

Las 7 maravillas del mundo

Índice

I	Teoría de la Probabilidad	5
1.	Revisión de conceptos básicos	5
1.1.	Definición de Espacio de Probabilidad	5
1.2.	Propiedades de la familia de sucesos	5
1.2.1.	Límites de sucesiones de conjuntos	5
1.2.2.	Familias monótonas	6
1.2.3.	Familias engendradas por una familia	7
1.2.4.	σ -álgebra de Borel en \mathbb{R}	7
1.2.5.	σ -álgebra de Borel en \mathbb{R}^k	8
1.3.	Primeras propiedades de la función de probabilidad P	8
1.3.1.	Probabilidad de la unión	9
1.3.2.	Subaditividad de la función de probabilidad	9
1.3.3.	Propiedades de límite de la función P	10
1.3.4.	Límites para familias monótonas	10
1.4.	Probabilidad condicionada	10
1.4.1.	Propiedades	11
1.4.2.	Teorema de la probabilidad compuesta	11
1.4.3.	Teorema de la probabilidad total	11
1.4.4.	Teorema de Bayes	12
2.	Independencia	13
2.1.	Independencia de dos sucesos	13
2.2.	Independencia entre n sucesos	13
2.3.	Independencia entre infinitos sucesos. Lema de Borell-Cantelli	14
2.4.	Independencia entre familias	16
2.5.	π -sistemas y d-sistemas	16
2.6.	Aplicación a la independencia	18
3.	Distribuciones en \mathbb{R} y \mathbb{R}^k	19
3.1.	Introducción. Sumas con un conjunto infinito de sumandos	19
3.2.	Distribuciones discretas en un espacio cualquiera	19
3.3.	Distribuciones en \mathbb{R} y \mathbb{R}^k . Distribuciones discretas	20
3.4.	La integral en las distribuciones continuas	20
3.5.	Medidas absolutamente continuas	21

3.6.	Distribuciones continuas en \mathbb{R} y \mathbb{R}^k	21
3.7.	Distribuciones uniformes	22
3.8.	Distribuciones singulares	22
3.9.	Descomposición de Lebesgue	23
4.	Función de distribución	26
4.1.	Funciones monótonas. Notaciones. Propiedades	26
4.2.	Definición de función de distribución de una dimensión. Función de distribución de una función de probabilidad en \mathbb{R}	26
4.2.1.	Distribuciones discretas	27
4.2.2.	Distribuciones continuas	27
4.3.	Ejemplo de distribución singular en \mathbb{R}	27
4.4.	Incremento mixto y continuidad lateral de funciones de dos variables	29
4.4.1.	Operador diferencia	29
4.4.2.	Incremento mixto	29
4.4.3.	Límites laterales en el plano	30
4.5.	Función de distribución en \mathbb{R}^2 . Función de distribución de una distribución de probabilidad en \mathbb{R}^2	30
4.5.1.	Distribuciones discretas	32
4.5.2.	Distribuciones continuas	32
4.6.	Funciones con varias variables. Incremento mixto. Continuidad lateral	32
4.7.	Función de distribución en \mathbb{R}^k . Función de distribución de una distribución de probabilidad en \mathbb{R}^k	33
4.7.1.	Distribución discreta	34
4.7.2.	Distribución continua	34
4.8.	Conjunto de continuidad de una función de distribución en \mathbb{R}^k	34
5.	Variable Aleatoria	35
5.1.	Función inversa. Función compuesta	35
5.2.	Notación. Variable aleatoria (de una dimensión)	37
5.3.	Distribución de una variable aleatoria	38
5.4.	Transformación o cambio de variable aleatoria. Funciones medibles Borel	39
5.5.	Cambio de variable aleatoria en distribuciones continuas	40
5.6.	Variables aleatorias con valores infinitos	41
6.	Variables aleatorias de k dimensiones	43
6.1.	Variable aleatoria de k dimensiones	43
6.2.	Distribución de una variable aleatoria de k dimensiones	43
6.3.	Transformación de variables aleatorias. Funciones medibles Borel	44
6.4.	Cambio de variable aleatoria en el caso continuo	45
6.5.	Independencia de variables aleatorias	46
6.5.1.	Caso de independencia	49

7. Distribuciones Marginales Y Condicionadas	50
7.1. Distribuciones marginales. Casos discreto y continuo	50
7.1.1. Distribuciones marginales en el caso discreto	50
7.1.2. Distribuciones marginales en el caso continuo	51
7.2. Distribuciones condicionadas en el caso discreto y en el caso continuo	51
7.2.1. Caso discreto	51
7.2.2. Distribuciones condicionadas en el caso continuo	52
7.3. Caso de independencia	52
7.3.1. Caso discreto	52
7.3.2. Caso continuo	53
8. Distribuciones condicionadas	54
8.1. Definición general de las distribuciones condicionadas	54
8.2. Casos particulares. Caso de independencia	55
8.2.1. Distribuciones discretas	55
8.2.2. Distribuciones continuas	56
8.3. Convolución	57
9. Funciones Beta y Gamma	58
9.1. Función Gamma	58
9.2. Función Beta	58
9.3. Relación entre las funciones gamma y beta	59
10. Esperanza Matemática	61
10.1. Variables aleatorias simples	61
10.2. Esperanza matemática de variables aleatorias simples	62
10.3. Propiedades de la esperanza matemática de variables aleatorias simples	63
10.4. Variables aleatorias no negativas	66
10.5. Propiedades de la esperanza matemática para variables aleatorias no negativas	69
10.6. Esperanza matemática de variables aleatorias con valores reales	71
10.7. Propiedades de la esperanza matemática	72
10.8. Variables casi seguramente (c. s.) iguales	75
11. Esperanza Matemática (continuación)	77
11.1. Variables aleatorias con valores complejos	77
11.2. Teoremas de límite	79
11.3. Integral de Lebesgue abstracta (material complementario)	80
11.4. Casos particulares	82
11.4.1. Esperanza matemática	82
11.5. Medida inducida por una transformación del espacio	82
11.6. Medidas en un mismo espacio	84
11.6.1. Ejemplo	85

12.Momentos (1 dimensión)	86
12.1. Momentos	86
12.2. Existencia de los momentos. Relación entre los momentos ordinarios y los momentos centrales	87
12.3. Desigualdad de Tchebychev	88
12.4. Momentos de las distribuciones simétricas	89
12.5. Otras características de las distribuciones	90
II Ampliación de Probabilidad y Procesos Estocásticos	93
13.Momentos	93
13.1. Momentos de distribuciones de varias dimensiones	93
13.2. Covarianza y coeficientes de correlación. Desigualdad de Schwarz	94
13.3. Curvas de regresión	97
13.4. Momentos de sumas	99
14.Funciones generatrices	100
14.1. Funciones generatrices. Introducción	100
14.2. Función generatriz de probabilidad	100
14.3. Función generatriz de momentos	101
14.4. Función generatriz de momentos factoriales	102
14.5. Función generatriz de suma de variables aleatorias independientes	103
15.Función Característica	104
15.1. Función característica. Primeras propiedades	104
15.2. Teorema de Inversión	106
15.3. Función característica integrable	108
15.4. Derivabilidad de la función característica	110
15.5. Distribuciones simétricas	111
16.Funciones generatrices (varias dimensiones)	113
16.1. Función generatriz de probabilidad de varias dimensiones	113
16.2. Función generatriz de momentos de variables aleatorias de varias dimensiones	114
16.3. Función característica	115
17.Algunas distribuciones notables	117
17.1. Distribución multinomial	117
17.2. Distribución normal	118
17.3. Distribución B (beta)	122
17.4. Distribución Γ (gamma)	123
17.5. Distribución χ^2 (ji cuadrado)	126
17.6. Distribución \mathcal{F} de Snédecor	127
17.7. Distribución t de Student	129

Parte I

Teoría de la Probabilidad

1. Revisión de conceptos básicos

1.1. Definición de Espacio de Probabilidad

Supongamos dado un conjunto Ω (**espacio muestral**) cuyos elementos (**puntos muestrales**) quedan indeterminados; el espacio Ω se supone no vacío. Asimismo supondremos fijada una familia \mathcal{S} (**familia de sucesos**) de subconjuntos (**sucesos**) de Ω . Además, suponemos dada una función $P : \mathcal{S} \rightarrow \mathbb{R}$.

