lim_{x \rightarrow -\infty} F(x, y) = 0, \quad \forall y \in \mathbb{R}, \\ F(x, -\infty) &= \lim_{y \rightarrow -\infty} F(x, y) = 0, \quad \forall x \in \mathbb{R}, \\ F(\infty, \infty) &= \lim_{x, y \rightarrow \infty} F(x, y) = 1. \end{aligned}$$

(c) $\forall (x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ con $x_1 < x_2$ e $y_1 < y_2$ se verifica

$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \geq 0.$$

En concreto, dado un vector aleatorio (X, Y) arbitrario, la función F definida por

$$F(x, y) = P(X \leq x, Y \leq y), \quad (x, y) \in \mathbb{R}^2,$$

resulta ser una función de distribución sobre \mathbb{R}^2 . Entonces, para valores $x_1 < x_2$ e $y_1 < y_2$, se tiene

$$\begin{aligned} P(x_1 < X \leq x_2, y_1 < Y \leq y_2) &= P(X \leq x_2, Y \leq y_2) \\ &\quad - P(X \leq x_1, Y \leq y_2) - P(X \leq x_2, Y \leq y_1) \\ &\quad + P(X \leq x_1, Y \leq y_1) \\ &= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1). \end{aligned}$$

Notemos una función $F(\cdot, \cdot)$ verificando las condiciones (a) y (b) de la definición 4.1 no necesariamente es la función de distribución de un vector aleatorio.

Ejemplo 4.1 Sea F la función definida por

$$F(x, y) = \begin{cases} 0, & \text{si } x < 0 \text{ o } x + y < 1 \text{ o } y < 0, \\ 1, & \text{en caso contrario.} \end{cases}$$

Entonces, F satisface (a) y (b) de la definición 4.1, pero no puede ser la función de distribución de un vector aleatorio (X, Y) puesto que

$$\begin{aligned} P\left(\frac{1}{3} < X \leq 1, \frac{1}{3} < Y \leq 1\right) &= F(1, 1) - F\left(1, \frac{1}{3}\right) - F\left(\frac{1}{3}, 1\right) \\ &\quad + F\left(\frac{1}{3}, \frac{1}{3}\right) \\ &= 1 - 1 - 1 + 0 = -1 < 0. \end{aligned}$$

Los anteriores elementos se extienden fácilmente al caso $n \geq 3$.

Definición 4.2 Se dice que $F(x_1, \dots, x_n)$ es la función de distribución de un vector aleatorio (X_1, \dots, X_n) (n -dimensional) si y sólo si F es no-decreciente y continua por la derecha con respecto a todas las componentes y satisface

$$(a) \quad F(-\infty, x_2, \dots, x_n) = F(x_1, -\infty, \dots, x_n) = \dots = F(x_1, x_2, \dots, -\infty) = 0, \\ \forall x_i \in \mathbb{R}, 1 \leq i \leq n, \text{ y } F(\infty, \infty, \dots, \infty) = 1.$$

$$(b) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n \text{ y } \forall \varepsilon_i > 0, 1 \leq i \leq n, \text{ se tiene}$$

$$\begin{aligned} &F(x_1 + \varepsilon_1, x_2 + \varepsilon_2, \dots, x_n + \varepsilon_n) \\ &- \sum_{i=1}^n F(\dots, x_{i-1} + \varepsilon_{i-1}, x_i, x_{i+1} + \varepsilon_{i+1}, \dots) \\ &+ \sum_{i,j=1, i < j}^n F(\dots, x_{i-1} + \varepsilon_{i-1}, x_i, x_{i+1} + \varepsilon_{i+1}, \dots, x_{j-1} + \varepsilon_{j-1}, x_j, x_{j+1} + \varepsilon_{j+1}, \dots) \\ &- \dots + (-1)^n F(x_1, \dots, x_n) \geq 0. \end{aligned}$$

4.3 Vectores aleatorios discretos y continuos

A continuación restringimos la atención a los casos discreto y continuo.

Definición 4.3 Sea (X, Y) un vector aleatorio 2-dimensional. Entonces,

(a) Se dice que (X, Y) es discreto si y sólo si toma valores sobre un conjunto de pares $(x_i, y_j) \in \mathcal{D}_{(X,Y)} \in \beta^2$ con probabilidades $p_{ij} = P(X = x_i, Y = y_j)$ tales que $\sum_{i,j} p_{ij} = 1$. En este caso, la función de distribución de (X, Y) viene dada por

$$F_{(X,Y)}(x, y) = P(X \leq x, Y \leq y) = \sum_{i,j: x_i \leq x, y_j \leq y} p_{ij}.$$

(b) Se dice que (X, Y) es continuo si y sólo si existe una función $f(x, y)$ no-negativa tal que $\forall (x, y) \in \mathbb{R}^2$ se tiene

$$F_{(X,Y)}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du.$$

Entonces, $f(x, y)$ es llamada función de densidad (conjunta) de (X, Y) .

En ambos casos, es claro que $F(\infty, \infty) = 1$. Además, en el caso continuo, si $f(x, y)$ es continua en el punto (x_0, y_0) , entonces

$$\left. \frac{\partial^2 F(x, y)}{\partial x \partial y} \right|_{(x, y) = (x_0, y_0)} = \left. \frac{\partial^2 F(x, y)}{\partial y \partial x} \right|_{(x, y) = (x_0, y_0)} = f(x_0, y_0).$$

Ejemplo 4.2 Un dado y una moneda equilibrados son lanzados. Sean X = "resultado obtenido en el lanzamiento del dado" e $Y = 0$ si se obtiene cruz, y 1 si se obtiene cara en el lanzamiento de la moneda. Tomamos $\mathcal{D}_{(X, Y)} = \{(i, j) : i \in \{1, \dots, 6\}, j \in \{0, 1\}\}$ y $p_{ij} = P(X = i, Y = j) = 1/12$, $(i, j) \in \mathcal{D}_{(X, Y)}$. Entonces,

$$F(x, y) = \begin{cases} 0, & \text{si } (x < 1) \text{ o si } (y < 0), \\ 1/12, & \text{si } (1 \leq x < 2, 0 \leq y < 1), \\ 1/6, & \text{si } (2 \leq x < 3, 0 \leq y < 1) \text{ o si } (1 \leq x < 2, 1 \leq y), \\ 1/4, & \text{si } (3 \leq x < 4, 0 \leq y < 1), \\ 1/3, & \text{si } (4 \leq x < 5, 0 \leq y < 1) \text{ o si } (2 \leq x < 3, 1 \leq y), \\ 5/12, & \text{si } (5 \leq x < 6, 0 \leq y < 1), \\ 1/2, & \text{si } (6 \leq x, 0 \leq y < 1) \text{ o si } (3 \leq x < 4, 1 \leq y), \\ 2/3, & \text{si } (4 \leq x < 5, 1 \leq y), \\ 5/6, & \text{si } (5 \leq x < 6, 1 \leq y), \\ 1, & \text{si } (6 \leq x, 1 \leq y). \end{cases}$$

Ejemplo 4.3 Sea (X, Y) un vector aleatorio con función de densidad

$$f(x, y) = \begin{cases} e^{-(x+y)}, & \text{si } 0 < x, 0 < y, \\ 0, & \text{en caso contrario.} \end{cases}$$

Entonces, $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du$; es decir, si $x > 0$ e $y > 0$,

$$F(x, y) = \int_0^x \int_0^y e^{-(u+v)} dv du = (1 - e^{-x})(1 - e^{-y}).$$

En caso contrario, $F(x, y) = 0$.

4.4 Distribuciones marginales y condicionadas

Distribuciones marginales

Sea (X, Y) un vector aleatorio 2-dimensional con función de masa

$$p_{ij} = P(X = x_i, Y = y_j), \quad (x_i, y_j) \in \mathcal{D}_{(X,Y)}.$$

Entonces,

$$\begin{aligned} p_{i\bullet} &= \sum_{j=1}^{\infty} p_{ij} = \sum_{j=1}^{\infty} P(X = x_i, Y = y_j) = P(X = x_i), \\ p_{\bullet j} &= \sum_{i=1}^{\infty} p_{ij} = \sum_{i=1}^{\infty} P(X = x_i, Y = y_j) = P(Y = y_j), \end{aligned}$$

definen las **funciones de masa marginales de X e Y** , respectivamente.

Ejemplo 4.2 (Continuación) En el ejemplo 4.2 del lanzamiento de un dado y una moneda equilibrados, se tiene

X:	1	2	3	4	5	6	$p_{\bullet j}$
Y:							
0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
$p_{i\bullet}$:	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	

Si (X, Y) es continuo con función de densidad $f(x, y)$, entonces las respectivas **funciones de densidad marginales de X e Y** son

$$\begin{aligned} f_1(x) &= \int_{-\infty}^{\infty} f(x, v) dv, \quad \forall x \in \mathbb{R}, \\ f_2(y) &= \int_{-\infty}^{\infty} f(u, y) du, \quad \forall y \in \mathbb{R}. \end{aligned}$$

En términos de funciones de distribución, las situaciones anteriores conducen a la **función de distribución marginal de X**

$$F_1(x) = F(x, \infty) = \begin{cases} \sum_{i: x_i \leq x} p_{i\bullet}, & \text{si } (X, Y) \text{ es discreto,} \\ \int_{-\infty}^x f_1(u) du, & \text{si } (X, Y) \text{ es continuo,} \end{cases}$$

y a la **función de distribución marginal de Y**

$$F_2(y) = F(\infty, y) = \begin{cases} \sum_{j: y_j \leq y} p_{\bullet j}, & \text{si } (X, Y) \text{ es discreto,} \\ \int_{-\infty}^y f_2(v) dv, & \text{si } (X, Y) \text{ es continuo.} \end{cases}$$

Con generalidad, dada la función de distribución $F(x_1, \dots, x_n)$ de un vector aleatorio (n -dimensional) (X_1, \dots, X_n) puede obtenerse cualquier función de distribución marginal k -dimensional desde F . En concreto, la función de distribución marginal de $(X_{i_1}, \dots, X_{i_k})$, con $1 \leq i_1 < \dots < i_k \leq n$, viene dada por

$$F(\infty, \dots, \infty, x_{i_1}, \infty, \dots, \infty, x_{i_k}, \infty, \dots, \infty).$$

Distribuciones condicionadas

Sea (X, Y) un vector aleatorio 2-dimensional discreto con función de masa $\{p_{ij} : \forall (x_i, y_j) \in \mathcal{D}_{(X,Y)}\}$. Desde la definición de probabilidad condicionada,

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

supuesto que $P(B) > 0$. Entonces, podemos definir la **función de masa condicionada de X , dado el suceso $\{Y = y_j\}$** , mediante

$$P(X = x_i | Y = y_j) = \frac{p_{ij}}{p_{\bullet j}},$$

supuesto que $p_{\bullet j} > 0$. Análogamente, la **función de masa condicionada de Y , dado el suceso $\{X = x_i\}$** , viene dada por

$$P(Y = y_j | X = x_i) = \frac{p_{ij}}{p_{i\bullet}},$$

supuesto que $p_{i\bullet} > 0$.

Si (X, Y) es continuo, entonces las probabilidades $P(X \in A | Y = y)$ y $P(Y \in B | X = x)$ no están definidas porque $P(Y = y) = 0$ y $P(X = x) = 0$. En este caso, la **función de distribución condicionada de X , dado que $Y = y$** , viene definida desde

$$\lim_{\varepsilon \rightarrow 0+} P(X \leq x | Y \in (y - \varepsilon, y + \varepsilon]),$$

supuesto que el límite existe. Si el límite existe, es denotado por $F_{X|Y}(x|y)$ y se define la **función de densidad condicionada de X , dado que $Y = y$** , como una función real $f_{X|Y}(x|y)$ no-negativa que satisface

$$F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(u|y) du, \quad \forall x \in \mathbb{R}.$$

Es fácil comprobar que, en cada punto (x, y) donde f es continua y $f_2(y) > 0$ y es continua, se tiene

$$\begin{aligned} F_{X|Y}(x|y) &= \lim_{\varepsilon \rightarrow 0+} \frac{P(X \leq x, y - \varepsilon < Y \leq y + \varepsilon)}{P(y - \varepsilon < Y \leq y + \varepsilon)} \\ &= \lim_{\varepsilon \rightarrow 0+} \frac{\int_{-\infty}^x \frac{1}{2\varepsilon} \int_{y-\varepsilon}^{y+\varepsilon} f(u, v) dv du}{\frac{1}{2\varepsilon} \int_{y-\varepsilon}^{y+\varepsilon} f_2(v) dv} \\ &= \int_{-\infty}^x \frac{f(u, y)}{f_2(y)} du, \end{aligned}$$

puesto que

$$\begin{aligned}\lim_{\varepsilon \rightarrow 0+} \int_{-\infty}^x \frac{1}{2\varepsilon} \int_{y-\varepsilon}^{y+\varepsilon} f(u, v) dv du &= \int_{-\infty}^x \left(\lim_{\varepsilon \rightarrow 0+} \frac{1}{2\varepsilon} \int_{y-\varepsilon}^{y+\varepsilon} f(u, v) dv \right) du \\ &= \int_{-\infty}^x f(u, y) du, \\ \lim_{\varepsilon \rightarrow 0+} \frac{1}{2\varepsilon} \int_{y-\varepsilon}^{y+\varepsilon} f_2(v) dv &= \lim_{\varepsilon \rightarrow 0+} \frac{F_2(y + \varepsilon) - F_2(y - \varepsilon)}{2\varepsilon} = f_2(y).\end{aligned}$$

Por lo tanto, existe una función de densidad condicionada de X , dado que $Y = y$, que viene dada por

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_2(y)}, \quad \text{cuando } f_2(y) > 0.$$

Este argumento es la demostración del siguiente teorema:

Teorema 4.1 Sean f la función de densidad conjunta del vector aleatorio (X, Y) de tipo continuo y f_2 la función de densidad marginal de Y . En cada punto (x, y) donde f es continua, $f_2(y) > 0$ y $f_2(y)$ es continua, la función de densidad condicionada de X , dado que $\{Y = y\}$, existe y admite la representación

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_2(y)}.$$

Desde el teorema 4.1 es posible establecer

$$\begin{aligned}F_1(x) &= F(x, \infty) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(u, v) dv du \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^x f(u, v) du dv = \int_{-\infty}^{\infty} f_2(v) F_{X|Y}(x|v) dv.\end{aligned}$$

Observación 4.1 La igualdad $F_1(x) = F_2(y) F_{X|Y}(x|y)$ no es correcta.

Observación 4.2 Expresiones paralelas son obtenidas para la distribución condicionada de Y , dado que $X = x$, mediante el intercambio de componentes.

Observación 4.3 Los argumentos anteriores son válidos para vectores aleatorios n -dimensionales, $n \geq 2$, supuesto que (X, Y) corresponde a una partición del tipo $X = (Z_1, \dots, Z_k)$ e $Y = (Z_{k+1}, \dots, Z_n)$. En este caso, es suficiente con reemplazar las integrales del tipo $\int_{-\infty}^{\infty}$ por $\int_{\mathbb{R}^k}$ e $\int_{\mathbb{R}^{n-k}}$, según el caso.

Una noción relacionada con el concepto de distribución condicionada es la noción de distribución truncada.

Definición 4.4 Sean X una variable aleatoria definida sobre (Ω, \mathcal{A}, P) y $B \in \beta$

tal que $0 < P(X \in B) < 1$. Entonces, la distribución condicionada $P(X \leq x|X \in B)$ definida para cualquier $x \in \mathbb{R}$ se denomina *distribución truncada de X sobre el conjunto B* .

Si X es discreta con función de masa $p_i = P(X = x_i)$, $i \geq 1$, entonces la distribución truncada de X sobre B viene dada por

$$P(X = x_i|X \in B) = \frac{P(X = x_i, X \in B)}{P(X \in B)} = \begin{cases} \frac{p_i}{\sum_{x_j \in B} p_j}, & \text{si } x_i \in B, \\ 0 & \text{si } x_i \notin B. \end{cases}$$

Similarmente, cuando X es continuo, se tiene

$$P(X \leq x|X \in B) = \frac{P(X \leq x, X \in B)}{P(X \in B)} = \frac{\int_{(-\infty, x] \cap B} f(u) du}{\int_B f(u) du},$$

y, como resultado, la función de densidad de la distribución truncada de X sobre B tiene la forma

$$h_B(x) = \begin{cases} \frac{f(x)}{\int_B f(u) du}, & \text{si } x \in B, \\ 0, & \text{si } x \notin B. \end{cases}$$

4.5 Independencia de variables aleatorias

La distribución conjunta de un vector aleatorio determina de forma única las distribuciones marginales de sus componentes. No obstante, el conocimiento de las distribuciones marginales no necesariamente determina la distribución conjunta, salvo en presencia de la propiedad de independencia.

Sean $F(x, y)$, $F_X(x)$ y $F_Y(y)$ las funciones de distribución de (X, Y) , X e Y , respectivamente.

Definición 4.5 (Variables aleatorias independientes) Se dice que X e Y son independientes si y sólo si $F(x, y) = F_X(x)F_Y(y)$, $\forall (x, y) \in \mathbb{R}^2$.

Es inmediato comprobar que

(i) Si X e Y son discretas, la independencia de X e Y equivale a la condición

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j), \quad \forall (x_i, y_j) \in \mathcal{D}_{(X, Y)}.$$

(ii) Si X e Y son continuas, la condición es

$$f(x, y) = f_X(x)f_Y(y), \quad \forall (x, y) \in \mathbb{R}^2.$$

En ambos casos, si X e Y son variables aleatorias independientes, entonces

$$\begin{aligned} F_{Y|X}(y|x) &= F_Y(y), \quad \forall y \in \mathbb{R}, \\ F_{X|Y}(x|y) &= F_X(x), \quad \forall x \in \mathbb{R}. \end{aligned}$$

Es sabido que cualquier conjunto de Borel $A \in \beta$ puede ser expresado como una cantidad numerable de uniones, intersecciones y diferencias de intervalos semiabiertos. Como consecuencia, se tiene la siguiente caracterización de la independencia de dos variables aleatorias.

Teorema 4.2 *X e Y son independientes si y sólo si*

$$P(X \in A_1, Y \in A_2) = P(X \in A_1)P(Y \in A_2), \quad \forall A_1, A_2 \in \beta.$$

Además, las transformaciones medibles de variables aleatorias independientes conservan la independencia.

Teorema 4.3 *Sean X e Y variables aleatorias independientes, y sean f y g funciones medibles Borel. Entonces, $f(X)$ y $g(Y)$ son independientes.*

Demostración.– Tenemos

$$\begin{aligned} P(f(X) \leq x, g(Y) \leq y) &= P(X \in f^{-1}(-\infty, x], Y \in g^{-1}(-\infty, y]) \\ &= P(X \in f^{-1}(-\infty, x])P(Y \in g^{-1}(-\infty, y]) \\ &= P(f(X) \leq x)P(g(Y) \leq y), \quad \forall (x, y) \in \mathbb{R}^2, \end{aligned}$$

por el teorema 4.2, lo cual equivale a la independencia de $f(X)$ y $g(Y)$. \square

Definición 4.5

- (a) *Una colección de variables aleatorias X_1, \dots, X_n se dice mutuamente independiente si y sólo si*

$$F(x_1, \dots, x_n) = \prod_{i=1}^n F_i(x_i), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n,$$

donde F es la función de distribución conjunta de (X_1, \dots, X_n) y F_i es la función de distribución marginal de X_i .

- (b) *Se dice que X_1, \dots, X_n son independiente dos-a-dos si y sólo si cada par de ellas son independientes.*
- (c) *Una sucesión $\{X_n : n \geq 1\}$ se dice independiente si y sólo si, para cada $n \geq 2$, las variables X_1, \dots, X_n son mutuamente independientes.*
- (d) *Dos vectores aleatorios (X_1, \dots, X_m) e (Y_1, \dots, Y_n) se dicen independientes si y sólo si*

$$F(x_1, \dots, x_m, y_1, \dots, y_n) = F_1(x_1, \dots, x_m)F_2(y_1, \dots, y_n),$$

$\forall (x_1, \dots, x_m) \in \mathbb{R}^m$ y $\forall (y_1, \dots, y_n) \in \mathbb{R}^n$, donde F , F_1 y F_2 son las funciones de distribución conjunta de $(X_1, \dots, X_m, Y_1, \dots, Y_n)$ y marginales de (X_1, \dots, X_m) e (Y_1, \dots, Y_n) , respectivamente.

Observemos que si la colección $\{X_1, \dots, X_n\}$ es mutuamente independiente, entonces cualquier subcolección $\{X_{i_1}, \dots, X_{i_k}\}$ también es mutuamente independiente. No obstante, la independencia dos-a-dos de $\{X_1, \dots, X_n\}$ no implica necesariamente que la colección $\{X_1, \dots, X_n\}$ sea mutuamente independiente.

Para finalizar, señalamos el siguiente resultado.

Teorema 4.4 Sean $X = (X_1, \dots, X_m)$ e $Y = (Y_1, \dots, Y_n)$ vectores aleatorios independientes. Entonces, la componente X_j de X y la componente Y_k de Y son independientes. Si h y g son funciones medibles Borel, entonces $h(X)$ y $g(Y)$ son independientes.

4.6 Transformaciones medibles de vectores aleatorios

Cuando $X = (X_1, \dots, X_n)$ es un vector aleatorio discreto no hay dificultad en establecer la distribución de una transformación medible de X , $U = g(X)$.

Sea $\mathcal{D}_X \subset \mathbb{R}^n$ el soporte de X ; es decir, $P(X \in \mathcal{D}_X) = 1$ de modo que todo punto de \mathcal{D}_X tiene masa estrictamente positiva. Sea la transformación

$$\begin{cases} u_1 = g_1(x_1, \dots, x_n), \\ \vdots \\ u_n = g_n(x_1, \dots, x_n), \end{cases}$$

que define una aplicación uno-a-uno $g : \mathcal{D}_X \longrightarrow \mathcal{D}_{g(X)}$. Entonces, para $u = (u_1, \dots, u_n) \in \mathcal{D}_{g(X)}$, se tiene

$$\begin{aligned} P(U = u) &= P(g_1(X) = u_1, \dots, g_n(X) = u_n) \\ &= P(X_1 = h_1(u), \dots, X_n = h_n(u)), \end{aligned}$$

donde

$$\begin{cases} x_1 = h_1(u_1, \dots, u_n), \\ \vdots \\ x_n = h_n(u_1, \dots, u_n), \end{cases}$$

define la transformación inversa $h = (h_1, \dots, h_n)$ de g . Además, $P(U = u) = 0$, para cada $u \notin \mathcal{D}_{g(X)}$.

Ejemplo 4.3 Consideramos la distribución Binomial negativa bi-variante, con función de masa

$$P(X = x, Y = y) = \frac{(x + y + k - 1)!}{x!y!(k - 1)!} p_1^x p_2^y (1 - p_1 - p_2)^k,$$

donde $x, y \in \{0, 1, \dots\}$, $k \geq 1$, y $p_1, p_2 \in (0, 1)$ satisfacen $p_1 + p_2 < 1$. Para determinar la función de masa de $U = X + Y$, introducimos

una variable aleatoria auxiliar $V = Y$, de forma que

$$\begin{cases} u = x + y, \\ v = y, \end{cases}$$

representa una transformación uno-a-uno definida desde $\mathcal{D}_{(X,Y)} = \{(x, y) : x, y \in \{0, 1, \dots\}\}$ en el conjunto $\mathcal{D}_{(U,V)} = \{(u, v) : u \in \{0, 1, \dots\}, v \in \{0, 1, \dots, u\}\}$. La aplicación inversa viene dada por

$$\begin{cases} x = u - v, \\ y = v. \end{cases}$$

Entonces, la función de masa de (U, V) es

$$\begin{aligned} P(U = u, V = v) &= P(X = u - v, Y = v) \\ &= \frac{(u + k - 1)!}{(u - v)!v!(k - 1)!} p_1^{u-v} p_2^v (1 - p_1 - p_2)^k, \end{aligned}$$

cuando $(u, v) \in \mathcal{D}_{(U,V)}$, y $P(U = u, V = v) = 0$, si $(u, v) \notin \mathcal{D}_{(U,V)}$.

Observación 4.4 Si la aplicación g no es una aplicación uno-a-uno de \mathcal{D}_X en \mathcal{D}_U , pero para cada $u \in \mathcal{D}_U$ existen un número finito (incluso numerable) de imágenes inversas $x(1), \dots, x(k(u))$, entonces

$$\begin{aligned} P(U = u) &= P(g_1(X) = u_1, \dots, g_n(X) = u_n) \\ &= \sum_{j=1}^{k(u)} P(g_1(x(j)) = u_1, \dots, g_n(x(j)) = u_n). \end{aligned}$$

El caso continuo es analíticamente más sofisticado. El teorema 4.5 es el homólogo multidimensional a los teoremas 3.11 y 3.12.

Teorema 4.5 Sea $X = (X_1, \dots, X_n)$ un vector aleatorio n -dimensional continuo con función de densidad conjunta $f_X(x_1, \dots, x_n)$. Asumimos que

(a) Las igualdades

$$\begin{cases} u_1 = g_1(x_1, \dots, x_n), \\ \vdots \\ u_n = g_n(x_1, \dots, x_n), \end{cases}$$

definen una transformación uno-a-uno de \mathbb{R}^n en \mathbb{R}^n ; es decir, se asume que existe la transformación inversa

$$\begin{cases} x_1 = h_1(u_1, \dots, u_n), \\ \vdots \\ x_n = h_n(u_1, \dots, u_n), \end{cases}$$

y está definida sobre el rango de la transformación.

- (b) Las transformaciones directa g e inversa h son continuas.
- (c) Las derivadas parciales $\frac{\partial x_i}{\partial u_j}$, para cualesquiera $1 \leq i, j \leq n$, existen y son continuas.
- (d) El Jacobiano J de la transformación inversa

$$J = \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right| = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \dots & \frac{\partial x_1}{\partial u_n} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} & \dots & \frac{\partial x_2}{\partial u_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial u_1} & \frac{\partial x_n}{\partial u_2} & \dots & \frac{\partial x_n}{\partial u_n} \end{vmatrix}$$

es no-nulo para cualquier punto (u_1, \dots, u_n) en el rango de la transformación

Entonces, el vector aleatorio $U = (U_1, \dots, U_n)$ tiene una función de distribución absolutamente continua y su función de densidad admite la representación

$$f_U(u_1, \dots, u_n) = f_X(h_1(u_1, \dots, u_n), \dots, h_n(u_1, \dots, u_n))|J|.$$

Ejemplo 4.4 Sean X_1, X_2 y X_3 variables aleatorias independientes e idénticamente distribuidas exponenciales de tasa $\lambda = 1$. Entonces, la función de densidad conjunta del vector aleatorio $X = (X_1, X_2, X_3)$ viene dada por

$$f_X(x_1, x_2, x_3) = e^{-(x_1+x_2+x_3)}, \quad x_1, x_2, x_3 > 0.$$

Sean

$$\begin{aligned} Y_1 &= X_1 + X_2 + X_3, \\ Y_2 &= \frac{X_1 + X_2}{X_1 + X_2 + X_3}, \\ Y_3 &= \frac{X_1}{X_1 + X_2}. \end{aligned}$$

Entonces, la transformación inversa

$$\begin{aligned} x_1 &= y_1 y_2 y_3, \\ x_2 &= y_1 y_2 (1 - y_3), \\ x_3 &= y_1 (1 - y_2), \end{aligned}$$

tiene por Jacobiano

$$J = \begin{vmatrix} y_2 y_3 & y_1 y_3 & y_1 y_2 \\ y_2(1 - y_3) & y_1(1 - y_3) & -y_1 y_2 \\ 1 - y_2 & -y_1 & 0 \end{vmatrix} = -y_1^2 y_2.$$

El recinto original $\mathcal{C}_X = \{(x_1, x_2, x_3) : x_1, x_2, x_3 > 0\}$ se transforma en $\mathcal{C}_{(Y_1, Y_2, Y_3)} = \{(y_1, y_2, y_3) : y_1 > 0, 0 < y_2 < 1, 0 < y_3 < 1\}$, de modo que

$$\begin{aligned} f_{(Y_1, Y_2, Y_3)}(y_1, y_2, y_3) &= f_X(y_1 y_2 y_3, y_1 y_2 (1 - y_3), y_1 (1 - y_2)) |J| \\ &= y_1^2 y_2 e^{-y_1}, \end{aligned}$$

para $y_1 > 0, 0 < y_2 < 1$ y $0 < y_3 < 1$.

Sean $X = (X_1, \dots, X_n)$ un vector aleatorio n -dimensional continuo, con función de densidad conjunta $f_X(x_1, \dots, x_n)$, y la transformación de \mathbb{R}^n en \mathbb{R}^n

$$\begin{cases} u_1 = g_1(x_1, \dots, x_n), \\ \vdots \\ u_n = g_n(x_1, \dots, x_n). \end{cases}$$

Supongamos que, para cada $u = (u_1, \dots, u_n)$, la transformación $g = (g_1, \dots, g_n)$ tiene un número finito $k = k(u)$ de inversas. Supongamos que \mathbb{R}^n puede descomponerse en k conjuntos disjuntos A_1, \dots, A_k tales que, para cada $1 \leq i \leq k$, la transformación g de A_i en \mathbb{R}^n es una transformación uno-a-uno con transformación inversa

$$\begin{cases} x_1 = h_{1i}(u_1, \dots, u_n), \\ \vdots \\ x_n = h_{ni}(u_1, \dots, u_n). \end{cases}$$

Además, supongamos que las derivadas parciales primeras son continuas y que cada Jacobiano

$$J_i = \begin{vmatrix} \frac{\partial h_{1i}}{\partial u_1} & \frac{\partial h_{1i}}{\partial u_2} & \dots & \frac{\partial h_{1i}}{\partial u_n} \\ \frac{\partial h_{2i}}{\partial u_1} & \frac{\partial h_{2i}}{\partial u_2} & \dots & \frac{\partial h_{2i}}{\partial u_n} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial h_{ni}}{\partial u_1} & \frac{\partial h_{ni}}{\partial u_2} & \dots & \frac{\partial h_{ni}}{\partial u_n} \end{vmatrix}$$

es no-nulo en el rango de la transformación. Entonces, una reelaboración del teorema 4.5 sobre los conjuntos A_1, \dots, A_k conduce a la siguiente expresión para la función de densidad conjunta de $U = g(X)$:

$$f_U(u_1, \dots, u_n) = \sum_{i=1}^k f_X(h_{1i}(u_1, \dots, u_n), \dots, h_{ni}(u_1, \dots, u_n)) |J_i|.$$

Ejemplo 4.5 Sea (X, Y) un vector aleatorio Normal bivalente; es decir, la función de densidad conjunta de (X, Y) es

$$\begin{aligned} f_{(X, Y)}(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right) \right\}, \end{aligned}$$

para $(x, y) \in \mathbb{R}^2$, donde $(\mu_1, \mu_2) \in \mathbb{R}^2$, $\sigma_1 > 0$, $\sigma_2 > 0$ y $|\rho| < 1$. Sean $U_1 = \sqrt{X^2 + Y^2}$ y $U_2 = X/Y$. Si $u_1 > 0$, entonces

$$\begin{cases} u_1 = \sqrt{x^2 + y^2}, \\ u_2 = \frac{x}{y}, \end{cases}$$

admite dos soluciones (x_1, y_1) y (x_2, y_2) , siendo

$$\begin{cases} x_1 = \frac{u_1 u_2}{\sqrt{1+u_2^2}}, \\ y_1 = \frac{u_1}{\sqrt{1+u_2^2}}, \end{cases} \quad \text{y} \quad \begin{cases} x_2 = \frac{-u_1 u_2}{\sqrt{1+u_2^2}}, \\ y_2 = \frac{-u_1}{\sqrt{1+u_2^2}}, \end{cases}$$

para cualquier $u_2 \in \mathbb{R}$. Los Jacobianos asociados vienen dados por

$$\begin{aligned} J_1 &= \begin{vmatrix} \frac{u_2}{\sqrt{1+u_2^2}} & \frac{u_1}{(1+u_2^2)^{3/2}} \\ \frac{1}{\sqrt{1+u_2^2}} & \frac{-u_1 u_2}{(1+u_2^2)^{3/2}} \end{vmatrix} = \frac{-u_1}{1+u_2^2}, \\ J_2 &= \begin{vmatrix} \frac{-u_2}{\sqrt{1+u_2^2}} & \frac{-u_1}{(1+u_2^2)^{3/2}} \\ \frac{-1}{\sqrt{1+u_2^2}} & \frac{u_1 u_2}{(1+u_2^2)^{3/2}} \end{vmatrix} = \frac{-u_1}{1+u_2^2}. \end{aligned}$$

Entonces,

$$\begin{aligned} f_{(U_1, U_2)}(u_1, u_2) &= \frac{u_1}{1+u_2^2} \left(f_{(X, Y)} \left(\frac{u_1 u_2}{\sqrt{1+u_2^2}}, \frac{u_1}{\sqrt{1+u_2^2}} \right) \right. \\ &\quad \left. + f_{(X, Y)} \left(\frac{-u_1 u_2}{\sqrt{1+u_2^2}}, \frac{-u_1}{\sqrt{1+u_2^2}} \right) \right), \end{aligned}$$

para $u_1 > 0$ y $u_2 \in \mathbb{R}$.

4.7 Momentos y matriz de varianza-covarianzas. Recta de regresión

Sean (X, Y) un vector aleatorio 2-dimensional y $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ una función medible. Para especificar la esperanza de $g(X, Y)$ debe distinguirse entre los casos discreto y continuo:

(i) Si (X, Y) es discreto y $\sum_{i,j} p_{ij} |g(x_i, y_j)| < \infty$, entonces

$$E[g(X, Y)] = \sum_{i,j} p_{ij} g(x_i, y_j).$$

(ii) Si (X, Y) es continuo e $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) |g(u, v)| dv du < \infty$, entonces

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) g(u, v) dv du.$$

Nos interesan funciones del tipo $g(x, y) = x^n y^m$.

Definición 4.6 (Momentos de un vector aleatorio)

(a) Si existe $E[X^n Y^m]$, llamamos momento (conjunto) de orden $n+m$ a cada valor

$$m_{nm} = E[X^n Y^m].$$

(b) Si existe $E[(X - E[X])^n (Y - E[Y])^m]$, llamamos momento (conjunto) central de orden $n+m$ a cada valor

$$\mu_{nm} = E[(X - E[X])^n (Y - E[Y])^m].$$

En particular, destacamos a las **medias**

$$m_{10} = E[X],$$

$$m_{01} = E[Y],$$

las **varianzas**

$$\mu_{20} = \text{Var}(X) = E[(X - E[X])^2],$$

$$\mu_{02} = \text{Var}(Y) = E[(Y - E[Y])^2],$$

y la **covarianza**

$$\text{Cov}(X, Y) = \mu_{11} = E[(X - E[X])(Y - E[Y])],$$

donde también podemos calcular $\text{Cov}(X, Y) = m_{11} - m_{10}m_{01}$. A continuación, analizamos el valor de $\text{Cov}(X, Y)$ bajo independencia. (La demostración del teorema 4.6 es inmediata, por lo que es omitida.)

Teorema 4.6

(a) Si X e Y son independientes, entonces $\text{Cov}(X, Y) = 0$; es decir, X e Y son incorreladas.

(b) Sean X_1, \dots, X_n variables aleatorias independientes tales que $E[|X_i|] < \infty$, $1 \leq i \leq n$. Entonces, la esperanza $E[\prod_{i=1}^n X_i]$ existe y $E[\prod_{i=1}^n X_i] = \prod_{i=1}^n E[X_i]$.

El teorema 4.7 muestra un procedimiento de cálculo para la varianza de la suma de variables aleatorias:

Teorema 4.7 Sean X_1, \dots, X_n variables aleatorias tales que $E[X_i^2] < \infty$, $1 \leq i \leq n$. Sean $a_1, \dots, a_n \in \mathbb{R}$ y $S = \sum_{i=1}^n a_i X_i$. Entonces, $\text{Var}(S)$ existe y viene dada por

$$\text{Var}(S) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_i a_j \text{Cov}(X_i, X_j).$$

Si, en particular, X_1, \dots, X_n son tales que $Cov(X_i, X_j) = 0, \forall i \neq j$, entonces

$$Var(S) = \sum_{i=1}^n a_i^2 Var(X_i).$$

Demostración.– Evaluamos

$$\begin{aligned} Var(S) &= E \left[\left(\sum_{i=1}^n a_i X_i - \sum_{i=1}^n a_i E[X_i] \right)^2 \right] = E \left[\left(\sum_{i=1}^n a_i (X_i - E[X_i]) \right)^2 \right] \\ &= E \left[\sum_{i=1}^n a_i^2 (X_i - E[X_i])^2 + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_i a_j (X_i - E[X_i])(X_j - E[X_j]) \right] \\ &= \sum_{i=1}^n a_i^2 Var(X_i) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_i a_j Cov(X_i, X_j), \end{aligned}$$

por la linealidad del operador esperanza. \square

Coeficiente de correlación

El coeficiente de correlación entre dos variables aleatorias X e Y es una medida del grado de dependencia lineal entre ellas.

Definición 4.7 *Supuesto que $E[X^2]$ y $E[Y^2]$ existen, se define el coeficiente de correlación entre X e Y como*

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

Entonces, se dice que X e Y son incorreladas si $\rho = 0$.

Notemos que ρ es una medida adimensional (no dependiente de las unidades de medida); es decir, como $Cov(aX, bY) = abCov(X, Y)$, se sigue que

$$\rho(aX, bY) = \rho(X, Y), \quad \forall a, b \neq 0.$$

Es claro que $\rho = 0$ si y sólo si $Cov(X, Y) = 0$. Además, si X e Y son independientes, entonces $\rho = 0$. No obstante, cuando $\rho = 0$, no necesariamente X e Y son independientes. La excepción es el caso Normal.

Teorema 4.8 *Sea (X, Y) un vector aleatorio Normal; es decir, con función de densidad*

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x,y)},$$

donde

$$Q(x, y) = \frac{1}{1-\rho^2} \left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right).$$

Entonces, X e Y son independientes si y sólo si $\rho = 0$, en cuyo caso, X e Y son Normales con parámetros (μ_1, σ_1^2) y (μ_2, σ_2^2) , respectivamente.

Demostración.— En primer lugar, escribamos

$$f_{(X,Y)}(x,y) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\} \\ \times \frac{1}{\sigma_2\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{(y-\beta_x)^2}{2\sigma_2^2(1-\rho^2)}\right\},$$

donde

$$\beta_x = \mu_2 + \rho\sqrt{\frac{\sigma_2^2}{\sigma_1^2}}(x-\mu_1).$$

Entonces, una simple sustitución muestra que $\rho = 0$ es una condición necesaria y suficiente para que $f_{(X,Y)}(x,y) = f_X(x)f_Y(y)$, $\forall (x,y) \in \mathbb{R}^2$, puesto que

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dy = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}, \\ f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)} = \frac{1}{\sigma_2\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{(y-\beta_x)^2}{2\sigma_2^2(1-\rho^2)}\right\},$$

con lo cual concluye la demostración. \square

La **matriz de varianza-covarianzas** de X e Y

$$\Sigma = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix}$$

agrupa la información sobre la dispersión de las variables aleatorias y la correlación entre ellas. Se trata de una matriz simétrica y semidefinida positiva. Por tanto, su determinante es no-negativo y $Cov^2(X,Y) \leq Var(X)Var(Y)$; es decir, $-1 \leq \rho \leq 1$, siendo ρ del mismo signo que $Cov(X,Y)$.