Todos estos elementos cumplirán seis axiomas, los axiomas S1, S2 y S3 para \mathcal{S} , que hacen que \mathcal{S} sea una σ -álgebra:

- **S1:** $\Omega \in \mathcal{S}$
- **S2:** $A \in \mathcal{S} \implies A^c \in \mathcal{S}$
- **S3:** $A_n \in \mathcal{S}, n = 1, 2, 3, \dots \implies \cup_{n=1}^{\infty} A_n \in \mathcal{S}$

y los axiomas P1, P2 y P3 para P , que hacen que P sea una **función de probabilidad**:

- **P1:** $P(\Omega) = 1$
- **P2:** $A \in \mathcal{S} \implies P(A) \geq 0$
- **P3:** $A_n \in \mathcal{S}, n = 1, 2, 3, \dots$ con $A_i \cap A_j = \emptyset, \forall i \neq j \implies P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$

1.2. Propiedades de la familia de sucesos

Si \mathcal{S} verifica S1, S2 y S3, entonces verifica:

- $\emptyset \in \mathcal{S}$
- $A_i \in \mathcal{S}, 1 \leq i \leq n \implies \cup_{i=1}^n A_i \in \mathcal{S}$
- $A_i \in \mathcal{S}, 1 \leq i \leq n \implies \cap_{i=1}^n A_i \in \mathcal{S}$
- $A_i \in \mathcal{S}, i = 1, 2, \dots \implies \cup_{i=1}^{\infty} A_i \in \mathcal{S}$
- $A, B \in \mathcal{S} \implies A \setminus B \in \mathcal{S}, A \Delta B \in \mathcal{S}$

1.2.1. Límites de sucesiones de conjuntos

Dada una sucesión de conjuntos de Ω

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

podemos definir dos nuevos conjuntos en Ω que llamamos **límite inferior** y **límite superior** de la sucesión de conjuntos dada y vienen dados por:

$$\liminf A_n = \{w \in \Omega : \exists m | \forall n \geq m \implies w \in A_n\}$$

$$\limsup A_n = \{w \in \Omega : \forall m, \exists n > m | w \in A_n\}$$

Intuitivamente, el límite inferior son los elementos que pertenecen a todos los A_n a partir de uno de ellos en adelante. El límite superior está formado por aquellos elementos que pertenecen a infinitos A_n .

Cuando ambos límites coinciden se dice que existe el **límite** de A_n y se escribe

$$\lim A_n = \liminf A_n = \limsup A_n$$

Teorema 1.1. *Los límites inferior y superior de A_n pueden obtenerse mediante las fórmulas*

$$\liminf A_n = \bigcup_{n \geq 1} \bigcap_{p \geq n} A_p$$

$$\limsup A_n = \bigcap_{n \geq 1} \bigcup_{p \geq n} A_p$$

por tanto, si $A_n \in \mathcal{S}$, entonces $\liminf A_n, \limsup A_n \in \mathcal{S}$. Y si existe el límite, este también estará en \mathcal{S} .

Teorema 1.2. *Para cualquier sucesión de conjuntos A_n :*

$$\liminf A_n \subset \limsup A_n$$

$$(\liminf A_n)^c = \limsup A_n^c$$

$$(\limsup A_n)^c = \liminf A_n^c$$

1.2.2. Familias monótonas

Una sucesión A_n de conjuntos de Ω se llama **monótona creciente** si se cumple $A_n \subset A_{n+1} \forall n$, y se dice **monótona decreciente** si $A_n \supset A_{n+1} \forall n$.

Las sucesiones monótonas son convergentes:

Teorema 1.3. *Si la sucesión A_n es monótona creciente, entonces existe*

$$\lim A_n = \bigcup_n A_n$$

Si la sucesión A_n es monótona decreciente, entonces existe

$$\lim A_n = \bigcap_n A_n$$

Diremos que la familia de conjuntos $\{A_t : t \in C\} \subset \mathcal{P}(\Omega)$, donde C es un subconjunto de la recta real, es una **familia monótona creciente** si

$$t_1, t_2 \in C, t_1 < t_2 \implies A_{t_1} \subset A_{t_2}$$

Una familia de la misma forma es una **familia monótona decreciente** si

$$t_1, t_2 \in C, t_1 < t_2 \implies A_{t_1} \supset A_{t_2}$$

Teorema 1.4. *Supongamos que $\{A_t : t \in T\} \subset \mathcal{P}(\Omega)$, $T \subset \mathbb{R}$, es una familia de conjuntos monótona creciente y existe un intervalo $I = (a, b) \subset T$. Entonces se verifican las siguientes propiedades:*

1. para cualquier sucesión (x_n) de números reales tal que $a < x_n < x_{n+1}$, $\forall n = 1, 2, \dots$ y $\lim x_n = b$ se tiene

$$\cup_{t < b} A_t = \cup_{n=1}^{\infty} A_{x_n}$$

por tanto, si la familia $\{A_t : t \in T\}$ está contenida en una σ -álgebra \mathcal{S} , entonces $\cup_{t < b} A_t \in \mathcal{S}$

2. para cualquier sucesión (y_n) tal que $b > y_n > y_{n+1}$ para $n = 1, 2, \dots$ y $\lim y_n = a$ se tiene

$$\cap_{t > a} A_t = \cap_{n=1}^{\infty} A_{y_n}$$

por tanto, si la familia $\{A_t : t \in T\}$ está contenida en una σ -álgebra \mathcal{S} , entonces $\cap_{t > a} A_t \in \mathcal{S}$

Este teorema tiene su análogo para intervalos I con $a = \infty$ ó $b = \infty$, así como para familias monótonas decrecientes.

1.2.3. Familias engendradas por una familia

Dada una familia $\mathcal{C} \subset \mathcal{P}(\Omega)$ se llama **σ -álgebra engendrada por \mathcal{C}** a la mínima σ -álgebra¹ en Ω que contenga a la familia \mathcal{C} . Se representa por $\sigma(\mathcal{C})$.

Teorema 1.5. *Dadas dos familias $\mathcal{C}_1, \mathcal{C}_2 \subset \mathcal{P}(\Omega)$, para que $\sigma(\mathcal{C}_1) = \sigma(\mathcal{C}_2)$ es necesario y suficiente que $\mathcal{C}_1 \subset \sigma(\mathcal{C}_2)$ y $\mathcal{C}_2 \subset \sigma(\mathcal{C}_1)$*

Si (Ω, \mathcal{O}) es un espacio topológico con la familia de abiertos \mathcal{O} , llamaremos **familia de Borel en Ω** a la σ -álgebra engendrada por la familia \mathcal{O} .

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

Llamamos **conjuntos de Borel en \mathbb{R}** a los conjuntos de la σ -álgebra engendrada por la familia \mathcal{O}_1 de los conjuntos abiertos en \mathbb{R} . La familia de los conjuntos de Borel se denota

$$\mathcal{B}_1 = \sigma(\mathcal{O}_1)$$

A veces es conveniente considerar \mathcal{B}_1 engendrada por una familia distinta de \mathcal{O}_1 . Llamemos

$$\mathcal{F}_1 = \{I \subset \mathbb{R} : I = (-\infty, x], x \in \mathbb{R}\}$$

Como $(x, \infty) \in \mathcal{O}_1 \subset \mathcal{B}_1$ resulta que

$$(-\infty, x] = (x, \infty)^c \in \mathcal{B}_1 \implies \mathcal{F}_1 \subset \mathcal{B}_1$$

Por otra parte, los intervalos $(a, b] = (-\infty, b] \setminus (-\infty, a] \in \sigma(\mathcal{F}_1)$ y $(a, b) = \cup_n (a, b - \frac{1}{n}] \in \sigma(\mathcal{F}_1)$, y como cualquier conjunto abierto en \mathbb{R} es una unión contable de intervalos abiertos (a, b) resulta

$$\mathcal{O}_1 \subset \sigma(\mathcal{F}_1)$$

¹Es fácil ver que la intersección arbitraria de σ -álgebras es una σ -álgebra, lo cual permite escribir la menor σ -álgebra como la intersección de todas las σ -álgebras que contengan al conjunto, siendo la intersección no vacía porque Ω es una σ -álgebra.