Recta de regresión

Los modelos de regresión tienen por finalidad determinar una relación funcional entre variables aleatorias X e Y , denotada por

$$y = h(x).$$

Asumamos que existen $E[(h(X))^2]$ y $E[Y^2]$. El principio de mínimos cuadrados consiste en elegir $h(x)$ de forma que se minimice la expresión

$$E[(Y - h(X))^2].$$

Si, por ejemplo, (X, Y) es un vector continuo, entonces debe minimizarse

$$\begin{aligned} E[(Y - h(X))^2] &= \int_{\mathbb{R}^2} (v - h(u))^2 f_{(X,Y)}(u, v) du dv \\ &= \int_{\mathbb{R}} f_X(u) \int_{\mathbb{R}} (v - h(u))^2 f_{Y|X}(v|u) dv du. \end{aligned}$$

Es claro que $E[(Y - h(X))^2]$ se minimiza cuando

$$\int_{\mathbb{R}} (v - h(u))^2 f_{Y|X}(v|u) dv$$

alcanza su mínimo. Por tanto, la función

$$h(x) = E[Y|X = x]$$

es la relación buscada.

Definición 4.8 Se llama *curva de regresión de Y sobre X* a

$$y = E[Y|X = x].$$

Análogamente, la *curva de regresión de X sobre Y* viene dada por $x = E[X|Y = y]$.

Habitualmente, se tiene interés en explicar la dependencia entre X e Y mediante una relación lineal

$$h(x) = a + bx,$$

donde $a, b \in \mathbb{R}$. Su utilidad radica en que muchas relaciones no-lineales pueden reducirse a relaciones lineales transformando adecuadamente las variables aleatorias originales. En cualquier caso, una relación lineal debe de ser considerada como una aproximación sencilla a un tipo de relación más compleja.

Sea (X, Y) un vector aleatorio 2-dimensional con $E[X^2] < \infty$ y $E[Y^2] < \infty$. Sean μ y Σ el vector de medias y la matriz de varianza-covarianzas; es decir,

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

Para aproximar linealmente Y en términos de X , el principio de mínimos cuadrados requiere elegir las constantes (a, b) de la relación $h(x) = a + bx$ de modo que se minimice la función

$$L(a, b) = E[(Y - (a + bX))^2].$$

Entonces, $L(a, b) = E[Y^2] + a^2 + b^2 E[X^2] - 2bE[XY] + 2abE[X] - 2aE[Y]$. Resolvemos para (a, b) el sistema de ecuaciones normales

$$\begin{aligned} \frac{\partial L(a, b)}{\partial a} &= 2(a + bE[X] - E[Y]) = 0, \\ \frac{\partial L(a, b)}{\partial b} &= 2(bE[X^2] - E[XY] + aE[X]) = 0. \end{aligned}$$

Entonces,

$$\begin{aligned} a &= \mu_2 - \frac{\sigma_{12}}{\sigma_1^2} \mu_1, \\ b &= \frac{\sigma_{12}}{\sigma_1^2}. \end{aligned}$$

Definición 4.9

(a) La recta de regresión de Y sobre X viene dada por

$$y = \mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(x - \mu_1).$$

Análogamente, la recta de regresión de X sobre Y es

$$x = \mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(y - \mu_2).$$

(b) Si existen $E[X^2]$ y $E[Y^2]$ y son finitas, entonces la cantidad σ_{12}/σ_1^2 es llamada coeficiente de regresión de Y sobre X . Análogamente, σ_{12}/σ_2^2 es llamado coeficiente de regresión de X sobre Y . Son denotados por

$$\begin{aligned} \gamma_{Y|X} &= \frac{\sigma_{12}}{\sigma_1^2}, \\ \gamma_{X|Y} &= \frac{\sigma_{12}}{\sigma_2^2}. \end{aligned}$$

Asociada a la recta de regresión de Y sobre X

$$y = \mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(x - \mu_1).$$

se define el **residual** como la variable aleatoria que representa la parte de Y que no está influenciada linealmente por X ; es decir,

$$Z_{Y|X} = Y - Y^*,$$

donde

$$Y^* = \mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(X - \mu_1).$$

Sus principales momentos son

$$E[Z_{Y|X}] = 0 \quad \text{y} \quad \text{Var}(Z_{Y|X}) = \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}.$$

Si dividimos la expresión de la **varianza residual** $\text{Var}(Z_{Y|X})$ por σ_2^2 obtenemos una relación entre el coeficiente de correlación y la varianza residual

$$1 - \rho^2 = \frac{\text{Var}(Z_{Y|X})}{\sigma_2^2}.$$

Propiedades fundamentales de la recta de regresión de Y sobre X son

- (i) $\rho^2 = 1$ si y sólo si $P(Y = Z_{Y|X}) = 1$; es decir, ajuste lineal perfecto.
- (ii) $\rho = 0$ si y sólo si $P(Y = \mu_2) = 1$; es decir, X e Y son incorreladas.

Dado que las rectas de regresión de Y sobre X , y de X sobre Y pueden expresarse, respectivamente, como

$$\begin{aligned} y - \mu_2 &= \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1), \\ y - \mu_2 &= \frac{1}{\rho} \frac{\sigma_2}{\sigma_1} (x - \mu_1), \end{aligned}$$

el valor absoluto de la pendiente de la recta de regresión de X sobre Y es superior al correspondiente valor para la recta de regresión de Y sobre X , puesto que $|\rho| \leq 1$. Por tanto, las dos rectas de regresión son ascendentes o descendentes simultáneamente, siendo la recta de regresión de X sobre Y más próxima a la vertical que la recta de regresión de Y sobre X , intersecándose en el punto $(x, y) = (\mu_1, \mu_2)$.

4.8 Función característica

Definición 4.10 Sea (X, Y) un vector aleatorio 2-dimensional. Llamamos función característica de (X, Y) a

$$\varphi_{(X,Y)}(t, u) = E \left[e^{i(tX + uY)} \right], \quad (t, u) \in \mathbb{R}^2,$$

donde i es la unidad imaginaria.

La función característica $\varphi_{(X,Y)}(t, u)$ de la definición 4.10 tiene propiedades análogas a aquellas asociadas a una variable aleatoria unidimensional. En particular, si existe el momento m_{kl} , entonces

$$i^{k+l} m_{kl} = \left. \frac{\partial^{k+l} \varphi_{(X,Y)}(t, u)}{\partial t^k \partial u^l} \right|_{(t,u)=(0,0)}.$$

Además, es claro que

$$\begin{aligned} \varphi_{(X,Y)}(t, 0) &= \varphi_X(t), \quad t \in \mathbb{R}, \\ \varphi_{(X,Y)}(0, u) &= \varphi_Y(u), \quad u \in \mathbb{R}. \end{aligned}$$

A continuación enunciamos dos resultados fundamentales sobre la función característica de un vector aleatorio.

Teorema 4.9 Sea (X, Y) un vector aleatorio 2-dimensional con función característica $\varphi_{(X,Y)}(t, u)$. Entonces, X e Y son independientes si y sólo si

$$\varphi_{(X,Y)}(t, u) = \varphi_X(t) \varphi_Y(u), \quad \forall (t, u) \in \mathbb{R}^2.$$

Teorema 4.10 Sean X_1, \dots, X_n variables aleatorias independientes, con funciones características $\varphi_{X_1}(t), \dots, \varphi_{X_n}(t)$, respectivamente. Entonces, la función característica de $Z = \sum_{j=1}^n X_j$ viene dada por

$$\varphi_Z(t) = \prod_{j=1}^n \varphi_{X_j}(t).$$

4.9 Distribuciones multidimensionales notables

Multinomial. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y sean K sucesos $A_1, \dots, A_K \in \mathcal{A}$ tales que

$$\bigcup_{j=1}^K A_j = \Omega \quad \text{y} \quad A_i \cap A_j = \emptyset, \quad \forall i \neq j.$$

Supongamos un experimento aleatorio, que se repite N veces, de modo que en cada repetición ocurre uno y sólo uno de los K sucesos anteriores. Supongamos que las N repeticiones son independientes y denotemos por p_j la probabilidad de que en una de las repeticiones ocurra el suceso A_j , para $1 \leq j \leq K$; es decir, $P(A_j) = p_j$, con $p_j \geq 0$ y $\sum_{j=1}^K p_j = 1$.

Sea (X_1, \dots, X_{K-1}) un vector aleatorio cuya componente X_j determina, en las N repeticiones, el número de veces que aparece el suceso A_j , para $1 \leq j \leq K-1$. Observemos que el suceso A_K aparece $N - \sum_{j=1}^{K-1} X_j$ veces, puesto que $N = \sum_{j=1}^K X_j$. Entonces, el vector (X_1, \dots, X_{K-1}) así definido tiene distribución multinomial $(N; p_1, \dots, p_{K-1})$ y su función de masa tiene la forma

$$P(X_1 = x_1, \dots, X_{K-1} = x_{K-1}) = \frac{N!}{x_1! x_2! \dots x_K!} p_1^{x_1} p_2^{x_2} \dots p_K^{x_K},$$

donde $x_K = N - (x_1 + \dots + x_{K-1})$ y $p_K = 1 - (p_1 + \dots + p_{K-1})$, para valores $x_j \in \{0, 1, \dots, K-1\}$ con $\sum_{j=1}^{K-1} x_j \leq N$.

Es sencillo observar que

$$\begin{aligned} E[X_j] &= Np_j, \quad 1 \leq j \leq K-1, \\ \text{Var}(X_j) &= Np_j(1-p_j), \quad 1 \leq j \leq K-1, \\ \text{Cov}(X_i, X_j) &= -Np_i p_j, \quad i \neq j. \end{aligned}$$

Entonces, tenemos que

$$\rho(X_i, X_j) = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}},$$

por lo que $\rho(X_i, X_j)$ no depende de N .

Es interesante tener en cuenta los siguientes resultados:

- Las distribuciones marginales de las variables aleatorias X_j , para $1 \leq j \leq K-1$, son binomiales (N, p_j) .
- Las variables aleatorias X_1, \dots, X_{K-1} son independientes.
- Si $X = (X_1, \dots, X_{K-1})$ e $Y = (Y_1, \dots, Y_{K-1})$ son dos vectores aleatorios independientes que siguen distribuciones multinomiales de parámetros $(N; p_1, \dots, p_{K-1})$ y $(M; p_1, \dots, p_{K-1})$, respectivamente, entonces $Z = X + Y$ es multinomial de parámetro $(N + M; p_1, \dots, p_{K-1})$.
- Si $X = (X_1, \dots, X_{K-1})$ e $Y = (Y_1, \dots, Y_{K-1})$ son dos vectores aleatorios no-negativos, independientes, y si $Z = X + Y$ es multinomial de parámetro $(N; p_1, \dots, p_{K-1})$, entonces X e Y siguen distribuciones multinomiales con los mismos parámetros.

La función característica del vector aleatorio multinomial $(N; p_1, \dots, p_{K-1})$ viene dada por

$$\begin{aligned}\varphi(t_1, \dots, t_{K-1}) &= E \left[e^{i(t_1 X_1 + \dots + t_{K-1} X_{K-1})} \right] \\ &= (p_1 e^{it_1} + \dots + p_{K-1} e^{it_{K-1}} + p_K)^n.\end{aligned}$$

Normal multivariante. Un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ sigue una distribución Normal_n si su función de densidad en \mathbb{R}^n viene dada por

$$f_X(\mathbf{x}) = \frac{\sqrt{|R|}}{(2\pi)^{n/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T R (\mathbf{x}-\mu)},$$

para $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n$, siendo $\mu = (\mu_1, \dots, \mu_n)^T$ y

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{pmatrix}$$

una matriz cuadrada de orden n , definida positiva de elementos constantes. Entonces,

- La distribución marginal de cada una de las componentes X_j es Normal_1 (unidimensional).
- La distribución conjunta de cualesquiera $k < n$ componentes $(X_{i_1}, \dots, X_{i_k})$ de \mathbf{X} sigue una distribución Normal_k (multivariante).
- El vector de medias de \mathbf{X} viene dado por μ , de modo que $E[X_j] = \mu_j$, para cada $1 \leq j \leq n$.

- La matriz de varianza-covarianzas de \mathbf{X} es la inversa de R ; es decir,

$$\Sigma = R^{-1} = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{pmatrix},$$

siendo

$$s_{ij} = \begin{cases} Cov(X_i, X_j), & \text{si } i \neq j, \\ Var(X_i), & \text{si } i = j. \end{cases}$$

La función característica del vector aleatorio Normal multivariante viene dada por

$$\varphi(\mathbf{t}) = e^{i\mathbf{t}^T \mu - \frac{1}{2} \mathbf{t}^T R^{-1} \mathbf{t}},$$

donde $\mathbf{t} = (t_1, \dots, t_n)^T \in \mathbb{R}^n$.

Es habitual emplear como parámetros de la distribución Normal_n al par $(\mu, \Sigma) = (\mu, R^{-1})$. Entonces, cualquier combinación lineal de la forma

$$\sum_{j=1}^n a_j X_j$$

sigue una distribución Normal_1 de media y varianza dadas, respectivamente, por

$$\sum_{j=1}^n a_j \mu_j \quad \text{y} \quad \sum_{i=1}^n \sum_{j=1}^n a_i a_j s_{ij}.$$

Para finalizar, señalamos que si \mathbf{X} es $\text{Normal}_n(\mu, \Sigma)$, entonces las variables aleatorias X_1, \dots, X_n son independientes si y sólo si $s_{ij} = Cov(X_i, X_j) = 0$, $\forall i \neq j$; es decir, si y sólo si $\Sigma = R^{-1}$ es una matriz diagonal.

Tema 5. Convergencias estocásticas

Índice

- 5.1 Introducción
- 5.2 Tipos de convergencia
 - 5.2.1 Convergencia en ley
 - 5.2.2 Convergencia en probabilidad
 - 5.2.3 Convergencia en media de orden p
 - 5.2.4 Convergencia casi segura
- 5.3 Relaciones entre los tipos de convergencia. Propiedades elementales
- 5.4 Ley débil de los grandes números
- 5.5 Ley fuerte de los grandes números
- 5.6 Teorema central del límite
- 5.7 Teorema de la convergencia monótona y teorema de la convergencia dominada
- Ejercicios

5.1 Introducción

El presente tema está dedicado al estudio de propiedades de convergencia de sucesiones de variables aleatorias, así como de las medidas de probabilidad inducidas por éstas. Los tres resultados fundamentales son conocidos con los nombres de "ley débil de los grandes números", "ley fuerte de los grandes números" y "teorema central del límite", y resultan ser de notable importancia en los métodos estadísticos.

De manera análoga al análisis matemático, distinguimos entre distintos tipos de convergencia que determinan las cualidades y las limitaciones de cada *forma* de entender la convergencia en estudio.

5.2 Tipos de convergencia

Existen dos formas de entender la noción de convergencia en el cálculo de probabilidades: por una parte, mediante la convergencia de una sucesión de medidas

de probabilidad y, por otra parte, desde la convergencia de una sucesión de variables aleatorias. La convergencia en ley pertenece a la primera forma, mientras que las convergencias en probabilidad, en media de orden p y casi segura son relativas a una sucesión de variables aleatorias.

5.2.1 Convergencia en ley

Comenzamos definiendo la convergencia en ley de una sucesión $\{F_n : n \geq 1\}$ de funciones de distribución sobre \mathbb{R} o, equivalentemente, de la correspondiente sucesión $\{\mu_{F_n} : n \geq 1\}$ de medidas asociadas a éstas.

Definición 5.1 Sean $\{F_n : n \geq 1\}$ una sucesión de funciones de distribución y F una función de distribución sobre \mathbb{R} . Se dice que F_n converge en ley hacia F cuando $n \rightarrow \infty$, y es denotado por $F_n \rightarrow F$ (en ley), si se satisface

$$\lim_{n \rightarrow \infty} F_n(x) = F(x), \quad (7)$$

en todo punto x de continuidad de F .

En el caso en que $\{X_n : n \geq 1\}$ sea una sucesión de variables aleatorias (unidimensionales) definidas sobre el espacio de probabilidad (Ω, \mathcal{A}, P) , con $\{F_n : n \geq 1\}$ siendo la correspondiente sucesión de funciones de distribución⁹, la definición 5.1 nos permite decir que X_n converge en ley o *en distribución* hacia X si existe una variable aleatoria X con función de distribución F tal que F_n converge en ley hacia F , cuando $n \rightarrow \infty$. En este caso, denotamos $X_n \rightarrow X$ (en ley).

Ejemplo 5.1 Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas con función de densidad

$$f(x) = \begin{cases} 1/\theta, & \text{si } 0 < x < \theta, \\ 0, & \text{en caso contrario,} \end{cases}$$

donde $\theta > 0$. La función de distribución de $M_n = \max(X_1, \dots, X_n)$ viene dada por

$$\begin{aligned} F_n(x) &= P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) \\ &= \prod_{j=1}^n P(X_j \leq x) = (F(x))^n = \left(\int_{-\infty}^x f(u) du \right)^n, \end{aligned}$$

de modo que

$$F_n(x) = \begin{cases} 0, & \text{si } x < 0, \\ (x/\theta)^n, & \text{si } 0 \leq x < \theta, \\ 1, & \text{si } \theta \leq x. \end{cases}$$

⁹Es decir, para cada $n \geq 1$, $F_n = F_{X_n}$

Entonces, es inmediato verificar que

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) = \begin{cases} 0, & \text{si } x < \theta, \\ 1, & \text{si } \theta \leq x, \end{cases}$$

siendo F la función de distribución de una variable aleatoria X degenerada en el punto $x = \theta$. Por lo tanto, $F_n \rightarrow F$ (en ley), cuando $n \rightarrow \infty$ y, equivalentemente, $X_n \rightarrow X$ (en ley), cuando $n \rightarrow \infty$.

Es de interés observar que la condición (7) por sí sola no asegura, salvo en casos particulares, la convergencia de la sucesión $\{F_n : n \geq 1\}$ de funciones de distribución hacia una función F de distribución.

Ejemplo 5.2 Sea

$$F_n(x) = \begin{cases} 0, & \text{si } x < n, \\ 1, & \text{si } n \leq x. \end{cases}$$

Es decir, para cada $n \geq 1$, F_n es la función de distribución de una variable aleatoria X_n degenerada en el punto $x = n$. Entonces, es claro que $\lim_{n \rightarrow \infty} F_n(x) = F(x) \equiv 0, \forall x \in \mathbb{R}$.

El siguiente ejemplo muestra que la convergencia en ley de una sucesión $\{X_n : n \geq 1\}$ de variables aleatorias no es suficiente para asegurar la convergencia de las sucesiones de momentos $\{E[X_n^k] : n \geq 1\}$, para $k \in \mathbb{N}$.

Ejemplo 5.3 Sea

$$F_n(x) = \begin{cases} 0, & \text{si } x < 0, \\ 1 - 1/n, & \text{si } 0 \leq x < n, \\ 1, & \text{si } n \leq x, \end{cases}$$

es decir, para cada $n \geq 1$, F_n es la función de distribución de una variable aleatoria X_n con función de masa

$$P(X_n = 0) = 1 - \frac{1}{n}, \quad P(X_n = n) = \frac{1}{n}$$

y, como consecuencia, $E[X_n^k] = n^{k-1}$, para $k \geq 1$.

Por otra parte, $F_n \rightarrow F$ (en ley), cuando $n \rightarrow \infty$ si F viene dada por

$$F(x) = \begin{cases} 0, & \text{si } x < 0, \\ 1, & \text{si } 0 \leq x, \end{cases}$$

y tiene momentos $E[X^k] = 0, \forall k \geq 1$. Entonces, no es cierto que $E[X_n^k] \rightarrow E[X^k]$, cuando $n \rightarrow \infty$, para $k \geq 1$.

El ejemplo 5.4 pone de manifiesto que, aunque se tiene la convergencia puntual de la sucesión $\{F_n : n \geq 1\}$ de funciones de distribución en los puntos de continuidad de F , no necesariamente se tiene la convergencia de las funciones de masa o de densidad asociadas, según el caso.