Por lo que se verifican las condiciones del teorema 5 y tenemos

$$\sigma(\mathcal{F}_1) = \sigma(\mathcal{O}_1) = \mathcal{B}_1$$

1.2.5. σ -álgebra de Borel en \mathbb{R}^k

Llamamos **conjuntos de Borel en \mathbb{R}^k** a los conjuntos de la σ -álgebra engendrada por la familia \mathcal{O}_k de los conjuntos abiertos en \mathbb{R}^k . Representaremos a la familia de los conjuntos de Borel en \mathbb{R}^k con la notación

$$\mathcal{B}_k = \sigma(\mathcal{O}_k)$$

De forma análoga a lo que ocurre en una dimensión, si definimos

$$\mathcal{F}_k = \{(-\infty, x_1] \times \dots \times (-\infty, x_k] : x_i \in \mathbb{R}, \forall i = 1, \dots, k\} = \left\{ I \subset \mathbb{R}^k : I = \prod_{i=1}^k (-\infty, x_i], x_i \in \mathbb{R} \right\}$$

se tiene que

$$\sigma(\mathcal{F}_k) = \sigma(\mathcal{O}_k) = \mathcal{B}_k$$

1.3. Primeras propiedades de la función de probabilidad P

- $P : \mathcal{S} \rightarrow \mathbb{R}$ es una **función de conjunto** ya que su dominio es una familia de conjuntos
- P es **no negativa**, por el axioma P2
- P es **numerablemente aditiva ó σ -aditiva**, por el axioma P3

Teorema 1.6. P *satisface:*

1. El suceso vacío, \emptyset , es un **suceso nulo**, o sea, $P(\emptyset) = 0$
2. $A_j \in \mathcal{S}$, $j = 1, \dots, n$ con $A_i \cap A_j = \emptyset$, $\forall i \neq j$, entonces

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

o sea, P es **finitamente aditiva**.

3. $A, B \in \mathcal{S}$, $A \subset B$, entonces

$$P(B \setminus A) = P(B) - P(A) \quad (P \text{ es } \mathbf{sustractiva})$$

$$P(A) \leq P(B) \quad (P \text{ es } \mathbf{monótona})$$

Además, $\forall A \in \mathcal{S}$

$$P(A) \leq 1$$

$$P(A^c) = 1 - P(A)$$

1.3.1. Probabilidad de la unión

Teorema 1.7. Para dos sucesos cualesquiera A y B se verifica

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

Teorema 1.8. Fórmula de inclusión-exclusión

Sean A_1, \dots, A_n n sucesos cualesquiera. Entonces

$$P(\cup_1^n A_i) = \sum_i^n P(A_i) - \sum_1^n P(A_i \cap A_j) + \sum_1^n P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(\cap_1^n A_i)$$

donde los índices se mueven en los conjuntos pertinentes. En el primer sumando, $1 \leq i \leq n$, en el segundo $1 \leq i < j \leq n$, en el tercero $1 \leq i < j < k \leq n$, y así sucesivamente.

1.3.2. Subaditividad de la función de probabilidad

Teorema 1.9. P se dice que es **numerablemente subaditiva** porque para cualquier sucesión A_n de sucesos se tiene

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

Teorema 1.10. P se dice que es **finitamente subaditiva** porque para n sucesos A_j cualesquiera, $1 \leq j \leq n$, se verifica

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

Así, como P posee las dos propiedades anteriores, se dice que es **contablemente subaditiva**.

Teorema 1.11. La unión contable de sucesos nulos es un suceso nulo.
La intersección contable de sucesos casi seguros es un suceso casi seguro.

1.3.3. Propiedades de límite de la función P

Teorema 1.12. Sea A_n una sucesión de sucesos monótona creciente, entonces existe

$$\lim P(A_n) = P(\lim A_n) = P(\cup_n A_n)$$

y se dice que P es **continua inferiormente**.

Si la sucesión es monótona decreciente, entonces también existe

$$\lim P(A_n) = P(\lim A_n) = P(\cap_n A_n)$$

y se dice que P es **continua superiormente**.

Teorema 1.13. Si A_n es una sucesión cualquiera de sucesos se tiene

$$P(\liminf A_n) \leq \liminf P(A_n) \leq \limsup P(A_n) \leq P(\limsup A_n)$$

Teorema 1.14. Si A_n es una sucesión de sucesos convergente, o sea, que tiene límite, entonces existe

$$\lim P(A_n) = P(\lim A_n)$$

y se dice que P es **continua**.

1.3.4. Límites para familias monótonas

Teorema 1.15. Sea $\{A_t : t \in T\}$, $(a, b) \subset T \subset \mathbb{R}$ una familia de sucesos monótona creciente. Entonces

$$\lim_{t \uparrow b} P(A_t) = P(\cup_{t < b} A_t)$$

$$\lim_{t \downarrow a} P(A_t) = P(\cap_{t > a} A_t)$$

Y este teorema tiene sus correspondientes análogos para intervalos infinitos y para familias monótonas decrecientes.

1.4. Probabilidad condicionada

Sean A, B dos sucesos cualesquiera con la única condición de que $P(B) > 0$. La **probabilidad de A condicionada por B** es

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

De esta definición se deduce que

$$P(A \cap B) = P(B)P(A|B), \quad (P(B) > 0)$$

Cuando $P(A) = P(A|B)$ los sucesos A y B se dice que son **independientes entre sí**. Lo mismo se dice si $P(B) = 0$.

1.4.1. Propiedades

A veces se denota

$$P_B(A) = P(A|B)$$

Así, vemos que, fijado B , P_B es una función de dominio \mathcal{S} .

Teorema 1.16. *Dado el espacio de probabilidad (Ω, \mathcal{S}, P) y un suceso B con $P(B) > 0$, entonces $(\Omega, \mathcal{S}, P_B)$ también es un espacio de probabilidad.*

Como $(\Omega, \mathcal{S}, P_B)$ es un espacio de probabilidad, entonces, si C es un suceso con $P_B(C) > 0$ podemos aplicar la definición de probabilidad condicionada en este nuevo espacio de probabilidad condicionada, para obtener

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

y vemos como no es necesario utilizar probabilidades con sucesivas condiciones, sino que basta con una sola condición para expresar probabilidades condicionadas sucesivamente.

1.4.2. Teorema de la probabilidad compuesta

Teorema 1.17. Teorema de la probabilidad compuesta

Sean A_1, \dots, A_n n sucesos cualesquiera y supongamos que

$$P(A_1 \cap \dots \cap A_{n-1}) > 0$$

Entonces se tiene

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|\cap_1^2 A_i) \cdot \dots \cdot P(A_n|\cap_1^{n-1} A_i)$$

1.4.3. Teorema de la probabilidad total

En el cálculo de probabilidades se presentan situaciones en las que existen n sucesos incompatibles cuya unión es Ω , o sea

$$H_1, \dots, H_n; H_i \cap H_j = \emptyset, \forall i \neq j; \cup_j H_j = \Omega$$

que poseen probabilidades conocidas $P(H_j)$ y para cada suceso A resultan también conocidas las probabilidades condicionadas $P(A|H_j)$. En esta situación se puede calcular la probabilidad de A a través de las probabilidades anteriores.

Los sucesos H_j se denominan **hipótesis**, las probabilidades $P(H_j)$ son las **probabilidades a priori de las hipótesis**, la probabilidad condicionada $P(A|H_j)$ es la **probabilidad de A en la hipótesis H_j** y la probabilidad $P(H_j|A)$ es la **probabilidad a posteriori** de la hipótesis H_j , cuando se conoce que ha ocurrido A .

Teorema 1.18. Teorema de la probabilidad total

Sean H_j n sucesos incompatibles de probabilidades no nulas y cuya unión es Ω . Entonces, para cualquier suceso A se tiene

$$P(A) = \sum_{1 \leq j \leq n} P(H_j) P(A|H_j)$$

Observación 1.1. Nótese que no es necesaria la hipótesis de que $P(H_j) > 0$, porque suprimiendo sumandos nulos obtendríamos la misma fórmula para las hipótesis no nulas.

Nótese, también que la fórmula funciona de igual forma para un conjunto numerable de hipótesis.

1.4.4. Teorema de Bayes

La fórmula de Bayes nos da la probabilidad a posteriori de una hipótesis H_j cuando se conoce que ha ocurrido un suceso A .

Teorema 1.19. Teorema de Bayes

Sean H_j , $1 \leq j \leq n$ n sucesos incompatibles de probabilidades positivas y cuya unión es Ω y A un suceso cualquiera de probabilidad positiva. Entonces, para cualquier suceso H_r , se obtiene la **fórmula de Bayes**

$$P(H_r|A) = \frac{P(H_r) P(A|H_r)}{\sum_{1 \leq j \leq n} P(H_j) P(A|H_j)}$$

Observación 1.2. Como antes, en el Teorema de Bayes se puede suprimir la condición $P(H_j) > 0 \forall j$ tomando en la suma del denominador solo las hipótesis no nulas. Además, esta fórmula se puede utilizar también en el caso en que los H_j son un conjunto numerable.

Teorema 1.20. Fórmula de Bayes generalizada

Sean H_j , $1 \leq j \leq n$, n sucesos incompatibles cuya unión es Ω y A un suceso tal que $P(A \cap H_j) > 0$, $\forall j$. Entonces, para cualquier suceso C se tiene

$$P(C|A) = \frac{\sum_{1 \leq j \leq n} P(H_j) P(A|H_j) P(C|A \cap H_j)}{\sum_{1 \leq j \leq n} P(H_j) P(A|H_j)}$$

Observación 1.3. Esta segunda fórmula es una generalización de la primera. La primera se obtiene de la segunda simplemente tomando $C = H_r$. Y de nuevo, la fórmula es válida para un conjunto numerable de hipótesis H_j . Además, la condición $P(A \cap H_j) > 0 \forall j$ puede sustituirse por $P(A) > 0$ si en las sumas del numerador y el denominador tenemos el cuidado de tomar solo los conjuntos de índices j tales que $P(H_j \cap A) > 0$.