Ejemplo 5.4 Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias con funciones de masa $p_n(x) = P(X_n = x) = 1$ si $x = 2 + 1/n$, y 0 en caso contrario. Entonces, se tiene que $\lim_{n \rightarrow \infty} p_n(x) = p(x) \equiv 0$, $\forall x \in \mathbb{R}$. Sin embargo, la sucesión de funciones de distribución¹⁰ $\{F_n : n \geq 1\}$ de las variables aleatorias $\{X_n : n \geq 1\}$ satisface

$$\lim_{n \rightarrow \infty} F_n(x) = F(x),$$

en todo punto x de continuidad de F cuando

$$F(x) = \begin{cases} 0, & \text{si } x < 2, \\ 1, & \text{si } 2 \leq x. \end{cases}$$

Es decir, F es la función de distribución de una variable aleatoria X degenerada en el punto $x = 2$ y tiene función de masa $p_X(x) = 1$ si $x = 2$, y 0 en caso contrario.

Los siguientes teoremas, que se proponen sin demostración, resumen algunos de los resultados de mayor importancia relativos a la convergencia en ley.

Teorema 5.1 Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias.

- (i) Para cada $n \geq 1$, sea $\{p_n(k) = P(X_n = k) : k \geq 0\}$ la función de masa de X_n , y sea $\{p(k) = P(X = k) : k \geq 0\}$ la función de masa de una variable aleatoria X . Entonces, $\lim_{n \rightarrow \infty} p_n(x) = p(x)$, $\forall x \in \mathbb{R}$, si y sólo si $X_n \rightarrow X$ (en ley), cuando $n \rightarrow \infty$.
- (ii) Para cada $n \geq 1$, sea f_n la función de densidad de X_n verificando que $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ para casi todo $x \in \mathbb{R}$ (es decir, excepto a lo sumo para una cantidad numerable de puntos x), donde f es la función de densidad de una variable aleatoria X . Entonces, $X_n \rightarrow X$ (en ley), cuando $n \rightarrow \infty$.

Teorema 5.2 Sean $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias y X una variable aleatoria tales que $X_n \rightarrow X$ (en ley), cuando $n \rightarrow \infty$. Entonces, $X_n + c \rightarrow X + c$ (en ley), cuando $n \rightarrow \infty$, $\forall c \in \mathbb{R}$, y $cX_n \rightarrow cX$ (en ley), cuando $n \rightarrow \infty$, $\forall c \neq 0$.

El teorema 5.3 establece una relación entre la convergencia en ley de una sucesión de funciones de distribución y la convergencia puntual de la correspondiente sucesión de funciones características.

¹⁰En este caso, se tiene

$$F_n(x) = \begin{cases} 0, & \text{si } x < 2 + \frac{1}{n}, \\ 1, & \text{si } 2 + \frac{1}{n} \leq x. \end{cases}$$

Teorema 5.3 (De continuidad de P. Levy) Sean $\{F_n : n \geq 1\}$ una sucesión de funciones de distribución sobre \mathbb{R} y $\{\varphi_n : n \geq 1\}$ la correspondiente sucesión de funciones características. Entonces,

- (i) Si $F_n \longrightarrow F$ (en ley), cuando $n \rightarrow \infty$, donde F es una función de distribución sobre \mathbb{R} con función característica φ , entonces

$$\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t), \quad \forall t \in \mathbb{R}.$$

- (ii) Si $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$, siendo φ una función continua en $t = 0$, entonces φ es la función característica de una función de distribución F tal que $F_n \longrightarrow F$ (en ley), cuando $n \rightarrow \infty$.

5.2.2 Convergencia en probabilidad

Definición 5.2 Sean $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias y X una variable aleatoria X definidas sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Se dice que $\{X_n : n \geq 1\}$ converge en probabilidad hacia X cuando $n \rightarrow \infty$, y es denotado por $X_n \longrightarrow X$ (en probabilidad), si y sólo si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\} = 0.$$

La definición de convergencia en probabilidad no ofrece información alguna acerca de la convergencia de las variables aleatorias X_n hacia X en el sentido estricto en que la convergencia es entendida en el análisis matemático. En concreto, si $X_n \longrightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, no necesariamente se tiene que, para $\varepsilon > 0$ dado, podamos encontrar un índice $N \in \mathbb{N}$ tal que $|X_n - X| < \varepsilon \forall n \geq N$. La definición 5.2 sólo establece la convergencia de la sucesión de probabilidades $P(|X_n - X| > \varepsilon)$ hacia 0 cuando $n \rightarrow \infty$.

Ejemplo 5.5 Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias con funciones de masa $P(X_n = 1) = 1/n$ y $P(X_n = 0) = 1 - 1/n$. Entonces,

$$P(|X_n| > \varepsilon) = \begin{cases} P(X_n = 1) = 1/n, & \text{si } 0 < \varepsilon < 1, \\ 0, & \text{si } 1 \leq \varepsilon. \end{cases}$$

Por lo tanto, $\lim_{n \rightarrow \infty} P(|X_n| > \varepsilon) = 0$ y $X_n \longrightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, donde $X \equiv 0$.

Las propiedades fundamentales de la convergencia en probabilidad de una sucesión de variables aleatorias aparecen recogidas en el teorema 5.4.

Teorema 5.4

- (a) $X_n \longrightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, si y sólo si $X_n - X \longrightarrow 0$ (en probabilidad), cuando $n \rightarrow \infty$.

- (b) Si $X_n \longrightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, y $X_n \longrightarrow Y$ (en probabilidad), cuando $n \rightarrow \infty$, entonces $P(X = Y) = 1$.
- (c) Si $X_n \longrightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, entonces $X_n - X_m \longrightarrow 0$ (en probabilidad), cuando $n, m \rightarrow \infty$.
- (d) Si $X_n \longrightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, e $Y_n \longrightarrow Y$ (en probabilidad), cuando $n \rightarrow \infty$, entonces $X_n \pm Y_n \longrightarrow X \pm Y$ (en probabilidad), cuando $n \rightarrow \infty$.
- (e) Si $X_n \longrightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, y $k \in \mathbb{R}$, entonces $kX_n \longrightarrow kX$ (en probabilidad), cuando $n \rightarrow \infty$.
- (f) Si $X_n \longrightarrow k$ (en probabilidad), cuando $n \rightarrow \infty$, con k constante, entonces $X_n^2 \longrightarrow k^2$ (en probabilidad), cuando $n \rightarrow \infty$.
- (g) Si $X_n \longrightarrow a$ (en probabilidad), cuando $n \rightarrow \infty$, e $Y_n \longrightarrow b$ (en probabilidad), cuando $n \rightarrow \infty$, con $a, b \in \mathbb{R}$, entonces $X_n Y_n \longrightarrow ab$ (en probabilidad), cuando $n \rightarrow \infty$.
- (h) Si $X_n \longrightarrow 1$ (en probabilidad), cuando $n \rightarrow \infty$, entonces $1/X_n \longrightarrow 1$ (en probabilidad), cuando $n \rightarrow \infty$.
- (i) Si $X_n \longrightarrow a$ (en probabilidad), cuando $n \rightarrow \infty$, e $Y_n \longrightarrow b$ (en probabilidad), cuando $n \rightarrow \infty$, siendo $a, b \in \mathbb{R}$ y $b \neq 0$, entonces $X_n/Y_n \longrightarrow a/b$ (en probabilidad), cuando $n \rightarrow \infty$.
- (j) Si $X_n \longrightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, e Y es una variable aleatoria, entonces $X_n Y \longrightarrow XY$ (en probabilidad), cuando $n \rightarrow \infty$.
- (k) Si $g : \mathbb{R} \longrightarrow \mathbb{R}$ es una función continua y $X_n \longrightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, entonces $g(X_n) \longrightarrow g(X)$ (en probabilidad), cuando $n \rightarrow \infty$.
- (l) Si $X_n \longrightarrow c$ (en probabilidad), cuando $n \rightarrow \infty$, siendo $c \in \mathbb{R}$ una constante, entonces $g(X_n) \longrightarrow g(c)$ (en probabilidad), cuando $n \rightarrow \infty$, para cualquier función g continua.

Demostración.— (Sólo probamos algunos de los apartados.)

(b) Sea $c > 0$. Entonces,

$$\begin{aligned} P(|X - Y| > c) &= P(|(X - X_n) + (X_n - Y)| > c) \\ &\leq P\left(|X_n - X| > \frac{c}{2}\right) + P\left(|X_n - Y| > \frac{c}{2}\right), \end{aligned}$$

de donde, tomando $n \rightarrow \infty$, se sigue que $P(|X - Y| > c) = 0$, $\forall c > 0$.

(k) Sea X una variable aleatoria. Entonces, para cada $\varepsilon > 0$, es posible encontrar una constante $k = k(\varepsilon)$ tal que

$$P(|X| > k) < \frac{\varepsilon}{2}.$$

Como g es continua, se tiene que g es uniformemente continua sobre $[-k, k]$, $\forall k \in \mathbb{R}^+$. Entonces, $\exists \delta = \delta(\varepsilon, k) > 0$ tal que $|g(x_n) - g(x)| < \varepsilon$, $\forall x \in [-k, k]$ y $|x_n - x| < \delta$.

Sean los conjuntos

$$\begin{aligned} A &= \{|X| \leq k\}, \\ B &= \{|X_n - X| < \delta\}, \\ C &= \{|g(X_n) - g(X)| < \varepsilon\}. \end{aligned}$$

Entonces, $\omega \in A \cap B$ implica $\omega \in C$, de modo que $A \cap B \subset C$ y

$$P(C^c) \leq P((A \cap B)^c) = P(A^c \cup B^c) \leq P(A^c) + P(B^c),$$

puesto que $C^c \subset (A \cap B)^c$; es decir,

$$P(|g(X_n) - g(X)| \geq \varepsilon) \leq P(|X| > k) + P(|X_n - X| \geq \delta) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2},$$

$\forall n \geq N$, donde $N = N(\varepsilon, \delta, k)$ es elegido de forma que $P(|X_n - X| \geq \delta) < \varepsilon/2$, $\forall n \geq N$.

(l) Se demuestra desde (g) y (k). \square

Observación 5.1 El ejemplo 5.3 sirve para mostrar que $X_n \rightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, pero no se tiene la convergencia $E[X_n^k] \rightarrow E[X^k]$ cuando $n \rightarrow \infty$, para $k \in \mathbb{N}$.

5.2.3 Convergencia en media de orden p

Definición 5.3 Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias tales que $E[|X_n|^p] < \infty$, $\forall n \geq 1$, para algún $p > 0$. Se dice que $\{X_n : n \geq 1\}$ converge en media de orden p hacia una variable aleatoria X , y se denota por $X_n \rightarrow X$ (m^p), si y sólo si $E[|X|^p] < \infty$ y

$$\lim_{n \rightarrow \infty} E[|X_n - X|^p] = 0.$$

Ejemplo 5.6 Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias definidas desde las funciones de masa

$$P(X_n = 0) = 1 - \frac{1}{n}, \quad P(X_n = 1) = \frac{1}{n}.$$

Entonces, $E[|X_n|^2] = 1/n \rightarrow 0$ cuando $n \rightarrow \infty$, lo cual implica que $X_n \rightarrow X$ (m^2), cuando $n \rightarrow \infty$, siendo X una variable aleatoria degenerada en $x = 0$.

Las desigualdades de Hölder y de Minkowski recogidas en los teoremas 5.5 y 5.6, respectivamente, son de interés en el establecimiento de la convergencia

en media de orden p de una sucesión de variables aleatorias. Estos resultados están basados en la desigualdad

$$|xy| \leq \frac{|x|^p}{p} + \frac{|y|^q}{q}, \quad (8)$$

donde $p, q \in \mathbb{R}^+$ son tales que $p > 1$ y $1/p + 1/q = 1$. Para demostrar (8), observamos que la función $h(x) = \ln x$ es cóncava para $x > 0$ y, como consecuencia, $\forall x_1, x_2 > 0$

$$\ln(tx_1 + (1-t)x_2) \geq t \ln x_1 + (1-t) \ln x_2.$$

Tomando exponenciales sobre ambos miembros de esta desigualdad,

$$tx_1 + (1-t)x_2 \geq x_1^t x_2^{1-t}.$$

Tomando $x_1 = |x|^p$, $x_2 = |y|^q$ y $t = 1/p$ (y, por tanto, $1-t = 1/q$), donde $p > 1$ y $1/p + 1/q = 1$, se obtiene

$$x_1^t x_2^{1-t} = (|x|^p)^{1/p} (|y|^q)^{1/q} \leq \frac{1}{p} |x|^p + \frac{1}{q} |y|^q.$$

Teorema 5.5 (Desigualdad de Hölder) Sean $p > 1$ y $q > 1$ tales que $1/p + 1/q = 1$. Entonces,

$$E[|XY|] \leq (E[|X|^p])^{1/p} (E[|Y|^q])^{1/q}. \quad (9)$$

Demostración.— En la desigualdad (8), tomamos

$$x = \frac{X(\omega)}{(E[|X|^p])^{1/p}} \quad \text{e} \quad y = \frac{Y(\omega)}{(E[|Y|^q])^{1/q}},$$

para cada $\omega \in \Omega$. Se obtiene entonces

$$\left| \frac{X(\omega)Y(\omega)}{(E[|X|^p])^{1/p} (E[|Y|^q])^{1/q}} \right| \leq \frac{|X(\omega)|^p}{p E[|X|^p]} + \frac{|Y(\omega)|^q}{q E[|Y|^q]}.$$

La relación (9) se sigue tomando esperanzas sobre esta desigualdad. \square

Para $p = 2$, la desigualdad (9) de Hölder viene dada por

$$E[|XY|] \leq \sqrt{E[X^2]E[Y^2]}$$

y es conocida como desigualdad de Cauchy-Schwartz.

Corolario 5.1 Si $X_n \longrightarrow X$ (m^2), cuando $n \rightarrow \infty$, entonces $\lim_{n \rightarrow \infty} E[X_n] = E[X]$ y $\lim_{n \rightarrow \infty} E[X_n^2] = E[X^2]$.

Demostración.— Desde la desigualdad de Cauchy-Schwartz, se tienen

$$\begin{aligned} |E[X_n - X]| &\leq E[|X_n - X|] = E[|1 \cdot (X_n - X)|] \leq \sqrt{E[|X_n - X|^2]}, \\ E[X_n^2] &\leq E[(X_n - X)^2] + E[X^2] + 2\sqrt{E[X^2]E[|X_n - X|^2]}, \end{aligned}$$

puesto que $E[X_n^2] = E[(X_n - X) + X]^2 = E[(X_n - X)^2] + E[X^2] + 2E[X(X_n - X)]$.

Entonces, el hecho de que $\lim_{n \rightarrow \infty} E[|X_n - X|^2] = 0$ implica el enunciado. \square

Corolario 5.2 Si $X_n \rightarrow X$ (m^2), cuando $n \rightarrow \infty$, e $Y_n \rightarrow Y$ (m^2), cuando $n \rightarrow \infty$, entonces $\lim_{n, m \rightarrow \infty} E[X_n Y_m] = E[XY]$.

Como consecuencia del corolario 5.1, la convergencia $X_n \rightarrow X$ (m^2), cuando $n \rightarrow \infty$, implica $\lim_{n \rightarrow \infty} \text{Var}(X_n) = \text{Var}(X)$. Similarmente, desde los corolarios 5.1 y 5.2, si $X_n \rightarrow X$ (m^2), cuando $n \rightarrow \infty$, e $Y_n \rightarrow Y$, (m^2) cuando $n \rightarrow \infty$, se concluye $\lim_{n, m \rightarrow \infty} \text{Cov}(X_n, Y_m) = \text{Cov}(X, Y)$.

Teorema 5.6 (Desigualdad de Minkowski) Para $p \geq 1$,

$$(E[|X + Y|^p])^{1/p} \leq (E[|X|^p])^{1/p} + (E[|Y|^p])^{1/p}. \quad (10)$$

Demostración. Para $p = 1$, la desigualdad (10) es inmediata. Cuando $p > 1$, se tiene

$$\begin{aligned} |X(\omega) + Y(\omega)|^p &= |X(\omega) + Y(\omega)| \cdot |X(\omega) + Y(\omega)|^{p-1} \\ &\leq |X(\omega)| \cdot |X(\omega) + Y(\omega)|^{p-1} + |Y(\omega)| \cdot |X(\omega) + Y(\omega)|^{p-1}. \end{aligned}$$

Tomando esperanzas,

$$E[|X + Y|^p] \leq E[|X| \cdot |X + Y|^{p-1}] + E[|Y| \cdot |X + Y|^{p-1}],$$

y aplicando (9), con $Y' = |X + Y|^{p-1}$, se tiene

$$E[|X + Y|^p] \leq \left((E[|X|^p])^{1/p} + (E[|Y|^p])^{1/p} \right) \left(E[|X + Y|^{(p-1)q}] \right)^{1/q}.$$

Excluyendo el caso trivial $E[|X + Y|^p] = 0$ y notando que $(p-1)q = p$, se tiene

$$E[|X + Y|^p] \leq \left((E[|X|^p])^{1/p} + (E[|Y|^p])^{1/p} \right) (E[|X + Y|^p])^{1/q},$$

desde donde se sigue la desigualdad (10). \square

Teorema 5.7 Si $X_n \rightarrow X$ (m^p), cuando $n \rightarrow \infty$, entonces $\lim_{n \rightarrow \infty} E[|X_n|^p] = E[|X|^p]$.

Demostración. Sea $p \in (0, 1]$. Entonces,

$$\begin{aligned} E[|X_n|^p] - E[|X|^p] &\leq E[|X_n - X|^p], \\ E[|X|^p] - E[|X_n|^p] &\leq E[|X_n - X|^p], \end{aligned}$$

de modo que $|E[|X_n|^p] - E[|X|^p]| \leq E[|X_n - X|^p]$.

Si $p > 1$, entonces la desigualdad de Minkowski permite observar que

$$\begin{aligned} (E[|X_n|^p])^{1/p} &\leq (E[|X_n - X|^p])^{1/p} + (E[|X|^p])^{1/p}, \\ (E[|X|^p])^{1/p} &\leq (E[|X_n - X|^p])^{1/p} + (E[|X_n|^p])^{1/p}, \end{aligned}$$

implicando $|(E[|X_n|^p])^{1/p} - (E[|X|^p])^{1/p}| \leq (E[|X_n - X|^p])^{1/p}$.

En ambos casos, el hecho de que $\lim_{n \rightarrow \infty} E[|X_n - X|^p] = 0$ implica el enunciado. \square

Como combinación de los teoremas 3.16 y 5.7 se obtiene que la condición $X_n \rightarrow X$ (m^p), cuando $n \rightarrow \infty$, implica $X_n \rightarrow X$ (m^q), cuando $n \rightarrow \infty$, y $\lim_{n \rightarrow \infty} E[|X_n|^q] = E[|X|^q]$ en el caso $q \leq p$.

5.2.4 Convergencia casi segura

Definición 5.4 Se dice que una sucesión $\{X_n : n \geq 1\}$ de variables aleatorias converge casi seguro o con probabilidad 1 hacia una variable aleatoria X cuando $n \rightarrow \infty$, y se denota por $X_n \rightarrow X$ (c.s.), si y sólo si

$$P\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\} = 1.$$

La verificación de la convergencia casi segura puede ser realizada desde el siguiente resultado.

Teorema 5.8 Sean $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias y X una variable aleatoria definidas sobre el espacio de probabilidad (Ω, \mathcal{A}, P) . Entonces, $X_n \rightarrow X$ (c.s.), cuando $n \rightarrow \infty$, si y sólo si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P\left(\bigcup_{m=n}^{\infty} \{\omega \in \Omega : |X_m(\omega) - X(\omega)| \geq \varepsilon\}\right) = 0.$$

El teorema 5.9 pone de manifiesto la diferencia entre convergencia casi segura y convergencia en probabilidad.

Teorema 5.9 (i) $X_n \rightarrow X$ (c.s.), cuando $n \rightarrow \infty$, si y sólo si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P\left\{\sup_{m \geq n} |X_m - X| > \varepsilon\right\} = 0.$$

(ii) Si $\forall \varepsilon > 0$ la serie $\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon)$ converge, entonces $X_n \rightarrow X$ (c.s.) cuando $n \rightarrow \infty$.

5.3 Relaciones entre los tipos de convergencia. Propiedades elementales

Nuestro siguiente objetivo es establecer relaciones entre los tipos de convergencia introducidos en la sección 5.2 y mostrar que las relaciones inversas no son necesariamente ciertas mediante contraejemplos.

Teorema 5.10 Si $X_n \longrightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, entonces $X_n \longrightarrow X$ (en ley), cuando $n \rightarrow \infty$.

Demostración.— Sean F_n y F las funciones de distribución de X_n y X , respectivamente. Entonces, desde la relación

$$\begin{aligned} \{X \leq x'\} &= \{X_n \leq x, X \leq x'\} \cup \{X_n > x, X \leq x'\} \\ &\subset \{X_n \leq x\} \cup \{X_n > x, X \leq x'\}, \end{aligned}$$

se observa que

$$F(x') \leq F_n(x) + P(X_n > x, X \leq x').$$

Dado que $X_n \longrightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, se tiene $X_n - X \longrightarrow 0$ (en probabilidad), cuando $n \rightarrow \infty$, desde el teorema 5.4.(a). Entonces, para $x' < x$,

$$P(X_n > x, X \leq x') \leq P(|X_n - X| > x - x') \longrightarrow 0, \text{ cuando } n \rightarrow \infty,$$

puesto que $X \leq x' < x < X_n$ implica $X_n - X > x - x'$. En definitiva, encontramos que

$$F(x') \leq \liminf_n F_n(x), \quad \forall x' < x.$$

Similarmente, se obtiene

$$\limsup_n F_n(x) \leq F(x''), \quad \forall x < x''.$$

Desde las dos desigualdades anteriores,

$$F(x') \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x''), \quad \text{si } x' < x < x''.$$

Dado que F tiene, a lo sumo, una cantidad numerable de puntos de discontinuidad, es posible elegir el punto x como un punto de continuidad de F . Entonces, se deduce $F(x) = \lim_{n \rightarrow \infty} F_n(x)$ tomando $x'' \downarrow x$ y $x' \uparrow x$. \square

Corolario 5.3 Sea $k \in \mathbb{R}$ una constante. Entonces, $X_n \longrightarrow k$ (en ley), cuando $n \rightarrow \infty$, si y sólo si $X_n \longrightarrow k$ (en probabilidad), cuando $n \rightarrow \infty$.

El resultado recíproco al teorema 5.10 no necesariamente es válido en el caso en que la constante k (en el corolario 5.3) sea reemplazada por una variable aleatoria X no-degenerada.

Ejemplo 5.7 Sean X_1, X_2, \dots variables aleatorias idénticamente distribuidas con distribución conjunta de (X_n, X) dada por

$$P(X = 0, X_n = 1) = P(X = 1, X_n = 0) = \frac{1}{2}.$$

Entonces, es sencillo comprobar que $X_n \rightarrow X$ (en ley), cuando $n \rightarrow \infty$. No obstante,

$$\begin{aligned} P\left(|X_n - X| > \frac{1}{2}\right) &= P(|X_n - X| = 1) \\ &= P(X_n = 0, X = 1) + P(X_n = 1, X = 0) \\ &= 1. \end{aligned}$$

Es decir, $X_n \rightarrow X$ (en ley), cuando $n \rightarrow \infty$, pero no es cierto que $X_n \rightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$.

Sin embargo, la convergencia en media de orden $p > 0$ sí implica la convergencia en probabilidad.

Teorema 5.11 Si $X_n \rightarrow X$ (m^p), cuando $n \rightarrow \infty$, con $p > 0$, entonces $X_n \rightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$.

Ejemplo 5.8 Para mostrar que el resultado recíproco al teorema 5.11 no se tiene necesariamente, supongamos que $\{X_n : n \geq 1\}$ es una sucesión de variables aleatorias definidas desde

$$P(X_n = 0) = 1 - \frac{1}{n^p}, \quad P(X_n = n) = \frac{1}{n^p}, \quad n \geq 1,$$

siendo $p > 0$. Entonces, $E[|X_n|^p] = 1$ y, como resultado, no es cierto que $X_n \rightarrow 0$ (m^p), cuando $n \rightarrow \infty$. No obstante, $X_n \rightarrow 0$ (en probabilidad), cuando $n \rightarrow \infty$, puesto que

$$P(|X_n| > \varepsilon) = \begin{cases} P(X_n = n) = \frac{1}{n^p}, & \text{si } \varepsilon < n, \\ 0, & \text{si } n \leq \varepsilon, \end{cases}$$

implica $\lim_{n \rightarrow \infty} P(|X_n| > \varepsilon) = 0$.

A continuación establecemos la relación entre la convergencia casi segura y la convergencia en probabilidad. Para ello, comenzamos observando que si $X_n \rightarrow 0$ (c.s.), entonces

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P\left(\sup_{m \geq n} |X_m| > \varepsilon\right) = 0.$$

Como consecuencia, tomados $\varepsilon > 0$ y $\eta > 0$, podemos encontrar $n_0 \in \mathbb{N}$ tal que

$$P\left(\sup_{n \geq n_0} |X_n| > \varepsilon\right) < \eta.$$

De forma equivalente, es posible escribir

$$\lim_{n_0 \rightarrow \infty} P \left(\bigcup_{n=n_0}^{\infty} \{|X_n| > \varepsilon\} \right) = 0.$$

Teorema 5.12 Si $X_n \rightarrow X$ (c.s.), cuando $n \rightarrow \infty$, entonces $X_n \rightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$.

Demostración.— De acuerdo con el argumento previo, tenemos que $X_n \rightarrow X$ (c.s.), cuando $n \rightarrow \infty$, implica que, para valores $\varepsilon > 0$ y $\eta > 0$ arbitrarios, es posible elegir $n_0 = n_0(\varepsilon, \eta)$ tal que

$$P \left(\bigcap_{n=n_0}^{\infty} \{|X_n - X| \leq \varepsilon\} \right) = 1 - P \left(\bigcup_{n=n_0}^{\infty} \{|X_n - X| > \varepsilon\} \right) \geq 1 - \eta.$$

Dado que

$$\bigcap_{n=n_0}^{\infty} \{|X_n - X| \leq \varepsilon\} \subset \{|X_n - X| \leq \varepsilon\}, \quad \forall n \geq n_0,$$

se tiene

$$P(|X_n - X| \leq \varepsilon) \geq P \left(\bigcap_{n=n_0}^{\infty} \{|X_n - X| \leq \varepsilon\} \right) \geq 1 - \eta,$$

es decir, $P(|X_n - X| > \varepsilon) < \eta$, $\forall n \geq n_0$, y por lo tanto se observa que $X_n \rightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$. \square

El ejemplo 5.9 muestra que el resultado recíproco al teorema 5.12 no es necesariamente cierto.

Ejemplo 5.9 Para cada $n \in \mathbb{N}$ existen enteros m y k (únicos) tales que

$$n = 2^k + m, \quad 0 \leq m < 2^k, \quad k \in \{0, 1, \dots\}.$$

Para cada $n \in \mathbb{N}$, definimos la variable aleatoria X_n sobre $\Omega = [0, 1]$ como

$$X_n(\omega) = \begin{cases} 2^k, & \text{si } \frac{m}{2^k} \leq \omega < \frac{m+1}{2^k}, \\ 0, & \text{en caso contrario.} \end{cases}$$

Sea la distribución de X_n aquella determinada desde la asignación $P(I) = \text{"longitud de } I"$ si $I \subset \Omega$ es un intervalo. Entonces,

$$P(X_n = 2^k) = \frac{1}{2^k} \quad \text{y} \quad P(X_n = 0) = 1 - \frac{1}{2^k}.$$

Es claro que no existe el valor $\lim_{n \rightarrow \infty} X_n(\omega)$ para $\omega \in \Omega$. Por tanto, X_n no converge casi seguro. No obstante,

$$P(|X_n| > \varepsilon) = P(X_n > \varepsilon) = \begin{cases} 0, & \text{si } 2^k \leq \varepsilon, \\ \frac{1}{2^k}, & \text{si } 0 < \varepsilon < 2^k, \end{cases}$$

para cada k fijo y, como resultado, $\lim_{n \rightarrow \infty} P(|X_n| > \varepsilon) = 0$.

Teorema 5.13 Sea $\{X_n : n \geq 1\}$ una sucesión (estríctamente) decreciente de variables aleatorias positivas tal que $X_n \rightarrow 0$ (en probabilidad), cuando $n \rightarrow \infty$. Entonces, $X_n \rightarrow 0$ (c.s.), cuando $n \rightarrow \infty$.

Observación 5.2 También es posible desmotrar que, cuando $X_n \rightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, existe una subsucesión $\{X_{n_k} : k \geq 1\}$ de la sucesión $\{X_n : n \geq 1\}$ que satisface $X_{n_k} \rightarrow X$ (c.s.), cuando $k \rightarrow \infty$.

Teorema 5.14 Sean $\{X_n : n \geq 1\}$ e $\{Y_n : n \geq 1\}$ sucesiones de variables aleatorias tales que $|X_n - Y_n| \rightarrow 0$ (en probabilidad), cuando $n \rightarrow \infty$, e $Y_n \rightarrow Y$ (en ley), cuando $n \rightarrow \infty$. Entonces, $X_n \rightarrow Y$ (en ley), cuando $n \rightarrow \infty$.

Demostración.— Sean $\varepsilon > 0$ y x un punto de continuidad de la función de distribución de Y . Entonces,

$$\begin{aligned} P(X_n \leq x) &= P(Y_n \leq x + Y_n - X_n) \\ &= P(Y_n \leq x + Y_n - X_n, Y_n - X_n \leq \varepsilon) \\ &\quad + P(Y_n \leq x + Y_n - X_n, Y_n - X_n > \varepsilon) \\ &\leq P(Y_n \leq x + \varepsilon) + P(Y_n - X_n > \varepsilon). \end{aligned}$$

Entonces,

$$\limsup_n P(X_n \leq x) \leq \limsup_n P(Y_n \leq x + \varepsilon).$$

Análogamente, se tiene

$$\liminf_n P(Y_n \leq x - \varepsilon) \leq \liminf_n P(X_n \leq x).$$

Dado que $\varepsilon > 0$ es arbitrario y x es un punto de continuidad de $P(Y \leq x)$ se observa que $\lim_{n \rightarrow \infty} P(X_n \leq x) = P(Y \leq x)$. \square

El corolario 5.4 es obtenido cuando se toman Y_n , X e Y idénticamente distribuidas en el teorema 5.14.

Corolario 5.4 Si $X_n \rightarrow X$ (en probabilidad), cuando $n \rightarrow \infty$, entonces $X_n \rightarrow X$ (en ley), cuando $n \rightarrow \infty$.

Para finalizar la sección, resumimos en el teorema 5.15 algunos resultados de interés.

Teorema 5.15 Sean $\{X_n : n \geq 1\}$ e $\{Y_n : n \geq 1\}$ sucesiones de variables aleatorias y $c \in \mathbb{R}$ una constante. Entonces,

- (a) Si $X_n \rightarrow X$ (en ley), cuando $n \rightarrow \infty$, e $Y_n \rightarrow c$ (en probabilidad), cuando $n \rightarrow \infty$, entonces $X_n \pm Y_n \rightarrow X \pm c$ (en ley), cuando $n \rightarrow \infty$.
- (b) Si $X_n \rightarrow X$ (en ley), cuando $n \rightarrow \infty$, e $Y_n \rightarrow c$ (en probabilidad), cuando $n \rightarrow \infty$, entonces $X_n Y_n \rightarrow cX$ (en ley), cuando $n \rightarrow \infty$, si $c \neq 0$, y $X_n Y_n \rightarrow 0$ (en probabilidad), cuando $n \rightarrow \infty$, si $c = 0$.
- (c) Si $X_n \rightarrow X$ (en ley), cuando $n \rightarrow \infty$, e $Y_n \rightarrow c$ (en probabilidad), cuando $n \rightarrow \infty$, entonces $X_n/Y_n \rightarrow X/c$ (en ley), cuando $n \rightarrow \infty$, si $c \neq 0$.

5.4 Ley débil de los grandes números

Sean $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias y $S_n = \sum_{j=1}^n X_j$, para $n \geq 1$. En esta sección nos preguntamos por la existencia de constantes $A_n \in \mathbb{R}$ y $B_n > 0$, con $B_n \rightarrow \infty$, tales que la sucesión

$$\left\{ \frac{S_n - A_n}{B_n} : n \geq 1 \right\}$$

converja hacia 0 en probabilidad, cuando $n \rightarrow \infty$. En el caso de existir, A_n y B_n son denominadas constantes de centralización y de normalización, respectivamente.

Definición 5.5 Se dice que una sucesión $\{X_n : n \geq 1\}$ de variables aleatorias obedece a la ley débil de los grandes números con respecto a la sucesión de constantes $\{B_n : n \geq 1\}$, con $B_n > 0$ y $B_n \rightarrow \infty$, si existe una sucesión $\{A_n : n \geq 1\}$ de constantes reales tal que

$$\frac{S_n - A_n}{B_n} \rightarrow 0 \text{ (en probabilidad), cuando } n \rightarrow \infty.$$

Teorema 5.16 Sean $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes dos-a-dos con $E[X_j] = \mu_j$ y $\text{Var}(X_j) = \sigma_j^2$, tales que $\sum_{j=1}^n \sigma_j^2 \rightarrow \infty$ cuando $n \rightarrow \infty$. Entonces,

$$\frac{S_n - \sum_{j=1}^n \mu_j}{\sum_{j=1}^n \sigma_j^2} \rightarrow 0 \text{ (en probabilidad), cuando } n \rightarrow \infty.$$

Demostración.—Aplicando el teorema 3.20 deducimos

$$P \left(\left| S_n - \sum_{j=1}^n \mu_j \right| > \varepsilon \sum_{j=1}^n \sigma_j^2 \right) \leq \frac{E \left[\left(S_n - \sum_{j=1}^n \mu_j \right)^2 \right]}{\varepsilon^2 \left(\sum_{j=1}^n \sigma_j^2 \right)^2} = \frac{1}{\varepsilon^2 \sum_{j=1}^n \sigma_j^2}.$$

El resultado se sigue desde $\lim_{n \rightarrow \infty} P\left(\left|S_n - \sum_{j=1}^n \mu_j\right| > \varepsilon \sum_{j=1}^n \sigma_j^2\right) = 0$, puesto que $\sum_{j=1}^n \sigma_j^2 \rightarrow \infty$, $\forall \varepsilon > 0$. \square

Observemos que, en el teorema 5.16, la condición $\lim_{n \rightarrow \infty} \sum_{j=1}^n \sigma_j^2 = \infty$ puede ser reemplazada por $\lim_{n \rightarrow \infty} n^{-2} \sum_{j=1}^n \sigma_j^2 = 0$.

Algunas aplicaciones sencillas del teorema 5.16 aparecen recogidas en el corolario 5.5.

Corolario 5.5

- (a) Si $\{X_n : n \geq 1\}$ es una sucesión de variables aleatorias idénticamente distribuidas con $E[X_j] = \mu$ y $\text{Var}(X_j) = \sigma^2 < \infty$, e incorreladas dos-a-dos, entonces

$$\frac{S_n - n\mu}{n\sigma^2} \longrightarrow 0 \text{ (en probabilidad), cuando } n \rightarrow \infty.$$

- (b) Si $\{X_n : n \geq 1\}$ es una sucesión de variables aleatorias independientes dos-a-dos con $E[X_j] = \mu_j$ y $\text{Var}(X_j) \leq \gamma < \infty$, $\forall j \geq 1$, entonces

$$\frac{S_n - \sum_{j=1}^n \mu_j}{n} \longrightarrow 0 \text{ (en probabilidad), cuando } n \rightarrow \infty.$$

- (c) Si, en las condiciones del apartado (b), las variables aleatorias son idénticamente distribuidas, entonces

$$\frac{S_n - n\mu}{n} \longrightarrow 0 \text{ (en probabilidad), cuando } n \rightarrow \infty.$$

Como consecuencia, si $\{X_n : n \geq 1\}$ es una sucesión de variables aleatorias idénticamente distribuidas, incorreladas dos-a-dos, con varianza finita, entonces

$$\frac{1}{n} \sum_{j=1}^n X_j \longrightarrow \mu \text{ (en probabilidad), cuando } n \rightarrow \infty,$$

lo cual justifica la definición frecuentista de probabilidad (véase la sección 2.3) como se muestra en el ejemplo 5.10.

Ejemplo 5.10 (Teorema de Bernoulli) Sean X_1, X_2, \dots variables aleatorias independientes e idénticamente distribuidas Bernoulli (p). Entonces, $E[X_j] = p$ y $\text{Var}(X_j) = p(1-p)$. Por lo tanto, el corolario 5.5 implica

$$\frac{1}{n} \sum_{j=1}^n X_j \longrightarrow p \text{ (en probabilidad), cuando } n \rightarrow \infty.$$

A continuación proponemos una condición necesaria y suficiente para la verificación de la ley débil de los grandes números cuando $B_n = n$.

Teorema 5.17 Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias. Denotemos $Y_n = n^{-1}S_n$. Entonces, una condición necesaria y suficiente para que $\{X_n : n \geq 1\}$ obedezca a la ley débil de los grandes números, con respecto a las constantes de normalización $B_n = n$, es

$$\lim_{n \rightarrow \infty} E \left[\frac{Y_n^2}{1 + Y_n^2} \right] = 0.$$

Este resultado se aplica frecuentemente a variables aleatorias que son sumas de n variables aleatorias, resultando ser de uso limitado bajo hipótesis diferentes. No obstante, desde esta condición necesaria y suficiente, el teorema 5.16 es deducido como un simple corolario.

Ejemplo 5.11 Sea (X_1, \dots, X_n) un vector aleatorio Normal_n con vector de medias nulo y matriz de varianzas-covarianzas $\Sigma = (\sigma_{ij})$, donde $\sigma_{ii} = E[X_i^2] = 1$ y $\sigma_{ij} = \text{Cov}(X_i, X_j) = \rho$ si $|j - i| = 1$, y 0 en caso contrario. Entonces, $S_n = \sum_{j=1}^n X_j$ es Normal_1 de media 0 y varianza σ^2 , donde $\sigma^2 = \text{Var}(S_n) = n + 2(n-1)\rho$. Entonces,

$$\begin{aligned} E \left[\frac{Y_n^2}{1 + Y_n^2} \right] &= E \left[\frac{S_n^2}{n^2 + S_n^2} \right] \\ &= \frac{2}{\sigma\sqrt{2\pi}} \int_0^\infty \frac{x^2}{n^2 + x^2} \exp \left\{ -\frac{x^2}{2\sigma^2} \right\} dx \\ &= \frac{2}{\sigma\sqrt{2\pi}} \int_0^\infty \frac{(n + 2(n-1)\rho)y^2}{n^2 + (n + 2(n-1)\rho)y^2} \exp \left\{ -\frac{y^2}{2} \right\} dy, \end{aligned}$$

donde la última igualdad se obtiene desde el cambio de variable $x = \sqrt{n + 2(n-1)\rho}y$. Notemos que

$$\begin{aligned} \frac{2}{\sigma\sqrt{2\pi}} \int_0^\infty \frac{(n + 2(n-1)\rho)y^2}{n^2 + (n + 2(n-1)\rho)y^2} \exp \left\{ -\frac{y^2}{2} \right\} dy \\ \leq \frac{n + 2(n-1)\rho}{n^2} \int_0^\infty \frac{2}{\sqrt{2\pi}} y^2 e^{-y^2/2} dy \longrightarrow 0, \end{aligned}$$

cuando $n \rightarrow \infty$. Por tanto, $\frac{1}{n} \sum_{j=1}^n X_j \longrightarrow 0$ (en probabilidad), cuando $n \rightarrow \infty$.

Para concluir se incluye un resultado debido a Khintchine.