2. Independencia

2.1. Independencia de dos sucesos

Dos sucesos A, B se dice que son **estocásticamente independientes** si se verifica

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

El orden no importa, es decir, si A y B son independientes, también podemos decir que B y A son independientes.

La dependencia es la negación de la independencia, se dirá que A y B son **estocásticamente dependientes** si

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

Nótese que si $P(B) > 0$, la independencia de A y B es equivalente a la condición

$$P(A) = P(A|B)$$

Teorema 2.1. *Si A y B son dos sucesos independientes, también A y B^c son independientes.*

Teorema 2.2. *Cualquier suceso nulo o casi seguro es independiente de cualquier otro suceso.*

2.2. Independencia entre n sucesos

Diremos que los n sucesos A_1, \dots, A_n son **independientes por parejas** si cada par A_i, A_j son independientes entre sí, o sea, si

$$P(A_i \cap A_j) = P(A_i)P(A_j), \quad \forall 1 \leq i < j \leq n,$$

y que son **estocásticamente independientes** si

$$P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i), \quad \forall I \subset \{1, \dots, n\}.$$

Más minuciosamente, son independientes si para cada i_1, i_2, \dots, i_r con

$$1 \leq i_1 < i_2 < \dots < i_r \leq n, \quad 2 \leq r \leq n,$$

se tiene

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_r}) = P(A_{i_1})P(A_{i_2}) \cdot \dots \cdot P(A_{i_r}).$$

Por lo que el número de condiciones que han de verificarse es

$$\binom{n}{2} + \binom{n}{3} + \dots + \binom{n}{n} = 2^n - n - 1.$$

Es obvio que la independencia implica la independencia por parejas, pero el recíproco no es cierto. Cuando los conjuntos no sean independientes diremos que son **dependientes**.

Teorema 2.3. *La independencia de n sucesos no depende del orden con que se enuncien. Los sucesos de un subconjunto del conjunto de n sucesos independientes son también independientes.*

Teorema 2.4. *Si los n sucesos A_1, \dots, A_n son independientes y se toma*

$$I \subset \{1, \dots, n-1\}, \quad A = \bigcap_{i \in I} A_i$$

entonces los sucesos A y A_n son independientes.

Teorema 2.5. *Sean A_1, \dots, A_{n-1} $n-1$ sucesos independientes y A_n un suceso cualquiera. Si para todo $I \subset \{1, \dots, n-1\}$, $A = \bigcap_{i \in I} A_i$ los dos sucesos A y A_n son independientes, entonces los n sucesos A_1, \dots, A_{n-1}, A_n son independientes.*

Teorema 2.6. *Si los n sucesos A_1, \dots, A_n son independientes, entonces los n sucesos A_1, \dots, A_n^c son independientes.*

Más generalmente, cada uno puede sustituirse por su complementario y el conjunto seguiría siendo independiente.

2.3. Independencia entre infinitos sucesos. Lema de Borell-Cantelli

Los sucesos de una familia $\{A_t : t \in T\}$ donde T es un conjunto cualquiera de índices (puede ser infinito) son **independientes** si los sucesos de cada conjunto finito

$$A_{t_1}, \dots, A_{t_n}, \quad n \geq 2, \quad \{t_1, \dots, t_n\} \subset T$$

son independientes.

Así pues, los sucesos de una sucesión A_1, \dots, A_n, \dots son independientes si $\forall n \geq 2$ los n sucesos A_1, \dots, A_n son independientes.

La **dependencia** es, de nuevo, la negación de la independencia.

Lema 2.1. De Borel-Cantelli

Cualesquiera que sean los sucesos de la sucesión $\{A_i\}_{i \in \mathbb{N}}$,

1. Si la serie

$$\sum_n P(A_n) \text{ converge}$$

entonces

$$P(\limsup A_n) = 0.$$

2. Si la serie

$$\sum_n P(A_n) \text{ diverge}$$

y, además, los sucesos de la sucesión son independientes, entonces

$$P(\limsup A_n) = 1.$$

Demostración. Veamos ambas afirmaciones:

1. Se tiene que

$$P(\limsup A_n) = P(\bigcap_n \bigcup_{p \geq n} A_p) \leq P(\bigcup_{p \geq n} A_p) \leq \sum_{p \geq n} P(A_p)$$

donde el último miembro es el resto de la serie completa, que es convergente. Por tanto, este último miembro tiende a 0 cuando $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \sum_{p \geq n} P(A_p) = 0$$

de donde obtenemos que

$$P(\limsup A_n) = 0$$

2. Aquí vamos a usar la desigualdad

$$1 - x \leq e^{-x} \quad (*)$$

válida $\forall x \in \mathbb{R}$.

Teniendo en cuenta la independencia y aplicando la desigualdad anterior con $x = P(A_p)$ tenemos que, para $n < m$

$$P(\bigcap_{p=n}^m A_p^c) \stackrel{A_n \text{ indep}}{\implies} \stackrel{A_n^c \text{ indep}}{=} \prod_{p=n}^m P(A_p^c) = \prod_{p=n}^m (1 - P(A_p)) \leq \prod_{p=n}^m e^{-P(A_p)} = e^{-\sum_{p=n}^m P(A_p)}$$

Tomando límites cuando $m \rightarrow \infty$ se obtiene, por la divergencia de la serie

$$P(\bigcap_{p=n}^{\infty} A_p^c) = 0, \forall n \implies P(\bigcup_n \bigcap_{p=n}^{\infty} A_p^c) = 0 \implies 0 = P(\liminf A_p^c) = P((\limsup A_n)^c) \implies P(\limsup A_n) = 1$$

Nótese que este paso podemos hacerlo porque P es continua y $\bigcap_{p=n}^{\infty} A_p^c$ convergente.

Para deducir (*) tomamos $f(x) = (1-x)e^x$, derivamos $f'(x) = -e^x + (1-x)e^x = e^x(-1+1-x) = -xe^x$, por lo que f es creciente en $(-\infty, 0)$ y decreciente en $(0, \infty)$, y en 0 presenta un máximo que vale $f(0) = 1$. Entonces

$$(1-x)e^x \leq 1 \implies 1-x \leq e^{-x}$$

□

2.4. Independencia entre familias

Además de la independencia entre sucesos es importante el concepto de independencia entre familias de sucesos. Diremos que n familias de sucesos

$$\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$$

son **familias independientes** si cada conjunto de n sucesos

$$A_1, A_2, \dots, A_n, \quad A_i \in \mathcal{A}_i, \quad i = 1, \dots, n$$

son sucesos independientes.

Las familias \mathcal{A}_t , $t \in T$ son independientes si para cada conjunto finito $\{t_1, t_2, \dots, t_n\} \subset T$, $n \geq 2$, las familias $\mathcal{A}_{t_1}, \dots, \mathcal{A}_{t_n}$ son independientes.

Proposición 2.1. *Dadas $n-1$ familias independientes $\mathcal{A}_1, \dots, \mathcal{A}_{n-1}$ para probar la independencia de las n familias $\mathcal{A}_1, \dots, \mathcal{A}_{n-1}, \mathcal{A}_n$ es suficiente probar, por el 2.5, que cada dos sucesos A, B con*

$$A = \bigcap_{i \in I} A_i, \quad I \subset \{1, \dots, n-1\}, \quad A_i \in \mathcal{A}_i, \quad \forall i \in I$$

$$B \in \mathcal{A}_n$$

son independientes.

Demostración. Por el 2.5, la condición del enunciado demuestra la independencia de los n sucesos A_1, \dots, A_{n-1}, B y esto nos proporciona la independencia de las n familias, pues A y B son arbitrarios. \square

2.5. π -sistemas y d-sistemas

Una familia \mathcal{C} de conjuntos de Ω se llama un **π -sistema** en Ω si es cerrada por intersecciones finitas, o sea, si se cumple que

$$A, B \in \mathcal{C} \implies A \cap B \in \mathcal{C}$$

Una familia \mathcal{D} de conjuntos de Ω se llama un **d-sistema** en Ω si se cumple:

1. $\Omega \in \mathcal{D}$
2. $A, B \in \mathcal{D}, B \subset A \implies A \setminus B \in \mathcal{D}$
3. $A_n \in \mathcal{D}, A_n \subset A_{n+1}, n = 1, 2, \dots \implies \bigcup_n A_n \in \mathcal{D}$

Dada una familia \mathcal{C} de subconjuntos de Ω se llama **π -sistema engendrado** por \mathcal{C} en Ω a la mínima familia de conjuntos de Ω que contiene a \mathcal{C} y es un π -sistema. Denotaremos esta familia por $\pi(\mathcal{C})$.