Teorema 5.18 (Ley débil de Khintchine) Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con media $E[X_j] = \mu$ finita. Entonces,

$$\frac{1}{n} \sum_{j=1}^n X_j \longrightarrow \mu \quad (\text{en probabilidad}), \text{ cuando } n \rightarrow \infty.$$

Ejemplo 5.12 Sean X_1, X_2, \dots variables aleatorias independientes e idénticamente distribuidas con $E[|X_j|^k] < \infty$, para $k \in \mathbb{N}$. Entonces, desde el teorema 5.18,

$$\frac{1}{n} \sum_{j=1}^n X_j^k \longrightarrow E[X_1^k] \quad (\text{en probabilidad}), \text{ cuando } n \rightarrow \infty.$$

Además, la condición $E[X_1^2] < \infty$ implica $E[X_1] < \infty$, de modo que

$$\left(\frac{1}{n} \sum_{j=1}^n X_j \right)^2 \longrightarrow \mu^2 \quad (\text{en probabilidad}), \text{ cuando } n \rightarrow \infty,$$

$$\frac{1}{n} \sum_{j=1}^n X_j^2 \longrightarrow E[X_1^2] \quad (\text{en probabilidad}), \text{ cuando } n \rightarrow \infty.$$

Es decir,

$$\frac{1}{n} \sum_{j=1}^n X_j^2 - \left(\frac{1}{n} \sum_{j=1}^n X_j \right)^2 \longrightarrow \text{Var}(X_1) \quad (\text{en probabilidad}),$$

cuando $n \rightarrow \infty$.

5.5 Ley fuerte de los grandes números

Una versión más fuerte, según el teorema 5.12, de los resultados dados en la sección 5.4 está basada en la convergencia casi segura.

Definición 5.6 Se dice que una sucesión $\{X_n : n \geq 1\}$ de variables aleatorias obedece a la ley fuerte de los grandes números con respecto a la sucesión de constantes $\{B_n : n \geq 1\}$, con $B_n > 0$ y $B_n \rightarrow \infty$, si existe una sucesión $\{A_n : n \geq 1\}$ de constantes reales tal que

$$\frac{S_n - A_n}{B_n} \longrightarrow 0 \quad (c.s.), \quad \text{cuando } n \rightarrow \infty.$$

Nos interesamos en el caso $B_n = n$, $n \geq 1$.

Comenzamos estudiando el lema de Borel-Cantelli.

Teorema 5.19 (Lema de Borel-Cantelli) Sean $\{A_n : n \geq 1\} \subset \mathcal{A}$ una sucesión de subconjuntos y $A = \limsup_n A_n$.

- (a) Si $\sum_{n=1}^{\infty} P(A_n) < \infty$, entonces $P(A) = 0$.
- (b) Si $\{A_n : n \geq 1\} \subset \mathcal{A}$ es una sucesión independiente y $\sum_{n=1}^{\infty} P(A_n) = \infty$, entonces $P(A) = 1$.

Demostración.— Para probar (a) evaluamos

$$\begin{aligned} P(A) &= P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = P\left(\lim_{n \rightarrow \infty} \bigcup_{k=n}^{\infty} A_k\right) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(A_k) = 0. \end{aligned}$$

El apartado (b) se demuestra de forma análoga. \square

Corolario 5.6 Si $\{A_n : n \geq 1\} \subset \mathcal{A}$ es una sucesión de subconjuntos independientes, entonces $P(A) \in \{0, 1\}$.

Una sencilla aplicación del lema de Borel-Cantelli conduce a la siguiente ley fuerte de los grandes números:

Teorema 5.20 Si X_1, X_2, \dots son variables aleatorias independientes e idénticamente distribuidas con media μ y momento de orden 4 finito, entonces

$$\frac{S_n}{n} \longrightarrow \mu \quad (c.s.), \text{ cuando } n \rightarrow \infty.$$

En el teorema 5.20, es posible asumir diferentes hipótesis que aseguren que el momento de orden 4 es finito; por ejemplo, que las variables aleatorias estén acotadas casi seguro.

Corolario 5.7 Si X_1, X_2, \dots son variables aleatorias independientes e idénticamente distribuidas tales que $P(|X_n| < K) = 1, \forall n \geq 1$, siendo $0 < K < \infty$, entonces

$$\frac{S_n}{n} \longrightarrow \mu \quad (c.s.), \text{ cuando } n \rightarrow \infty.$$

El estudio de las leyes fuertes de los grandes números implica el uso de un buen número de resultados auxiliares relativos a las nociones de "sucesiones equivalentes", "sucesiones equivalentes en cola", la desigualdad de Kolmogorov, el lema de Kronecker, etc. Las demostraciones resultan ser laboriosas y superan los objetivos de la asignatura. Por este motivo, en el resto de la sección nos restringimos a enunciar los resultados y a comentar su posible aplicación evitando argumentos teóricos.

El criterio de Cauchy evita las hipótesis de independencia e idéntica distribución entre las variables aleatorias de la sucesión $\{X_n : n \geq 1\}$.

Teorema 5.21 (Criterio de Cauchy) $X_n \longrightarrow X \quad (c.s.),$ cuando $n \rightarrow \infty$, si y sólo si $\lim_{n \rightarrow \infty} P(\sup_m |X_{n+m} - X_n| \leq \varepsilon) = 1$.

La combinación del teorema 5.21 con la desigualdad de Kolmogorov permite deducir el siguiente corolario:

Corolario 5.8 Si $\{X_n : n \geq 1\}$ es una sucesión de variables aleatorias independientes verificando

$$\sum_{n=1}^{\infty} \frac{\text{Var}(X_n)}{B_n^2} < \infty,$$

con $B_n \uparrow \infty$, entonces

$$\frac{S_n - E[S_n]}{B_n} \longrightarrow 0 \quad (\text{c.s.}), \text{ cuando } n \rightarrow \infty.$$

Ejemplo 5.13 Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes con varianzas uniformemente acotadas; es decir, $\text{Var}(X_n) \leq K < \infty, \forall n \geq 1$. Entonces,

$$\sum_{n=1}^{\infty} \frac{\text{Var}(X_n)}{n^2} \leq K \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

lo cual implica, desde el corolario 5.8, el comportamiento

$$\frac{S_n - E[S_n]}{n} \longrightarrow 0 \quad (\text{c.s.}), \text{ cuando } n \rightarrow \infty.$$

Como aplicación del ejemplo 5.13 es posible demostrar la ley fuerte de Borel, que incide sobre la definición frecuentista de la noción de probabilidad.

Corolario 5.9 (Ley fuerte de Borel de los grandes números) Para una sucesión $\{X_n : n \geq 1\}$ de variables aleatorias independientes e idénticamente distribuidas Bernoulli (p), se verifica

$$\frac{S_n}{n} \longrightarrow p \quad (\text{c.s.}), \text{ cuando } n \rightarrow \infty.$$

Demostración.— Es una aplicación inmediata del ejemplo 5.13 observando que $E[X_n] = p$ y $\text{Var}(X_n) = p(1-p) \leq 1/4$ para todo $p \in (0, 1)$. \square

Si aludimos a resultados donde no se asumen condiciones sobre los momentos de orden superior a 1, entonces tenemos la ley fuerte de Kolmogorov.

Teorema 5.22 (Ley fuerte de Kolmogorov de los grandes números) Sean X_1, X_2, \dots variables aleatorias independientes e idénticamente distribuidas. Entonces,

$$\frac{S_n}{n} \longrightarrow \mu \quad (\text{c.s.}), \text{ cuando } n \rightarrow \infty,$$

con $\mu < \infty$ si y sólo si $E[|X_j|] < \infty$, en cuyo caso $E[X_j] = \mu$.

5.6 Teorema central del límite

Sean $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias definidas sobre (Ω, \mathcal{A}, P) y $S_n = \sum_{j=1}^n X_j$. En las secciones 5.4 y 5.5 hemos tratado la convergencia de la sucesión

$$\frac{S_n - A_n}{B_n},$$

cuando $n \rightarrow \infty$, hacia una variable aleatoria degenerada mediante las leyes débil y fuerte de los grandes números, que involucran convergencias en probabilidad y casi segura, respectivamente. En la presente sección analizamos la convergencia de $(S_n - A_n)/B_n$ hacia una variable aleatoria no-degenerada mediante una convergencia en ley.

Supongamos que, para una adecuada elección de las sucesiones de constantes $\{A_n : n \geq 1\}$ y $\{B_n > 0 : n \geq 1\}$, se observa que

$$\frac{S_n - A_n}{B_n} \longrightarrow Y \quad (\text{en ley}), \text{ cuando } n \rightarrow \infty,$$

siendo Y una variable aleatoria definida sobre (Ω, \mathcal{A}, P) . Podemos preguntarnos por las características generales de Y ; es decir, por el tipo de distribución de probabilidad de Y . La respuesta es sencilla: podemos obtener cualquier distribución *límite* dependiendo de las hipótesis probabilísticas hechas sobre las variables de $\{X_n : n \geq 1\}$. Por ejemplo, si X_1 es una variable aleatoria con función de distribución F y X_2, X_3, \dots son degeneradas en el punto $x = 0$, entonces

$$\frac{S_n - A_n}{B_n} \longrightarrow X_1 \quad (\text{en ley}), \text{ cuando } n \rightarrow \infty,$$

en el supuesto $A_n = 0$ y $B_n = 1$.

El siguiente resultado es fundamental en el cálculo de probabilidades y se restringe al caso de mayor utilidad práctica, donde la ley de probabilidad límite es Normal.

Teorema 5.23 (Central del límite de Lindeberg) *Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes no-degeneradas con funciones de distribución $\{F_n : n \geq 1\}$. Denotemos*

$$\mu_n = E[X_n], \quad \sigma_n^2 = \text{Var}(X_n) \quad \text{y} \quad s_n^2 = \sigma_1^2 + \dots + \sigma_n^2.$$

En el caso en que, para todo $n \geq 1$, F_n sea absolutamente continua con función de densidad f_n , asumamos la condición

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n \int_{\{|x - \mu_k| > \varepsilon s_n\}} (x - \mu_k)^2 f_k(x) dx = 0, \quad \forall \varepsilon > 0. \quad (11)$$

En el caso en que, para todo $n \geq 1$, X_n sea una variable aleatoria discreta con soporte $\{x_{nj} : j \geq 1\}$ y función de masa $\{p_{nj} = P(X_n = x_{nj}) : j \geq 1\}$, asumamos la condición

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n \sum_{|x_{kj} - \mu_k| > \varepsilon s_n} (x_{kj} - \mu_k)^2 p_{kj} = 0, \quad \forall \varepsilon > 0.$$

Entonces,

$$\frac{S_n - E[S_n]}{s_n} \longrightarrow Z \quad (\text{en ley}), \text{ cuando } n \rightarrow \infty,$$

donde Z es $\text{Normal}(0, 1)$.

Ejemplo 5.14 Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas no-degeneradas con función de distribución (común) F , siendo $\mu = E[X_n]$ y $\sigma^2 = \text{Var}(X_n) < \infty$. Entonces, si la función de distribución F tiene función de densidad f , entonces

$$\mu = \int_{-\infty}^{\infty} x f(x) dx \quad \text{y} \quad \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

Por lo tanto, la condición (11) es de la forma

$$\frac{1}{n\sigma^2} \sum_{k=1}^n \int_{\{|x-\mu| > \varepsilon \sigma \sqrt{n}\}} (x - \mu)^2 f(x) dx \longrightarrow 0,$$

cuando $n \rightarrow \infty$ ¹¹.

El ejemplo 5.14 es la prueba, en el caso absolutamente continuo, del siguiente corolario:

Corolario 5.10 Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con función de distribución (común) F , siendo $\mu = E[X_n]$ y $\sigma^2 = \text{Var}(X_n) < \infty$. Entonces,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \longrightarrow Z \quad (\text{en ley}), \text{ cuando } n \rightarrow \infty,$$

donde Z es $\text{Normal}(0, 1)$.

¹¹Notemos que

$$\int_{\{|x-\mu| > \varepsilon \sigma \sqrt{n}\}} (x - \mu)^2 f(x) dx = E[Z_n],$$

siendo $Z_n = (X - \mu)^2 I_{\{|X-\mu| > \varepsilon \sigma \sqrt{n}\}}$, donde X es una variable aleatoria con función de distribución F . Entonces, es claro que $Z_n \rightarrow 0$ (c.s.), cuando $n \rightarrow \infty$, y $|Z_n| \leq (X - \mu)^2$ con $E[(X - \mu)^2] = \sigma^2 < \infty$. Por el teorema de la convergencia dominada (teorema 5.25), se observa que $\lim_{n \rightarrow \infty} E[Z_n] = 0$ y, como resultado, se tiene

$$\lim_{n \rightarrow \infty} \frac{1}{n\sigma^2} \sum_{k=1}^n \int_{\{|x-\mu| > \varepsilon \sigma \sqrt{n}\}} (x - \mu)^2 f(x) dx = \lim_{n \rightarrow \infty} \frac{1}{n\sigma^2} \sum_{k=1}^n E[Z_n] = \lim_{n \rightarrow \infty} \frac{E[Z_n]}{\sigma^2} = 0.$$

5.7 Teorema de la convergencia monótona y teorema de la convergencia dominada

El presente tema finaliza con dos resultados — los teoremas de la convergencia monótona y de la convergencia dominada — que analizan la convergencia de expresiones integrales relativas a una sucesión de variables aleatorias.

Teorema 5.24 (De la convergencia monótona) Sea $\{X_n : n \geq 1\}$ una sucesión no-decreciente de variables aleatorias no-negativas. Si $X_n \rightarrow X$ (c.s.), cuando $n \rightarrow \infty$, entonces $\lim_{n \rightarrow \infty} E[X_n] = E[X]$.

Ejemplo 5.15 Sean X_1, X_2, \dots variables aleatorias no-negativas. Para cada $n \geq 1$, definimos la variable aleatoria no-negativa $S_n = \sum_{j=1}^n X_j$, que satisface $S_n \uparrow \sum_{j=1}^{\infty} X_j$. Aplicando el teorema 5.24 se tiene

$$E[\lim_{n \rightarrow \infty} S_n] = \lim_{n \rightarrow \infty} E[S_n],$$

es decir, $E[\sum_{j=1}^{\infty} X_j] = \sum_{j=1}^{\infty} E[X_j]$.

Teorema 5.25 (De la convergencia dominada) Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias tal que $|X_n| \leq Y$, $\forall n \geq 1$, donde Y es una variable aleatoria con media finita, y $X_n \rightarrow X$ (c.s.), cuando $n \rightarrow \infty$. Entonces, X tiene media finita y $\lim_{n \rightarrow \infty} E[X_n] = E[X]$.

Los teoremas 5.24 y 5.25 aseguran que, bajo condiciones apropiadas, el límite de las esperanzas de las variables aleatorias X_n es la esperanza de la variable aleatoria $\lim_{n \rightarrow \infty} X_n$. Es posible obtener resultados más generales sustituyendo la función $\lim_{n \rightarrow \infty}$ por las funciones \liminf o \limsup ; es decir, prescindiendo de la convergencia (c.s.) de la sucesión de variables aleatorias.

Teorema 5.26 (Lema de Fatou) Sean $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias y X una variable aleatoria definidas sobre (Ω, \mathcal{A}, P) .

(a) Si $X \leq X_n$, $\forall n \geq 1$, y $E[X] > -\infty$, entonces

$$E\left[\liminf_n X_n\right] \leq \liminf_n E[X_n].$$

(b) Si $X_n \leq X$, $\forall n \geq 1$, y $E[X] < \infty$, entonces

$$\limsup_n E[X_n] \leq E\left[\limsup_n X_n\right].$$

El lema de Fatou es importante para demostrar el teorema de la convergencia dominada.

Demostración (del teorema 5.25).— Dadas $\{X_n : n \geq 1\}$ e Y tales que $|X_n| \leq Y$, $\forall n \geq 1$, se tienen las desigualdades requeridas para aplicar el teorema 5.26.(a) ($-Y \leq X_n$, $\forall n \geq 1$, con $E[-Y]$ finita) y 5.26.(b) ($X_n \leq Y$, $\forall n \geq 1$, con $E[Y]$ finita). Como resultado,

$$E[\liminf_n X_n] \leq \liminf_n E[X_n] \leq \limsup_n E[X_n] \leq E[\limsup_n X_n],$$

donde $\liminf_n X_n = \limsup_n X_n$ dado que existe $\lim_{n \rightarrow \infty} X_n$. Esto demuestra que $E[\lim_{n \rightarrow \infty} X_n] = \lim_{n \rightarrow \infty} E[X_n]$. \square

Hoja 0

1.- (1.a) Supongamos un alfabeto de n letras. Determinar cuántas iniciales diferentes se pueden formar con dos letras.

(1.b) Determinar cuántas letras debería tener el alfabeto para que un millón de personas puedan ser identificadas mediante tres iniciales.

2.- Determinar cuántos números de tres cifras pueden tenerse si se emplean sólo cifras impares. Calcular cuántos de estos números son menores que 500.

3.- Se eligen cinco bolas entre diez disponibles, siendo ordenadas en cinco cajas. Determinar de cuántas maneras distintas pueden colocarse.

4.- Calcular en cuántos subconjuntos de tres elementos de cinco posibles aparece un elemento específico.

5.- Se dispone de dos mesas para tres y seis personas. Calcular de cuántas formas pueden distribuirse nueve invitados.

6.- Demostrar la identidad $2^n = \sum_{i=0}^n \binom{n}{i}$.

7.- Calcular de cuántas formas pueden ordenarse las letras de la palabra 'catarata'.

8.- Calcular cuántas señales diferentes, cada una de seis banderas colocadas en una línea vertical, pueden formarse con cuatro banderas rojas idénticas y dos banderas azules idénticas.

9.- Si se dispone de cuatro grupos de individuos (tres americanos, cuatro franceses, cuatro daneses y dos italianos), calcular de cuántas formas posibles pueden sentarse en una fila de sillas cuando aquellos de igual nacionalidad se sientan correlativamente.

10.- Calcular cómo pueden distribuirse nueve juguetes entre cuatro niños cuando el menor de éstos recibe tres juguetes y cada uno de los restantes recibe dos juguetes.

- 11.-** Con las vocales 'a, e, i, o, u' y las consonantes 'b, c, d, f', calcular el número de palabras de nueve letras distintas que pueden formarse. Calcular este número cuando no hay vocales juntas.
- 12.-** Se rellena un test de doce items, donde las respuestas son 'verdadero' y 'falso'. Se ha decidido contestar a seis items de forma aleatoria. Determinar el número de formas en que puede hacerse.
- 13.-** Calcular cuántos números naturales de cuatro cifras hay con todas las cifras diferentes.
- 14.-** Calcular cuántos números de tres cifras pueden formarse con los dígitos '1, 3, 5, 7 y 9'. Calcular cuánto suman todos ellos.
- 15.-** Supongamos el conjunto de dígitos $\{1, 2, 3, 4, 5, 6\}$.
- (15.a)** Calcular en cuántas ordenaciones de los elementos del conjunto aparecen el '1' y el '2' seguidos.
- (15.b)** Calcular en cuántas ordenaciones de los elementos del conjunto aparecen el '1' y el '2' ordenados.
- 16.- (16.a)** Calcular la cantidad de números de tres cifras que son capicúas.
(16.b) Análogo en el caso de números de cinco cifras.
- 17.-** Determinar cuántas soluciones tiene la ecuación $x + y + z = 8$ con x, y, z enteros mayores que cero. Generalizar el resultado para $x_1 + \dots + x_k = n$.
- 18.-** Para transmitir señales desde una isla a la costa, se dispone de 6 luces blancas y 6 luces rojas colocadas en el vértice de un hexágono. En cada vértice no puede haber encendida más que una luz (blanca o roja) y el número mínimo de luces encendidas es tres. Determinar cuántas señales se pueden realizar.