Dada una familia \mathcal{C} de subconjuntos de Ω se llama **d-sistema engendrado** por \mathcal{C} en Ω al mínimo d-sistema que contiene a \mathcal{C} . Denotaremos esta familia por $d(\mathcal{C})$.

Es evidente que toda σ -álgebra es un π -sistema y un d-sistema. En sentido contrario tenemos el siguiente resultado:

Teorema 2.7. *Una familia \mathcal{C} de conjuntos de Ω que es a la vez un π -sistema y un d-sistema en Ω es también una σ -álgebra en Ω .*

Demostración. Basta con probar los axiomas:

- **S1:** \mathcal{C} d-sistema $\implies \Omega \in \mathcal{C}$
- **S2:** $\Omega \supset A \in \mathcal{C} \implies A^c = \Omega \setminus A \in \mathcal{C}$
- **S3:** $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{C} \implies B_n := \bigcup_{i=1}^n A_i = \left(\bigcap_{i=1}^n A_i^c\right)^c \in \mathcal{C} \xrightarrow{B_n \subset B_{n+1}} \bigcup_{n=1}^{\infty} A_n = \bigcup_n B_n \in \mathcal{C}$

□

Teorema 2.8. Teorema π - λ de Dynkin.

Sea \mathcal{C} cualquier π -sistema de conjuntos de Ω . El d-sistema que engendra \mathcal{C} y la σ -álgebra que engendra \mathcal{C} son iguales:

$$d(\mathcal{C}) = \sigma(\mathcal{C})$$

Si un d-sistema \mathcal{D} contiene a un π -sistema \mathcal{C} , también contiene a la σ -álgebra $\sigma(\mathcal{C})$.

Demostración. Comenzamos viendo que en general tenemos que $\sigma(\mathcal{C}) \supset d(\mathcal{C})$, ya que $\sigma(\mathcal{C})$ es un d-sistema que contiene a \mathcal{C} . Por lo tanto, la peculiaridad está en que si \mathcal{C} es un π -sistema entonces $\sigma(\mathcal{C}) \subset d(\mathcal{C})$. Además, la segunda afirmación es equivalente a la primera, ya que si $\sigma(\mathcal{C}) \subset d(\mathcal{C})$ entonces $\mathcal{C} \subset \mathcal{D} \implies \sigma(\mathcal{C}) \subset d(\mathcal{C}) \subset \mathcal{D}$ y si $\sigma(\mathcal{C}) \subset \mathcal{D}$ para todo $\mathcal{D} \supset \mathcal{C}$ entonces también lo estará para $d(\mathcal{D})$. Probaremos la inclusión que falta viendo que $d(\mathcal{C})$ también es un π -sistema y por tanto una σ -álgebra que contiene a \mathcal{C} .

Dado $A \in d(\mathcal{C})$, sea

$$\mathcal{D}_A := \{B \in d(\mathcal{C}) : B \cap A \in d(\mathcal{C})\} \subset d(\mathcal{C}).$$

Veamos que es un d-sistema:

1. $\Omega \in \mathcal{C} \subset d(\mathcal{C}) \wedge \Omega \cap A = A \in \mathcal{C} \subset d(\mathcal{C}) \implies \Omega \in \mathcal{D}_A$
2. $\forall B_1, B_2 \in \mathcal{D}_A, B_1 \subset B_2$

$$(B_1 \cap A) \subset (B_2 \cap A) \implies (B_2 \setminus B_1) \cap A = (B_2 \cap A) \setminus (B_1 \cap A) \in d(\mathcal{C}) \implies B_2 \setminus B_1 \in \mathcal{D}_A$$

3. $\forall \{B_n\}_{n \in \mathbb{N}} \in \mathcal{D}_A,$

$$B_n \uparrow B \implies (B_n \cap A) \uparrow (B \cap A) \in d(\mathcal{C}) \implies B \in \mathcal{D}_A$$

Si tomamos $A \in \mathcal{C} \subset d(\mathcal{C})$, por ser \mathcal{C} un π -sistema, es claro que $\mathcal{C} \subset \mathcal{D}_A$; y al ser \mathcal{D}_A un d-sistema, se tiene $d(\mathcal{C}) \subset \mathcal{D}_A$. Así, por la arbitrariedad de A , para todo $B \in d(\mathcal{C})$ y $A \in \mathcal{C}$ se da la condición

$$B \cap A \in d(\mathcal{C}). \tag{1}$$

Pero esto justamente significa que $\mathcal{C} \subset \mathcal{D}_B$ tomando cualquier $B \in d(\mathcal{C})$; y repitiendo el argumento tenemos $d(\mathcal{C}) \subset \mathcal{D}_B$, dándose pues la condición (1) para cualquier $A \in d(\mathcal{C})$. Y por tanto, $d(\mathcal{C})$ es un π -sistema. □

2.6. Aplicación a la independencia

Teorema 2.9. *Dados n π -sistemas independientes $\mathcal{C}_1, \dots, \mathcal{C}_n$ las n σ -álgebras que engendran $\sigma(\mathcal{C}_1), \dots, \sigma(\mathcal{C}_n)$ también son independientes.*

Demostración. Vamos a comenzar probando la independencia de las n familias

$$\sigma(\mathcal{C}_1), \mathcal{C}_2, \dots, \mathcal{C}_n$$

en las que solo se ha tomado una σ -álgebra.

Para comprobar la independencia de estas n familias, por la proposición 2.8, como $\mathcal{C}_2, \dots, \mathcal{C}_n$ son familias independientes, basta comprobar que si

$$I \subset \{2, \dots, n\}, A = \bigcap_{i \in I} A_i, A_i \in \mathcal{C}_i, \forall i \in I$$

entonces $\sigma(\mathcal{C}_1), \{A\}$ son independientes.

Para este propósito, definimos la familia

$$\mathcal{F} = \{C \in d(\mathcal{C}_1) : P(A \cap C) = P(A)P(C)\}$$

esta familia contiene a \mathcal{C}_1 , pues \mathcal{C}_1 es independiente de los demás π -sistemas, y es un d-sistema:

1. $P(\Omega \cap A) = P(A) = P(A) \cdot 1 = P(A)P(\Omega) \implies \Omega \in \mathcal{F}$
2. $C_1, C_2 \in \mathcal{F}, C_1 \subset C_2 \implies P((C_2 \setminus C_1) \cap A) = P((C_2 \cap A) \setminus (C_1 \cap A)) = P(C_2 \cap A) - P(C_1 \cap A) = P(C_2)P(A) - P(C_1)P(A) = (P(C_2) - P(C_1))P(A) = P(C_2 \setminus C_1)P(A) \implies C_2 \setminus C_1 \in \mathcal{F}$
3. $C_n \in \mathcal{F}, C_n \uparrow C \implies (C_n \cap A) \uparrow (C \cap A) \implies P(C \cap A) = \lim_n P(C_n \cap A) = \lim_n P(C_n)P(A) = P(C)P(A) \implies C \in \mathcal{F}$

Por lo tanto, se tiene que

$$\mathcal{F} = d(\mathcal{C}_1)$$

y por el teorema 2.10, tenemos que

$$\mathcal{F} = \sigma(\mathcal{C}_1)$$

lo que, por la definición de \mathcal{F} nos demuestra que $\sigma(\mathcal{C}_1), \{A\}$ son independientes, como queríamos ver.

Así, vemos como dados los n π -sistemas independientes, podemos cambiar uno de ellos por la σ -álgebra que engendra y el conjunto sigue siendo independiente. Ahora bien, esta σ -álgebra engendada es también un π -sistema, por lo que se puede reiterar este proceso hasta tomar todas las σ -álgebras engendradas, dándonos el resultado del teorema. \square

3. Distribuciones en \mathbb{R} y \mathbb{R}^k

3.1. Introducción. Sumas con un conjunto infinito de sumandos

Teorema 3.1. Sea T un conjunto infinito cualquiera y $a : T \rightarrow \mathbb{R}$ una función que toma valores reales no negativos ($a(t) \geq 0$ para todo $t \in T$). Si existe una constante K tal que para todo subconjunto finito $F \subset T$ se verifica que

$$\sum_{t \in F} a(t) \leq K$$

entonces el conjunto $D = \{t \in T : a(t) > 0\}$ es, a lo sumo, numerable.