Hoja 1 – Introducción al Cálculo de Probabilidades

1.- Dados el conjunto $B \subset \Omega$ y las sucesiones $\{A_n : n \geq 1\} \subset \mathcal{P}(\Omega)$ y $\{B_n : n \geq 1\} \subset \mathcal{P}(\Omega)$, se pide

- (1.a) Demostrar la igualdad $\limsup A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$.
- (1.b) Si $A_n \downarrow$, demostrar que existe $\lim_{n \rightarrow \infty} A_n$ y que $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n$.
- (1.c) Demostrar que $\limsup A_n = \limsup A_{2n} \cup \limsup A_{2n-1}$ y $\liminf A_n = \liminf A_{2n} \cap \liminf A_{2n-1}$.
- (1.d) Demostrar que $\limsup(B - A_n) = B - \liminf A_n$ y $\liminf(B - A_n) = B - \limsup A_n$.
- (1.e) Demostrar que $(\limsup A_n)^c = \liminf A_n^c$ y $(\liminf A_n)^c = \limsup A_n^c$.
- (1.f) Demostrar que $\limsup(A_n \cup B_n) = \limsup A_n \cup \limsup B_n$ y $\liminf(A_n \cap B_n) = \liminf A_n \cap \liminf B_n$.

2.- Determinar los límites inferiores y superiores de $\{A_n : n \geq 1\}$ cuando:

- (2.a) $A_{2n-1} = \mathbb{Q} \cap \left[\frac{1}{n}, \frac{5n}{2n+2}\right]$ y $A_{2n} = (\mathbb{R} - \mathbb{Q}) \cap \left(-\frac{2}{n}, \frac{7n+3}{9n}\right]$.
- (2.b) $A_{3n-2} = \left(\frac{n-1}{5n+3}, \frac{2n-1}{n}\right]$, $A_{3n-1} = \left(\frac{3n}{5n+1}, \frac{3n+2}{n}\right)$ y $A_{3n} = \left[1, \frac{2n^2+1}{n+2}\right)$.
- (2.c) $A_n = \left\{x \in \mathbb{R} : \frac{1}{n} \leq x \leq 3 - \frac{1}{n}\right\}$.

3.- Supongamos $\Omega = \mathbb{R}^2$. Calcular los conjuntos $\lim_{n \rightarrow \infty} E_n$, $\lim_{n \rightarrow \infty} E_n^c$, $\lim_{n \rightarrow \infty} F_n$, $\lim_{n \rightarrow \infty} F_n^c$ y $\lim_{n \rightarrow \infty}(E_n \cap F_n^c)$, donde

$$\begin{aligned} E_n &= \left\{ (x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1 + \frac{1}{n} \right\}, \\ F_n &= \left\{ (x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq \frac{n}{n+1} \right\}. \end{aligned}$$

4.- Calcular $\lim_{n \rightarrow \infty} A_n$ en los siguientes casos:

(4.a) $A_n = \{(x, y) \in \mathbb{R}^2 : \frac{1}{n} \leq x^2 + y^2 \leq 4 - \frac{1}{n}\}.$

(4.b) $A_n = \{(x, y) \in \mathbb{R}^2 : 0 \leq x^2 + y^2 \leq \frac{1}{n}\}.$

5.- Estudiar la convergencia de la sucesión $\{A_n : n \geq 1\} \subset \mathcal{P}(\Omega)$ en los siguientes casos:

(5.a) $A_n = (-\frac{1}{n}, 1]$ si n es par, y $(-1, \frac{1}{n}]$ si n es impar.

(5.b) $A_n = (0, 1 - \frac{1}{n}]$ si n es impar, y $[\frac{1}{n}, 1)$ si n es par.

Hoja 2 – Probabilidad

1.- Sean Ω un espacio muestral y $\mathcal{A} \subset \mathcal{P}(\Omega)$ una σ -álgebra. Para $A \in \mathcal{A}$ fijado, se define

$$\mathcal{A}_A = \{B \subset \Omega : B = A \cap C \text{ con } C \in \mathcal{A}\}.$$

Demostrar que $\mathcal{A}_A \subset \mathcal{P}(A)$ es σ -álgebra.

2.- Sea $\{A_n : n \geq 1\} \subset \mathcal{A}$ una sucesión monótona decreciente, donde $\mathcal{A} \subset \mathcal{P}(\Omega)$ es σ -álgebra. Demostrar que existe $\lim_{n \rightarrow \infty} P(A_n)$ y que $\lim_{n \rightarrow \infty} P(A_n) = P(\lim_{n \rightarrow \infty} A_n)$.

3.- Sea $\{A_n : n \geq 1\} \subset \mathcal{A}$ una sucesión convergente, donde $\mathcal{A} \subset \mathcal{P}(\Omega)$ es σ -álgebra. Demostrar que existe $\lim_{n \rightarrow \infty} P(A_n)$ y que $\lim_{n \rightarrow \infty} P(A_n) = P(\lim_{n \rightarrow \infty} A_n)$.

4.- Sean $\mathcal{A}_1 \subset \mathcal{P}(\Omega_1)$ y $\mathcal{A}_2 \subset \mathcal{P}(\Omega_2)$ dos σ -álgebras, y sea $E \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Demostrar que $E_{\omega_1} \in \mathcal{A}_2$, para cada $\omega_1 \in \Omega_1$, y que $E^{\omega_2} \in \mathcal{A}_1$, para cada $\omega_2 \in \Omega_2$.

5.- Sean $\mathcal{A} \subset \mathcal{P}(\Omega)$ una σ -álgebra y $\{A_n : n \geq 1\} \subset \mathcal{A}$ una sucesión. Demostrar las siguientes desigualdades:

(5.a) $P(\liminf A_n) \leq \liminf P(A_n)$.

(5.b) $\limsup P(A_n) \leq P(\limsup A_n)$.

Además, resolver el ejercicio 3 desde (5.a) y (5.b).

6.- Sean (Ω, \mathcal{A}) un espacio probabilizable y $\{P_n : n \geq 1\}$ una sucesión de medidas de probabilidad sobre (Ω, \mathcal{A}) . Se define $P : \mathcal{A} \rightarrow [0, 1]$ por

$$P(A) = \sum_{n=1}^{\infty} a_n P_n(A),$$

donde $a_n \geq 0$ y $\sum_{n=1}^{\infty} a_n = 1$. Demostrar que P es una medida de probabilidad sobre (Ω, \mathcal{A}) .

7.- Sea (Ω, \mathcal{A}) el espacio probabilizable con $\Omega = \{0, 1, 2, \dots\}$ y $\mathcal{A} = \mathcal{P}(\Omega)$. Demostrar que las funciones de conjunto P dadas en (7.a) y (7.b) son medidas de probabilidad sobre (Ω, \mathcal{A}) y determinar las probabilidades de los conjuntos de (7.c).

(7.a) $P(A) = \sum_{x \in A} e^{-\lambda} \lambda^x / x!$, siendo $\lambda > 0$.

(7.b) $P(A) = \sum_{x \in A} p(1-p)^x$, siendo $p \in (0, 1)$.

(7.c) $A = \{x > 2\}$, $B = \{x < 3\}$, $C = \{3 < x < 6\}$, $A \cap B$, $A \cup B$, $B \cap C$, $A \cap C^c$ y $B \cap C^c$.

8.- Sean \mathbb{N}_p y \mathbb{N}_i los conjuntos de números naturales pares e impares, respectivamente. Se define la familia de subconjuntos

$$\mathcal{C} = \{A \cup \mathbb{N}_i : A \subset \mathbb{N}_p\}.$$

Estudiar si $\mathcal{C} \subset \mathcal{P}(\mathbb{N})$ tiene estructura de álgebra.

9.- Demostrar que $\mathcal{C} = \{A \subset \mathbb{N} : A \text{ finito o } A^c \text{ finito}\}$ tiene estructura de álgebra, pero no es una σ -álgebra de $\mathcal{P}(\mathbb{N})$.

10.- Sean $\Omega = \{0, 1, 2, \dots\}$ y $\mathcal{C} = \{A \subset \Omega : 2n \in A \text{ si y sólo si } 2n+1 \in A\}$. Demostrar que $\mathcal{C} \subset \mathcal{P}(\Omega)$ es σ -álgebra.

11.- Sean Ω un conjunto no-numerable y $\mathcal{C} = \{A \subset \Omega : A \text{ o } A^c \text{ es numerable}\}$. Estudiar si $\mathcal{C} \subset \mathcal{P}(\Omega)$ es σ -álgebra.

12.- De los 30 temas de un examen, un alumno conoce 18 temas. Se proponen dos tipos de examen:

1. Se eligen 3 temas al azar y el alumno debe contestar 2.
2. Se eligen 5 temas al azar y el alumno debe contestar 3.

Determinar el tipo de examen más favorable al alumno.

13.- Calcular la probabilidad de que, al tirar una moneda n veces, obtengamos la k -ésima cara en la n -ésima tirada.

14.- Se tiene un manojó de N llaves donde sólo una de ellas abre una puerta. Suponiendo que cada llave probada es retirada del manojó, determinar la probabilidad de que la puerta se abra en el k -ésimo intento. (Nota: todas las llaves deben ser probadas, por lo que es necesario hacer N intentos.) Determinar la probabilidad de este suceso cuando las llaves no se retiran del manojó.

15.- En una urna se introducen n bolas, cada una de ellas de color blanco o negro con igual probabilidad. Se extraen k bolas con reemplazamiento desde

la urna. Determinar la probabilidad de que la urna sólo contenga bolas blancas si las k bolas extraídas han sido blancas.

16.- Se consideran $N + 1$ urnas idénticas numeradas desde 0 hasta N , conteniendo N bolas. En concreto, la i -ésima urna contiene i bolas negras y $N - i$ bolas blancas, para $0 \leq i \leq N$. Se escoge una urna al azar y se extraen n bolas una-a-una con reemplazamiento. Si las n bolas extraídas resultan ser negras, determinar la probabilidad de que, al extraer la $(n + 1)$ -ésima bola, ésta sea de color negro.

17.- De una urna con a bolas blancas y b bolas negras se extraen k bolas al azar. ¿Cuál es la probabilidad de que entre las k bolas haya exactamente r bolas blancas?

18.- Se lanzan 2 dados n veces. ¿Cuál es la probabilidad de obtener al menos un 6 doble?

19.- Una urna contiene 5 bolas rojas, 3 verdes, 2 amarillas y 4 blancas. Se extraen 8 bolas al azar. Calcular la probabilidad de que:

(19.a) Exactamente sean 2 rojas, 2 verdes, 1 amarilla y 3 blancas.

(19.b) Estén todas las bolas blancas.

(19.c) Haya, al menos, una bola roja.

20.- Con una moneda se juega a cara o cruz. Se suspenden los lanzamientos cuando por primera vez la diferencia entre el número de caras y el número de cruces es, en valor absoluto, igual a 3. Calcular la probabilidad de que los lanzamientos se suspendan en la sexta tirada o antes.

21.- Los jugadores A, B y C participan en el siguiente juego: se lanza un dado y A gana si sale 1 ó 3; B gana si sale 4 ó 5; C gana si sale 6; y si sale 2 se vuelve a lanzar el dado. Calcular la probabilidad de que gane A, de que gane C y de que gane B.

22.- Una urna contiene 5 bolas negras y 4 blancas, otra urna contiene 4 bolas negras y 5 blancas. Supongamos que se trasladan 2 bolas de la primera a la segunda urna y, a continuación se extrae una bola de la segunda urna. ¿Cuál es la probabilidad de que la bola extraída sea blanca?

23.- Se sabe que al lanzar cinco monedas aparecieron al menos dos caras. ¿Cuál es la probabilidad de que el número de caras exacto fuese tres?

24.- Disponemos de una moneda y dos dados A y B. A tiene 4 caras rojas y 2 blancas, y B tiene 2 caras rojas y 4 blancas. Se lanza una moneda, si sale cara se lanza repetidas veces el dado A y si sale cruz se hace lo mismo con el dado B.

(24.a) Calcular la probabilidad de que en el primer lanzamiento la cara observada sea roja.

(24.b) Sabiendo que en los dos primeros lanzamientos hemos observado dos caras rojas, ¿cuál es la probabilidad de que el dado lanzado sea el A?

25.- Se lanzan dos monedas, si el resultado es CC se extraen dos bolas de una urna U_1 que contiene 3 bolas rojas, 2 blancas y 4 negras. Si el resultado es CX se extraen de U_2 que contiene 2 rojas, 1 blanca y 5 negras, y si sale XX o XC las bolas se extraen de U_3 que contiene 6 rojas, 4 blancas y 6 negras. Si las dos bolas extraídas resultaron ser una blanca y otra roja, ¿cuál es la probabilidad de que sean de U_1 ? ¿y de U_2 ?

26.- Determinada batería antiaérea disparaba sobre un avión. Para derribar el aparato bastaba con alcanzar ambos reactores o la cabina del piloto. Sean p_1 la probabilidad de alcanzar el primer reactor, p_2 la probabilidad de alcanzar el segundo y p_3 la probabilidad de dar en la cabina del piloto. Se supone que estos tres puntos sensibles eran tocados uno independientemente del otro. Determinar la probabilidad de que dicho avión hubiese sido derribado.

27.- El volumen diario de producción de 3 plantas diferentes de una fábrica es de 500 unidades en la primera, 1000 en la segunda y 2000 en la tercera. Sabiendo que el porcentaje de unidades defectuosas producidas en cada planta es del 1%, 0.8% y 2% respectivamente, determinar la probabilidad de que

(27.a) Extraída una unidad al azar resulte ser no defectuosa.

(27.b) Habiendo sido extraída una unidad defectuosa, haya sido producida en la primera planta.

28.- Se supone que n bolas se colocan al azar en n cajas numeradas. Calcular las probabilidades de que:

(28.a) La caja 1 esté vacía.

(28.b) Alguna caja esté vacía.

(28.c) La caja 1 sea la única vacía.

(28.d) Hay una única caja vacía.

29.- En un colegio electoral de 42 electores, 15 han votado la lista A y el resto la B. Se seleccionan 10 electores, calcúlese la probabilidad de que, como máximo, tres de ellos hayan votado la lista A.

30.- De una baraja española (40 cartas) repartida en su totalidad entre 4 jugadores, hallar la probabilidad de que haya como mínimo un jugador cuya mano sean cartas todas del mismo palo.

31.- Se tiene una moneda y una urna. Se lanza la moneda al aire. Si sale cara, se introduce una bola negra en la urna; si sale cruz, la bola introducida es de color blanco. El proceso se repite 10 veces. A continuación, se extrae una bola de la urna, se mira su color y se devuelve a la urna. El procedimiento se repite 10 veces. Finalizado éste se ha observado que las 10 bolas extraídas eran de color blanco. ¿Cuál es la probabilidad de que la urna contenga sólo bolas blancas?

32.- Dos de las cuatro válvulas de un aparato, que funcionan independientemente, han fallado. Calcular la probabilidad de que hayan sido la 1 y la 2, si la probabilidad de que falle la válvula i es $i/10$.

Hoja 3 – Variables aleatorias unidimensionales

1.- Sean $\mathcal{C} \subset \mathcal{P}(\Omega_2)$ y $f : \Omega_1 \longrightarrow \Omega_2$ una aplicación.

(1.a) Demostrar que $f^{-1}(A^c) = (f^{-1}(A))^c$, $f^{-1}(\cup_{i \in I} A_i) = \cup_{i \in I} f^{-1}(A_i)$ y $f^{-1}(\cap_{i \in I} A_i) = \cap_{i \in I} f^{-1}(A_i)$.

(1.b) Demostrar que $\sigma(f^{-1}(\mathcal{C})) = f^{-1}(\sigma(\mathcal{C}))$.

2.- Sea una urna con 3 bolas blancas, 2 negras y 1 verde. Se extraen 3 bolas al azar y se considera ξ definida como el "número de bolas blancas extraídas". Construir la función de probabilidad inducida por ξ y su función de distribución.

3.- Sea Ω el espacio muestral asociado al lanzamiento de una moneda equilibrada en tres ocasiones, con puntos muestrales denotados por ω_{ijk} con $i, j, k \in \{C, X\}$. Sea

$$\mathcal{A} = \{\phi, A_1, A_2, A_3, A_1 \cup A_2, A_1 \cup A_3, A_2 \cup A_3, A_1 \cup A_2 \cup A_3\}$$

una σ -álgebra sobre $\mathcal{P}(\Omega)$, donde $A_1 = \{\omega_{ccx}, \omega_{cxc}\}$, $A_2 = \Omega - (A_1 \cup A_3)$ y $A_3 = \{\omega_{xxx}\}$. Sea $X(\omega)$ el "número de caras obtenidas en el resultado ω ". Estudiar si X es una variable aleatoria.

4.- Sean $\Omega = \mathbb{Z}_+$, $\mathcal{A} = \mathcal{P}(\Omega)$ y $P(\{\omega\}) = 2^{-\omega}$, $\forall \omega \in \Omega$. Se define $\xi(\omega)$ como el "resto de ω (módulo k)". Demostrar que ξ es una variable aleatoria y determinar los valores $P(\xi = r)$, para $r \in \{0, 1, \dots, k-1\}$.

5.- Sea $f : \mathbb{R} \longrightarrow \mathbb{R}^+ = [0, \infty)$ una función Riemann-integrable tal que $\int_{-\infty}^{\infty} f(u) du = 1$. Demostrar que $F(x) = \int_{-\infty}^x f(u) du$ define una función de distribución absolutamente continua.

6.- La duración T de las conferencias telefónicas en una central es una variable aleatoria con función de densidad

$$f(t) = \begin{cases} \alpha e^{-kt}, & \text{si } t > 0, \\ 0, & \text{si } t \leq 0. \end{cases} \quad (k > 0)$$

- (6.a) Calcular α para que f sea función de densidad.
- (6.b) Si $1/k = 2$ minutos, calcular la probabilidad de que una conversación dure más de 3 minutos.
- (6.c) Probabilidad de que una conversación dure entre 3 y 6 minutos.

7.- Sea F la función definida por

$$F(x) = \begin{cases} 0, & \text{si } x < 1/3, \\ x^{2\alpha-1}, & \text{si } 1/3 \leq x < \alpha, \\ 1, & \text{si } \alpha \leq x. \end{cases}$$

- (7.a) Determinar los valores de α para que F sea función de distribución.
- (7.b) Determinar α para que F sea discreta. Análogo para el caso absolutamente continuo.
- (7.c) Calcular $P(\liminf A_n)$ y $P(\limsup A_n)$ cuando $A_n = [\frac{1}{3} + \frac{1}{3^{n+1}}, \alpha)$.

8.- Sea

$$F(x) = \begin{cases} 0, & \text{si } x < 1, \\ (x+1)/10, & \text{si } 1 \leq x < 3/2, \\ 1/3 + (x-3/2)/3, & \text{si } 3/2 \leq x < 5/2, \\ 1, & \text{si } 5/2 \leq x. \end{cases}$$

- (8.a) Comprobar que F es función de distribución. Determinar las funciones de distribución F_1 discreta y F_2 absolutamente continua tales que $F(x) = \lambda F_1(x) + (1-\lambda)F_2(x)$, $\forall x \in \mathbb{R}$.
- (8.b) Evaluar $P_F(\mathbb{Q})$ y $P_F(\mathbb{R} - \mathbb{Q})$.
- (8.c) Dada la sucesión de subconjuntos

$$A_{2n-1} = \left(\frac{1}{n+1}, \frac{5n+2}{n+1} \right),$$

$$A_{2n} = \left(\frac{4n+3}{n}, \frac{8n+1}{n+1} \right),$$

se pide evaluar $P(\liminf A_n)$ y $P(\limsup A_n)$.

9.- Demostrar que la función

$$F(x) = \begin{cases} 0, & \text{si } x < 0, \\ 1 - 2^{-x-1} - 2^{-[x]-1}, & \text{si } x \geq 0, \end{cases}$$

es una función de distribución de tipo mixto, donde $[x]$ denota la parte entera de x .

10.- Sea la función de densidad de una variable aleatoria X con distribución Cauchy

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

Determinar la distribución de la variable aleatoria $Y = 3X + 1$.

11.- Sea X una variable aleatoria con función de distribución

$$F(x) = \begin{cases} 0, & \text{si } x < -1, \\ 0.2(x+1), & \text{si } -1 \leq x < 0, \\ 0.3, & \text{si } 0 \leq x < 1/2, \\ 0.3 + 2(x-0.5)^2, & \text{si } 1/2 \leq x < 1, \\ 0.9 + 0.2(x-1), & \text{si } 1 \leq x < 3/2, \\ 1, & \text{si } 3/2 \leq x. \end{cases}$$

(11.a) Comprobar que F es función de distribución. Determinar las funciones de distribución F_1 discreta y F_2 absolutamente continua tales que $F(x) = \lambda F_1(x) + (1-\lambda)F_2(x)$, $\forall x \in \mathbb{R}$.