3.2. Distribuciones discretas en un espacio cualquiera

Supondremos en esta sección que $w \in \Omega \implies \{w\} \in \mathcal{S}$.

Teorema 3.2. Sea (Ω, \mathcal{S}, P) un espacio de probabilidad y $C \in \mathcal{S}$. Si para cada $\omega \in C$ se tiene que $P(\{\omega\}) > 0$, entonces C es contable.

Una distribución de probabilidad (Ω, \mathcal{S}, P) diremos que es una **distribución discreta** si existe un conjunto contable D en el que está concentrada la probabilidad, $P(D) = 1$. En dicho caso, se puede definir la función $p : \Omega \rightarrow \mathbb{R}$ mediante

$$p(\omega) = \begin{cases} P(\{\omega\}) & \text{para } \omega \in D \\ 0 & \text{para } \omega \notin D \end{cases}$$

y la llamaremos **función puntual de probabilidad** en Ω (o **función de punto de probabilidad**).

La función $P : \mathcal{S} \rightarrow \mathbb{R}$ es una **función de conjunto**, ya que su dominio tiene como elementos conjuntos de Ω . Por el contrario, la función p que hemos definido antes diremos que es una **función de punto**, ya que su dominio es Ω .

La función $p : \Omega \rightarrow \mathbb{R}$ posee las siguientes propiedades:

- D es contable
- $p(w) \geq 0$, $\forall w \in D$, $p(w) = 0$, $\forall w \notin D$
- $\sum_{w \in D} p(w) = 1$

Será una **función puntual de probabilidad** cualquier función que verifica estas propiedades.

Está claro que p queda determinada por P , y el siguiente teorema nos asegura que P también queda determinada por p .

Teorema 3.3. Sea Ω un conjunto cualquiera, p una función puntual de probabilidad, \mathcal{S} una σ -álgebra en Ω que contiene los conjuntos unitarios contenidos en D y $P : \mathcal{S} \rightarrow \mathbb{R}$ definida por

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

para cada $A \in \mathcal{S}$. Entonces (Ω, \mathcal{S}, P) es una distribución de probabilidad discreta.

3.3. Distribuciones en \mathbb{R} y \mathbb{R}^k . Distribuciones discretas

Una distribución de probabilidad en la recta está formada por una terna $(\mathbb{R}, \mathcal{S}, P)$, donde \mathcal{S} es una σ -álgebra en \mathbb{R} y P una función de probabilidad de dominio \mathcal{S} .

Para formar un espacio de probabilidad en \mathbb{R} tomaremos la σ -álgebra de Borel \mathcal{B}_1 .

Distribuciones discretas en \mathbb{R} . Una distribución de probabilidad $P : \mathcal{B}_1 \rightarrow \mathbb{R}$ en la recta \mathbb{R} diremos que es una **distribución discreta** si existe un conjunto contable $D = \{x_1, x_2, \dots\} \subset \mathbb{R}$ en el que está concentrada la probabilidad, $P(D) = 1$. Esta definición no es más que un caso particular de la definición general de distribución discreta. Llamaremos **función puntual de probabilidad** de una distribución discreta real a la función

$$p : \mathbb{R} \rightarrow \mathbb{R}$$

$$p(x) = P(\{x\})$$

Que verifica

$$p(x) \geq 0, \quad \sum_{x \in \mathbb{R}} p(x) = 1$$

Y para cada conjunto B de Borel de la recta se tiene que $P(B) = \sum_{x \in B} p(x)$.

Distribuciones discretas en \mathbb{R}^k . Una distribución de probabilidad $P : \mathcal{B}_k \rightarrow \mathbb{R}^k$ en \mathbb{R}^k diremos que es una **distribución discreta** o de **tipo discreto** si hay un conjunto contable $D = \{a_1, a_2, \dots\} \subset \mathbb{R}^k$ en el que está concentrada la probabilidad, $P(D) = 1$.

3.4. La integral en las distribuciones continuas

Dado un conjunto $A \subset \Omega$, la función $I_A : \Omega \rightarrow \mathbb{R}$ definida por

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

se llama **función indicador** (o **función característica**) del conjunto A .

Sea $\mu : \mathcal{F} \rightarrow \mathbb{R}^2$ una medida y F una familia de funciones de Ω en \mathbb{R} que contiene a las funciones indicador $I_A : \Omega \rightarrow \mathbb{R}$ para todo $A \in \mathcal{F}$, cerrada por productos por constantes reales, por sumas finitas (con ciertas restricciones) y por límites monótonos. Sea $f \in F$. Llamamos “**integral de Lebesgue** de la función f relativa a la medida μ ” a

$$\int f \, d\mu = \int f(\omega) \, d\mu(\omega) = \int_{\Omega} f(\omega) \, d\mu(\omega)$$

que es un valor real. Sus propiedades son:

² $\mathbb{R} = \mathbb{R} \cup \{-\infty, +\infty\}$.

1. $\int I_A d\mu = \mu(A)$.
2. $\int cf d\mu = c \int f d\mu$ para todo $c \in \mathbb{R}$.
3. $\int (f_1 + f_2) d\mu = \int f_1 d\mu + \int f_2 d\mu$.
4. $\int f d\mu \geq 0$ si $f \geq 0$.
5. Si $f_n \geq 0$ y $f_n \uparrow f$ entonces $\int f_n d\mu \rightarrow \int f d\mu$.

Cuando $\Omega = \mathbb{R}^k$ y se toma la medida de Lebesgue $\lambda_k : \mathcal{B}_k \rightarrow \bar{\mathbb{R}}$ se obtiene la **integral de Lebesgue** en \mathbb{R}^k . Dicha medida es la extensión de la función contenido (longitud, área, volumen) de un intervalo de \mathbb{R}^k a una medida definida en la σ -álgebra de Borel en \mathbb{R}^k .

Nota: Información complementaria en el punto XIV.3 del libro.

3.5. Medidas absolutamente continuas

Supongamos que Ω es un espacio cualquiera y \mathcal{F} una σ -álgebra en Ω . Si se tienen dos medidas cualesquiera en Ω μ y λ , entonces se dice que μ es **absolutamente continua** respecto de λ si verifica

$$C \in \mathcal{F} \quad \lambda(C) = 0 \implies \mu(C) = 0$$

Teorema 3.4. Teorema de Radon-Nikodym.

Sean $\mu, \lambda : \mathcal{F} \rightarrow \bar{\mathbb{R}}$ dos medidas con dominio la misma σ -álgebra \mathcal{F} en Ω y la medida λ finita (o σ -finita)^a. Si la medida μ es absolutamente continua respecto de λ entonces existe una función $f : \Omega \rightarrow \bar{\mathbb{R}}$ tal que para todo $A \in \mathcal{F}$, se tiene que

$$\mu(A) = \int_A f(\omega) d\lambda(\{\omega\})$$

^aUna medida λ es finita si existe una familia finita de conjuntos disjuntos $\{E_n\}_{n \in \{1, \dots, n\}}$ tal que $\Omega = \bigcup_{i=1}^n E_i$ y $\lambda(E_i) < \infty, \forall i \in \{1, \dots, n\}$. Es σ -finita si la familia anterior es numerable.

Vale una propiedad recíproca

Teorema 3.5. Dada una medida $\lambda : \mathcal{F} \rightarrow \bar{\mathbb{R}}$, si la medida $\mu : \mathcal{F} \rightarrow \bar{\mathbb{R}}$ se obtiene mediante la integral

$$\mu(A) = \int_A f(\omega) d\lambda(\{\omega\})$$

entonces μ es absolutamente continua respecto de λ .

Notemos que la noción de medida absolutamente continua siempre depende de otra medida.