(11.b) Determinar la distribución de $Y = |X|$.

12.- Sea X una variable aleatoria exponencial de tasa $\lambda = 1$. Se define $Y = X^2$ si $0 \leq X < 2$, $Y = 4$ si $2 \leq X < 3$, $Y = -4(X-4)$ si $3 \leq X < 4$, e $Y = 0$ si $4 \leq X$. Determinar la distribución de Y .

13.- Sea X una variable aleatoria discreta con soporte \mathcal{D}_X y función de masa p_X . Sean $g : \mathbb{R} \rightarrow \mathbb{R}$ una función medible e $Y = g(X)$ la variable aleatoria transformada. Demostrar que Y es una variable aleatoria discreta con soporte $\mathcal{D}_Y = g(\mathcal{D}_X)$ y función de masa

$$p_Y(y) = \sum_{\{x \in \mathbb{R} : g(x)=y\} \cap \mathcal{D}_X} p_X(x), \quad \text{si } y \in \mathcal{D}_Y.$$

14.- Sea (\mathbb{R}, β, P) un espacio de probabilidad, siendo P la medida de probabilidad relativa a la función de distribución

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

Sean las variables aleatorias $\xi_1 = X^2$, $\xi_2 = X^3$ y $\xi_3 = e^{-X}$, supuesto que X es una variable aleatoria con función de distribución F . Calcular las funciones de distribución inducidas por cada una de ellas.

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

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

(15.a) Determinar los valores de k tales que f es una función de densidad.

(15.b) Calcular la esperanza, la moda y la mediana de X .

(15.c) ¿Para qué valores de k es máxima la varianza de X ?

16.- Sea X una variable aleatoria 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 $Var(X)$.

17.- Sea

$$f(x) = \begin{cases} 1, & \text{si } 0 < x < 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

la función de densidad de una variable aleatoria X y sea $Y = -2 \ln X$. Calcular la distribución de Y .

18.-

(18.a) Determinar la función generatriz de momentos de una variable aleatoria X discreta con momentos $E[X^n] = \frac{1}{n+1}$.

(18.b) Determinar la función generatriz de momentos de una variable aleatoria X uniforme sobre el intervalo (a, b) .

(18.c) Se sabe que la función generatriz de momentos de una variable aleatoria X discreta viene dada por

$$M(t) = \frac{1 - \alpha}{1 - \alpha e^t}, \quad \alpha \in (0, 1).$$

Determinar la función de masa de X .

19.- Dada la función

$$\varphi(t) = \frac{1}{2 - e^{it}},$$

se pide:

(19.a) Comprobar que φ es la función característica de una variable aleatoria X discreta y hallar la función de masa, la media y la varianza de X .

(19.b) Escribir la función característica de $Y = 1 + 2X$, así como su media y su varianza.

(19.c) Si $Y_1 = (Y+2)^2$ e $Y_2 = X^2+5$, calcular $P(Y_1 \in [10, 50])$ y $P(Y_2 \in [6, 10])$.

20.- Determinar la función característica de las variables aleatorias que cumplen:

(20.a) Toda la probabilidad está concentrada en un punto.

(20.b) Toda la probabilidad está concentrada en n puntos.

21.- Demostrar que $\varphi(t) = \frac{se^{nt}}{t}$ es la función característica de una variable aleatoria.

22.- El porcentaje de piezas defectuosas fabricadas por una máquina es el 4%. Las piezas se empaquetan en lotes de 100. El cliente rechaza el lote si contiene más de dos piezas defectuosas. Calcular el porcentaje de lotes rechazados que puede esperar el fabricante si el proceso de fabricación no ha sufrido modificaciones.

23.- En dos grupos de segundo año de carrera se ha medido el coeficiente de inteligencia de los alumnos. En el grupo A la media fue 100 y la desviación típica 10, mientras que en el grupo B estas medidas fueron 105 y 12, respectivamente. Supongamos que ambos grupos tienen el mismo número de alumnos. Se escoge un alumno al azar y se comprueba que su coeficiente es mayor que 120. Calcular, suponiendo normalidad, la probabilidad de que el citado alumno pertenezca al grupo B.

Hoja 4 – Variables aleatorias multidimensionales

1.- Estudiar si

$$F(x, y) = \begin{cases} 1, & \text{si } x + 2y \geq 1, \\ 0, & \text{si } x + 2y < 1, \end{cases}$$

es una función de distribución en \mathbb{R}^2 .

2.- Dada la variable aleatoria 2-dimensional (X, Y) tal que

$$P(X = 1, Y = 0) = P(X = 0, Y = 1) = P(X = 1, Y = 1) = \frac{1}{3},$$

calcular la función de distribución conjunta de (X, Y) y las funciones de distribución marginales de X e Y .

3.- Sea (X, Y) una variable aleatoria 2-dimensional con distribución uniforme sobre el recinto

$$\mathcal{C} = \left\{ (x, y) \in \mathbb{R}^2 : y \leq \frac{x}{3}, x \leq 3, y \geq 0 \right\}.$$

Calcular la función de densidad conjunta de (X, Y) , la función de distribución conjunta de (X, Y) y las distribuciones marginales de X e Y .

4.- Dada la variable aleatoria 2-dimensional (X, Y) con función de densidad conjunta

$$f(x, y) = \begin{cases} 2^{-1} \sin(x + y), & \text{si } 0 \leq x \leq \pi/2, 0 \leq y \leq \pi/2, \\ 0, & \text{en caso contrario,} \end{cases}$$

calcular:

(4.a) Las esperanzas de X e Y .

(4.b) La matriz de varianza-covarianzas de (X, Y) .

5.- Sea (X, Y) una variable aleatoria 2-dimensional con función de densidad $f(x, y) = 24y(1 - x - y)$ sobre el recinto $\mathcal{C} = \{(x, y) \in \mathbb{R}^2 : x + y \leq 1, x \geq 0, y \geq 0\}$. Calcular:

(5.a) La función de distribución conjunta.

(5.b) Las funciones de densidad marginales.

(5.c) Las funciones de densidad condicionadas.

6.- Se sitúan de forma aleatoria e independiente N puntos en el intervalo $(0, T)$. Si X representa la distancia de 0 al primer punto e Y denota la distancia de 0 al segundo punto, entonces calcular la distribución conjunta y las correspondientes marginales de (X, Y) .

7.- La probabilidad de que desde un huevo nazca un insecto es p . En una flor, el número de huevos puestos por estos insectos sigue una distribución Poisson de media λ .

(7.a) Calcular la distribución del número de insectos que nace en una flor.

(7.b) Se ha observado una flor y se ha constatado que el número de insectos que han nacido en ella ha sido n . Calcular la distribución del número de huevos que había en la flor.

8.- Sea ξ una variable aleatoria discreta con función de masa

$$p(x) = \frac{1}{2^x}, \quad x = 1, 2, \dots$$

Supongamos que $\eta/\xi = x$ es una variable aleatoria con función de densidad

$$f(y/x) = x(1 - y)^{x-1}, \quad 0 < y < 1.$$

Determinar la distribución de η .

9.- Sea X una variable aleatoria con distribución Exponencial de parámetro $\lambda = 1$. Supongamos que $Y/X = x$ es una variable aleatoria con función de densidad

$$f(y/x) = \begin{cases} xy^{-(x+1)}, & \text{si } y > 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

Calcular:

(9.a) La función de densidad conjunta.

(9.b) La función de densidad marginal de Y .

(9.c) La función de densidad de $X/Y = y$.

10.- Se considera la variable aleatoria 2-dimensional (ξ, η) , donde ξ sigue una distribución Poisson de parámetro λ y $\eta/\xi = x$ sigue una distribución Normal de media $x^2 - 2x$ y varianza σ^2 . Calcular la esperanza de η y la matriz de varianza-covarianzas.

11.- Calcular la matriz de varianza-covarianzas de la variable aleatoria 2-dimensional con función característica

$$\phi(t, u) = \frac{9e^{3i(t+2u)-(t^2+4u^2)/2}}{(3-it)^2}.$$

12.- Se considera la variable aleatoria 2-dimensional (ξ, η) con función de masa

$$\begin{aligned} P(-1, -1) &= \frac{1}{16}, & P(-1, 0) &= \frac{3}{16}, & P(-1, 1) &= 0, \\ P(0, -1) &= \frac{1}{16}, & P(0, 0) &= \frac{1}{4}, & P(0, 1) &= \frac{3}{16}, \\ P(1, -1) &= \frac{1}{8}, & P(1, 0) &= \frac{1}{16}, & P(1, 1) &= \frac{1}{16}. \end{aligned}$$

Demostrar que $\phi_{\xi+\eta}(t) = \phi_{\xi}(t)\phi_{\eta}(t)$ y mostrar que, sin embargo, ξ y η no son independientes.

13.- Sea (X_1, X_2) una variable aleatoria 2-dimensional con función de densidad

$$f(x_1, x_2) = \begin{cases} 1, & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

Se considera la transformación (Y_1, Y_2) , donde $Y_1 = \max\{X_1, X_2\}$ e $Y_2 = \min\{X_1, X_2\}$. Calcular la función de densidad de (Y_1, Y_2) .

14.- Sea (X, Y) una variable aleatoria 2-dimensional con función de densidad uniforme en el recinto

$$\mathcal{C} = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq 1\}.$$

Calcular las distribuciones de Z y (Z, T) , donde $Z = X + Y$ y $T = X - Y$.

15.- Sea (ξ_1, ξ_2) una variable aleatoria 2-dimensional con función de densidad

$$f(x, y) = \begin{cases} 4xy, & \text{si } 0 < x < 1, 0 < y < 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

Hallar la distribución conjunta de (η_1, η_2) , donde $\eta_1 = \xi_1 + \xi_2$ y $\eta_2 = \max\{\xi_1, \xi_2\}$.

16.- Se consideran las variables aleatorias X e Y independientes, donde $X \sim \text{Uniforme}(1, 3)$ e Y tiene función de densidad

$$f_Y(y) = \begin{cases} 2^{-1}e^{2-y}, & \text{si } y > 2, \\ 0, & \text{en caso contrario.} \end{cases}$$

Calcular:

(16.a) La distribución conjunta de (Z, W) , donde $Z = X/Y$ y $W = XY$.

(16.b) La distribución marginal de Z .

17.- Sean X_1, \dots, X_{n+1} variables aleatorias independientes e idénticamente distribuidas según una distribución Bernoulli de parámetro p . Sean

$$V_n = \prod_{i=1}^n X_i \quad \text{y} \quad V_{n+1} = \prod_{i=1}^{n+1} X_i.$$

Calcular la distribución conjunta de (V_n, V_{n+1}) , su función característica y deducir si V_n y V_{n+1} son o no independientes.

18.- Sean X e Y variables aleatorias independientes e idénticamente distribuidas con distribución Exponencial de parámetro $\lambda = 1$. Calcular la distribución de $T = |X - Y|$.

19.- Sea (ξ_1, ξ_2) una variable aleatoria con función de densidad

$$f(x_1, x_2) = \begin{cases} 2^{-1}e^{-x_1}, & \text{si } x_1 > 0, -1 < x_2 < 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

Hallar la distribución de la variable aleatoria $\eta = |\xi_1 + \xi_2|$.

20.- Sea (X, Y) una variable aleatoria 2-dimensional con función de densidad

$$f(x, y) = \begin{cases} k^2 e^{-ky}, & \text{si } 0 < x < y, \\ 0, & \text{en caso contrario.} \end{cases}$$

(20.a) Hallar la probabilidad del recinto $[0, 1] \times [0, 1]$.

(20.b) Determinar los valores de k para que f sea una función de densidad.

(20.c) Demostrar que X e $Y - X$ son independientes.

(20.d) Escribir la curva (general) de regresión y la recta de regresión de Y sobre X .

21.- Consideremos la variable aleatoria 2-dimensional (X, Y) con distribución uniforme en el recinto limitado por las rectas $y = x$, $x = -y$, $y = 1$ e $y = -1$; es decir, $f(x, y) = 1/2$ para $(x, y) \in \mathcal{C} = \{(x, y) \in \mathbb{R}^2 : |x| < |y| < 1\}$.

(21.a) Calcular la curva (general) de regresión de Y sobre X .

(21.b) Estudiar si X e Y son independientes y/o incorreladas.

22.- Consideremos las variables aleatorias X e Y , con rectas de regresión

$$\begin{cases} 3x + 2y - 26 = 0, \\ 6x + y - 31 = 0. \end{cases}$$

Calcular las medias marginales y el coeficiente de correlación. Identificar cuál es la recta de regresión de Y sobre X .

23.- Se considera una variable aleatoria (X, Y) con distribución $Normal_2$. Se sabe que las rectas de regresión tienen por ecuaciones $x - y + 2 = 0$, $3x - 10y + 40 = 0$; y que la suma de las varianzas de X e Y es 1.3. Determinar la correspondiente función de densidad.

Hoja 5 – Convergencias estocásticas

1.- Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas Normales de media 0 y varianza 1. Estudiar la convergencia en probabilidad de la sucesión $\{Y_n : n \geq 1\}$, donde $Y_n = n^{-1} \sum_{i=1}^n X_i^2$.

2.- Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas Bernoulli(p), con $0 < p < 1$. Sea la variable aleatoria $V_n = X_1 X_2 \dots X_n$. Estudiar la convergencia en ley, la convergencia en probabilidad y la convergencia casi segura de la sucesión $\{V_n : n \geq 1\}$.

3.- Sea (X, Y) una variable aleatoria con función de densidad conjunta

$$f(x, y) = \begin{cases} 4^{-1} \exp\{-x/2\}, & \text{si } x > 0, -1 < y < 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

Se pide:

(3.a) Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con ley $\mathcal{L}(X)$. Hallar la distribución de

$$Y_n = \sum_{i=1}^n X_i.$$

Estudiar la convergencia en ley de la sucesión $\{V_n : n \geq 1\}$, donde $V_n = n^{-2} Y_n$.

(3.b) Sea Z_n una variable aleatoria tal que $Z_n/X = x$ es Poisson de parámetro nx . Calcular la distribución de Z_n y estudiar la convergencia en ley de la sucesión $\{Z_n : n \geq 1\}$.

4.- Demostrar que si $\{X_n : n \geq 1\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas que converge en ley a la constante c , entonces la sucesión $\{e^{X_n} : n \geq 1\}$ converge en ley hacia e^c . Usar este resultado para estudiar la convergencia en ley y en probabilidad de

$$Y_n = \left(\prod_{i=1}^n X_i \right)^{1/n},$$

donde $\{X_n : n \geq 1\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas Uniformes sobre el intervalo $(0, 1)$.

5.- Sea $\{X_n : n \geq 0\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas y definamos

$$Y_n = \sum_{i=0}^n \frac{X_i}{2^{n+1-i}}.$$

Estudiar la convergencia en ley de $\{Y_n : n \geq 0\}$ cuando:

(5.a) $X_i \sim \text{Normal}(0, \sigma^2)$, $i \geq 0$.

(5.b) $X_i \sim \text{Cauchy}(0, 1)$, $i \geq 0$.

6.- Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes con funciones de densidad

$$f_{X_n}(x) = \frac{1}{\pi} \frac{n}{1 + (nx)^2}.$$

Estudiar la convergencia en probabilidad y la convergencia en media de orden 2 de $\{X_n : n \geq 1\}$.

7.- Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes con distribuciones Poisson de parámetro n . Demostrar que

$$\frac{X_n - n}{\sqrt{n}}$$

converge en ley a X , donde X es una variable aleatoria Normal $(0, 1)$.

8.- Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas Uniforme sobre $(0, 1)$. Consideremos

$$\begin{aligned} Y_n &= \max(X_1, \dots, X_n), \\ Z_n &= nY_n, \\ T_n &= n^\alpha(1 - Y_n). \end{aligned}$$

Estudiar la convergencia en ley de las sucesiones $\{Y_n : n \geq 1\}$, $\{Z_n : n \geq 1\}$ y $\{T_n : n \geq 1\}$.

9.- Estudiar si las siguientes sucesiones de variables aleatorias independientes cumplen la ley débil y/o la ley fuerte de los grandes números:

(9.a) $\{X_n : n \geq 1\}$, donde

$$\begin{aligned} P(X_n = 2^n) &= P(X_n = -2^n) = \frac{1}{2^{2n+1}}, \\ P(X_n = 0) &= 1 - \frac{1}{2^{2n}}. \end{aligned}$$

(9.b) $\{X_n : n \geq 1\}$, donde

$$P\left(X_n = \sqrt{\ln(n+\alpha)}\right) = P\left(X_n = -\sqrt{\ln(n+\alpha)}\right) = \frac{1}{2}, \quad \alpha \in \mathbb{N}.$$

10.- Estudiar la convergencia en ley de la sucesión $\{Y_n : n \geq 1\}$, donde $Y_n = \sum_{i=1}^n X_i$ y $\{X_n : n \geq 1\}$ es una sucesión de variables aleatorias independientes con función de masa

$$P\left(X_i = \frac{1}{2^i}\right) = P\left(X_i = -\frac{1}{2^i}\right) = \frac{1}{2}.$$

11.- Sea X una variable aleatoria Exponencial de tasa λ . Sea

$$Y_n(\omega) = \begin{cases} 0, & \text{si } X(\omega) \in [0, 1 - 1/n), \\ 1, & \text{si } X(\omega) \in [1 - 1/n, n), \\ n, & \text{si } X(\omega) \in [n, \infty). \end{cases}$$

Estudiar la convergencia en ley, en probabilidad, en media de orden 2 y casi segura de la sucesión $\{Y_n : n \geq 1\}$.

12.- Una empresa está especializada en la producción de tuercas. La experiencia en los últimos años muestra que la probabilidad de producir una tuerca defectuosa es 0.035. Las tuercas se empaquetan en cajas de 100 unidades. Determinar el porcentaje de cajas sin más de dos tuercas defectuosas.

13.- Una urna contiene diez bolas numeradas del cero al nueve. De la urna se extraen n bolas con reemplazamiento.

(13.a) Señalar qué dice la ley de los grandes números sobre la aparición de ceros.

(13.b) Determinar cuántas extracciones será preciso efectuar para que, con probabilidad 0.95, la frecuencia relativa de aparición de ceros esté comprendida entre 0.09 y 0.11.

(13.c) Utilizar el teorema central del límite para calcular la probabilidad de que, entre los n números elegidos, el 'cinco' aparezca entre $(n - 3\sqrt{n})/10$ y $n + 3\sqrt{n}$. Particularizar para $n = 25$ y $n = 100$.

14.- Sea $\{X_n : n \geq 1\}$ una sucesión de variables aleatorias independientes donde X_i tiene función de masa

$$\begin{aligned} P(X_i = -2) &= P(X_i = -1) = P(X_i = 1) = P(X_i = 2) = \frac{1}{4i}, \\ P(X_i = 0) &= 1 - \frac{1}{i}. \end{aligned}$$

Estudiar la convergencia en ley, en probabilidad, en media cuadrática y casi seguro. Analizar si se cumple la ley fuerte de los grandes números.

15.- Sea ξ una variable aleatoria Uniforme sobre el intervalo $(-1/\theta, 1/\theta)$. Se definen las variables aleatorias $\eta = \theta\xi$ y

$$X_i(\omega) = \begin{cases} -1, & \text{si } \eta(\omega) \in [-1, -1/i), \\ 0, & \text{si } \eta(\omega) \in [-1/i, 1/i), \\ 1, & \text{si } \eta(\omega) \in [1/i, 1). \end{cases}$$

Estudiar la convergencia en probabilidad, casi segura y en media cuadrática de la sucesión $\{X_n : n \geq 1\}$.

16.- La duración de una bombilla se distribuye Exponencial con media 1 mes. Cada vez que una bombilla se estropea es reemplazada inmediatamente por otra nueva. Determinar cuál es el número mínimo de bombillas que deben tenerse para que, con probabilidad 0.95, tengamos luz durante un intervalo de tiempo de dos años.