3.6. Distribuciones continuas en \mathbb{R} y \mathbb{R}^k

Llamaremos **función de densidad** en \mathbb{R} a una función $f : \mathbb{R} \rightarrow \bar{\mathbb{R}}$ integrable Lebesgue si cumple las condiciones

$$f(x) \geq 0 \quad \int_{\mathbb{R}} f(x) dx = 1$$

Una distribución en la recta $(\mathbb{R}, \mathcal{B}_1, P)$ se dice **absolutamente continua** (respecto a la medida de Lebesgue en \mathbb{R}) o simplemente **continua** si existe una función de densidad tal que

$$P((a, b)) = \int_a^b f(x) dx = \int_{(a,b)} f(x) dx = \int_{\mathbb{R}} f(x) I_{(a,b)}(x) dx$$

que puede generalizarse para cualquier conjunto de Borel $B \in \mathcal{B}_1$

$$P(B) = \int_B f(x) dx = \int_{\mathbb{R}} f(x) I_B(x) dx$$

Lo dicho para las distribuciones en \mathbb{R} se generaliza para \mathbb{R}^k . Decimos que $(\mathbb{R}^k, \mathcal{B}_k, P)$ es una distribución *absolutamente continua* o de *tipo continuo* en \mathbb{R}^k si existe una función $f : \mathbb{R}^k \rightarrow \mathbb{R}$ tal que

$$f(x) \geq 0 \qquad \int_{\mathbb{R}^k} f(x) dx = 1$$

a la que llamamos *función de densidad* (de probabilidad) en \mathbb{R}^k , que para conjunto $A \in \mathcal{B}_k$ cumple

$$P(A) = \int_A f(x) dx = \int \dots \int_A f(x_1, \dots, x_k) dx_1 \dots dx_k.$$

3.7. Distribuciones uniformes

Una distribución $(\mathbb{R}, \mathcal{B}_1, P)$ de tipo continuo es **uniforme** en un conjunto $B \in \mathcal{B}_1$ si su función de densidad es constante en B y nula fuera de B . Es decir, para cierto $k \in \mathbb{R}$,

$$f(x) = k I_B(x).$$

Como la función de densidad f debe cumplir la condición

$$\int_{\mathbb{R}} f(x) dx = \int_{\mathbb{R}} k I_B(x) dx = k \int_B dx = k \lambda_1(B) = 1,$$

la constante k debe valer $k = \frac{1}{\lambda_1(B)}$, donde λ_1 es la medida de Lebesgue en la recta.

De modo más general una distribución $(\mathbb{R}^k, \mathcal{B}_k, P)$ de tipo continuo es *uniforme* en un conjunto $B \in \mathcal{B}_k$ si su función de densidad es constante en B y nula fuera de B . Con los mismos cálculos anteriores tenemos $f(x) = k I_B(x)$ donde la constante k viene dada por $k = \frac{1}{\lambda_k(B)}$, siendo λ_k la medida de Lebesgue en \mathbb{R}^k .

3.8. Distribuciones singulares

Sean $\mu : \mathcal{F} \rightarrow \overline{\mathbb{R}}$ y $\lambda : \mathcal{F} \rightarrow \overline{\mathbb{R}}$ dos medidas cualesquiera con dominio la misma σ -álgebra \mathcal{F} en Ω . La medida μ es una **medida singular** respecto de la medida λ si existe un conjunto $C \in \mathcal{F}$ tal que $\lambda(C) = 0$ y $\mu(C^c) = 0$.

Una distribución $(\mathbb{R}, \mathcal{B}_1, P)$ en la recta tal que para cada conjunto unitario $\{x\} \subset \mathbb{R}$ vale $P(\{x\}) = 0$ (no tiene componente discreta) diremos que es una **distribución singular** si existe un conjunto $C \in \mathcal{B}_1$, con $\lambda_1(C) = 0$ y $P(C) = 1$.

Una distribución $(\mathbb{R}^k, \mathcal{B}_k, P)$ tal que para cada conjunto unitario $\{x\} \subset \mathbb{R}^k$ vale $P(\{x\}) = 0$ diremos que es una *distribución singular* si existe un conjunto $C \in \mathcal{B}_k$, con $\lambda_k(C) = 0$ y $P(C) = 1$.

En \mathbb{R}^2 definir una distribución singular es tan fácil como tomar una que tenga la probabilidad concentrada en un segmento y que dentro de dicho conjunto sea de tipo continuo.

3.9. Descomposición de Lebesgue

En \mathbb{R}^k hay tres clases disjuntas de distribuciones: las discretas, las absolutamente continuas y las singulares. Cualquier otra distribución es una combinación lineal convexa de tres distribuciones, una de cada tipo.

Teorema 3.6. *Dada una distribución cualquiera $(\mathbb{R}^k, \mathcal{B}_k, P)$, existe un número $\alpha \in [0, 1]$ y dos distribuciones $(\mathbb{R}^k, \mathcal{B}_k, P_d)$ y $(\mathbb{R}^k, \mathcal{B}_k, P_c)$ tales que P_d es de tipo discreto, P_c tiene la propiedad $P_c(\{x\}) = 0, \forall x \in \mathbb{R}^k$, y se cumple para todo $B \in \mathcal{B}_k$ que*

$$P(B) = \alpha P_d(B) + (1 - \alpha) P_c(B).$$

Esta descomposición es única.

Demostración. Si P es de tipo discreto o $P(\{x\}) = 0, \forall x \in \mathbb{R}^k$ la demostración es trivial. Excluidos estos casos, sea

$$D := \left\{ x \in \mathbb{R}^k : P(\{x\}) > 0 \right\} = \{a_1, a_2, \dots\}$$

Este conjunto es contable (3.2) y vale forzosamente $\alpha := P(D) \in (0, 1)$, por los casos excluidos. Definimos las funciones

$$P_d(B) = \frac{1}{\alpha} P(B \cap D)$$

$$P_c(B) = \frac{1}{1 - \alpha} P(B \cap D^c).$$

1. P_d es una función de probabilidad de tipo discreto pues

- **P1:** $P_d(\Omega) = \frac{1}{\alpha} P(\Omega \cap D) = \frac{1}{\alpha} P(D) = \frac{1}{\alpha} \alpha = 1$,
- **P2:** $B \in \mathcal{B}_k \implies P_d(B) = \frac{1}{\alpha} P(B \cap D) \geq 0$,
- **P3:** $B_n \in \mathcal{B}_k, n = 1, 2, 3, \dots$ con $B_i \cap B_j = \emptyset, \forall i \neq j \implies$

$$P_d\left(\bigcup_{n=1}^{\infty} B_n\right) = \frac{1}{\alpha} P\left(\left(\bigcup_{n=1}^{\infty} B_n\right) \cap D\right) = \frac{1}{\alpha} P\left(\bigcup_{n=1}^{\infty} B_n \cap D\right) = \sum_{n=1}^{\infty} \frac{1}{\alpha} P(B_n \cap D) = \sum_{n=1}^{\infty} P_d(B_n), \text{ y}$$

- $P_d(D) = \frac{1}{\alpha} P(D \cap D) = \frac{1}{\alpha} P(D) = 1$;

2. P_c es una función de probabilidad con $P(\{x\}) = \frac{1}{1 - \alpha} P(\{x\} \cap D^c) = 0, \forall x \in \mathbb{R}^k$.

- **P1:** $P_c(\Omega) = \frac{1}{1 - \alpha} P(\Omega \cap D^c) = \frac{1}{1 - \alpha} P(D^c) = \frac{1}{1 - \alpha} (1 - \alpha) = 1$,
- **P2:** $B \in \mathcal{B}_k \implies P_c(B) = \frac{1}{1 - \alpha} P(B \cap D^c) \geq 0$,
- **P3:** $B_n \in \mathcal{B}_k, n = 1, 2, 3, \dots$ con $B_i \cap B_j = \emptyset, \forall i \neq j \implies$

$$P_c\left(\bigcup_{n=1}^{\infty} B_n\right) = \frac{1}{1 - \alpha} P\left(\left(\bigcup_{n=1}^{\infty} B_n\right) \cap D^c\right) = \sum_{n=1}^{\infty} \frac{1}{1 - \alpha} P(B_n \cap D^c) = \sum_{n=1}^{\infty} P_c(B_n).$$

Así, obtenemos la igualdad $\alpha P_d(B) + (1 - \alpha) P_c(B) = \alpha \frac{1}{\alpha} P(B \cap D) + (1 - \alpha) \frac{1}{1 - \alpha} P(B \cap D^c) = P(B)$.

Unicidad. Supongamos que existe otra descomposición

$$P(B) = \alpha' Q_d(B) + (1 - \alpha') Q_c(B)$$

con D' contable, $Q_d(D') = 1$ y $Q_c(\{x\}) = 0$ para todo $x \in \mathbb{R}^k$. Por la definición anterior del conjunto D , vale

$$\alpha' Q_d(\{x\}) = P(\{x\}) = \alpha P_d(\{x\}) \begin{cases} > 0 & \text{si } x \in D \\ = 0 & \text{si } x \notin D \end{cases}$$

y por tanto $D \subset D'$. Por lo tanto

$$1 = Q_d(D') = Q_d(D' \cap D) + Q_d(D' \cap D^c) = Q_d(D) + \sum_{x \in D' \setminus D} Q_d(\{x\}) = Q_d(D).$$

Así, también se tiene $\alpha = P(D) = \alpha' Q_d(D) = \alpha'$, de donde obtenemos finalmente las siguientes conclusiones

1. $\forall B \in \mathcal{B}_k, P_d(B) = \frac{1}{\alpha} P(B \cap D) \stackrel{Q_c(D) = \sum_{x \in D} Q_c(\{x\}) = 0}{=} \frac{1}{\alpha} \alpha' Q_d(B \cap D) \stackrel{Q_d(D) = 1}{=} Q_d(B)$
2. $\forall B \in \mathcal{B}_k, P(B) = \alpha P_d(B) + (1 - \alpha) P_c(B) = \alpha' Q_d(B) + (1 - \alpha') Q_c(B) \implies P_c(B) = Q_c(B)$.

□

Teorema 3.7. *Dada una distribución $(\mathbb{R}^k, \mathcal{B}_k, P_c)$, tal que $P_c(\{x\}) = 0, \forall x \in \mathbb{R}^k$, entonces existe una constante $\beta \in [0, 1]$ y dos distribuciones $(\mathbb{R}^k, \mathcal{B}_k, P_s)$ y $(\mathbb{R}^k, \mathcal{B}_k, P_{ac})$ tales que P_s es singular y P_{ac} es absolutamente continua y se cumple para todo $B \in \mathcal{B}_k$ que*

$$P_c(B) = \beta P_s(B) + (1 - \beta) P_{ac}(B).$$

Esta descomposición es única.

Demostración. De nuevo excluimos los casos triviales correspondientes a P_c singular o absolutamente continua. Llamemos

$$\beta := \sup \{P_c(A) \in \mathbb{R} : A \in \mathcal{B}_k, \lambda_k(A) = 0\}$$

Veamos que existe $C \in \mathcal{B}_k$ tal que $P_c(C) = \beta$ y $\lambda_k(C) = 0$. En efecto, por la definición de β existe una sucesión $\{A_n\}_n$ tal que

$$A_n \in \mathcal{B}_k, \lambda_k(A_n) = 0, P_c(A_n) \rightarrow \beta$$

y para $C = \bigcup_n A_n$ se tiene $\lambda_k(C) \leq \sum_n \lambda_k(A_n) = 0$ y $P_c(C) \geq P_c(A_n) \rightarrow \beta$, luego $P_c(C) = \beta$ (por ser el supremo).

Si $\beta = 0$, la distribución sería absolutamente continua (los conjuntos de medida nula tienen probabilidad nula) y si $\beta = 1$ sería singular (por definición). Así, tenemos $\beta \in (0, 1)$ y definiendo las funciones

$$P_s(B) = \frac{1}{\beta} P_c(B \cap C)$$

$$P_{ac}(B) = \frac{1}{1 - \beta} P_c(B \cap C^c).$$

Es claro que son dos medidas de probabilidad en \mathbb{R}^k (y si no lo tienes claro repites la demo anterior). Para ver que P_s es singular basta tener en cuenta que $\lambda_k(C) = 0$ y $P_s(C) = \frac{1}{\beta}P_c(C) = \frac{1}{\beta}\beta = 1$. Para ver que P_{ac} es absolutamente continua respecto de λ_k , tomemos $B \in \mathcal{B}_k$ con $\lambda_k(B) = 0$, si no valiese $P_{ac}(B) = 0$ tendríamos $P_c(B \cap C^c) > 0$ y el conjunto $C' := C \cup (B \cap C^c)$ cumpliría $\lambda_k(C') = 0$ y $P_c(C') = P_c(C) + P_c(B \cap C^c) > \beta$, incumpliendo la definición de β (supremo). Análogamente a la demostración anterior, es claro que $P_c(B) = \beta P_s(B) + (1 - \beta) P_{ac}(B)$.

La unicidad no entra en el examen de 2020. □

Teorema 3.8. *Dada una distribución cualquiera $(\mathbb{R}^k, \mathcal{B}_k, P)$, existen tres números $\{\alpha_i\}_{i \in \{1,2,3\}}$ tal que*

$$\alpha_i \geq 0 \text{ y } \sum_i \alpha_i = 1$$

y tres distribuciones $(\mathbb{R}^k, \mathcal{B}_k, P_d)$, $(\mathbb{R}^k, \mathcal{B}_k, P_s)$ y $(\mathbb{R}^k, \mathcal{B}_k, P_{ac})$ tales que P_d es discreta, P_s singular y P_{ac} absolutamente continua tales que para todo $B \in \mathcal{B}_k$, se tiene

$$P(B) = \alpha_1 P_d(B) + \alpha_2 P_s(B) + \alpha_3 P_{ac}(B).$$

Esta descomposición es única.

Demostración. Basta con aplicar el los dos teoremas anteriores y llamar $\alpha_1 = \alpha$, $\alpha_2 = (1 - \alpha)\beta$ y $\alpha_3 = (1 - \alpha)(1 - \beta)$, pues

$$\begin{aligned} P(B) &= \alpha P_d(B) + (1 - \alpha) P_c(B) = \alpha P_d(B) + (1 - \alpha) (\beta P_s(B) + (1 - \beta) P_{ac}(B)) = \\ &= \alpha P_d(B) + (1 - \alpha) \beta P_s(B) + (1 - \alpha)(1 - \beta) P_{ac}(B). \end{aligned}$$

□

4. Función de distribución

4.1. Funciones monótonas. Notaciones. Propiedades

Dada una función real $G : \mathbb{R} \rightarrow \mathbb{R}$, el **límite por la derecha** en un punto $x \in \mathbb{R}$ y el **límite por la izquierda** se denotan

$$G(x+) = \lim_{t \downarrow x} G(t) \quad G(x-) = \lim_{t \uparrow x} G(t)$$

y de la misma forma puede denotarse (abusando de notación) los límites hacia el $-\infty$ y $+\infty$. Nótese que estas nociones suponen la existencia de estos límites.

La función G será **continua por la derecha en un punto** $x \in \mathbb{R}$ si $G(x+) = G(x)$ y será **continua por la derecha en un conjunto** $C \subset \mathbb{R}$ si lo es $\forall x \in C$. Si G es continua por la derecha en \mathbb{R} se dirá que G es **continua por la derecha**.

La definición de **continuidad por la izquierda** es completamente análoga.

Una función real G es **monótona creciente** si

$$x_1 < x_2 \implies G(x_1) \leq G(x_2)$$

Para una función real G monótona creciente los dos límites laterales siempre existen y se verifica

$$G(x-) \leq G(x) \leq G(x+), \quad \forall x \in \mathbb{R}$$

Denotamos por $Cont G$ el **conjunto de puntos de continuidad** de G . El complementario de este conjunto se denomina **conjunto de puntos de discontinuidad** de G .

De esta definición, se tiene que, si G es monótona creciente, entonces

$$Cont G = \{x : G(x-) = G(x+)\}$$

Teorema 4.1. *Si $G : \mathbb{R} \rightarrow \mathbb{R}$ es una función monótona creciente el conjunto de sus puntos de discontinuidad es, a lo más, numerable.*

4.2. Definición de función de distribución de una dimensión. Función de distribución de una función de probabilidad en \mathbb{R}

Llamaremos **función de distribución** en \mathbb{R} a una función $F : \mathbb{R} \rightarrow \mathbb{R}$ que verifique:

1. F es monótona creciente
2. F es continua por la derecha
3. $F(-\infty) = 0$, $F(+\infty) = 1$

Entre las funciones de probabilidad $P : \mathcal{B}_1 \rightarrow \mathbb{R}$ y las funciones de distribución $F : \mathbb{R} \rightarrow \mathbb{R}$ existe una relación como vamos a ver.

Teorema 4.2. *Dada una función de probabilidad $P : \mathcal{B}_1 \rightarrow \mathbb{R}$ la función $F : \mathbb{R} \rightarrow \mathbb{R}$ definida por*

$$F(x) = P((-\infty, x]), \quad \forall x \in \mathbb{R}$$

*es una función de distribución, denominada **función de distribución de la función de probabilidad P** .*

Teorema 4.3. Si $F(x) = P((-\infty, x])$ es la función de distribución de la función de probabilidad P , entonces se verifica

$$P((-\infty, x)) = F(x-)$$

además, para las probabilidades de los intervalos finitos vale

$$P((a, b]) = F(b) - F(a), \quad a \leq b$$

$$P([a, b]) = F(b) - F(a-), \quad a \leq b$$

$$P((a, b)) = F(b-) - F(a), \quad a < b$$

$$P([a, b)) = F(b-) - F(a-), \quad a \leq b$$

$$P(\{a\}) = F(a) - F(a-)$$

Teorema 4.4. Dada una función de distribución en \mathbb{R} , $F : \mathbb{R} \rightarrow \mathbb{R}$, existe una función de probabilidad $P : \mathcal{B}_1 \rightarrow \mathbb{R}$, de la cual F es función de distribución. Es decir

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

Teorema 4.5. Dada una distribución F , la función de probabilidad P , de la cual F es función de distribución es única.

Por tanto, la correspondencia entre las funciones de probabilidad P y las funciones de distribución F es una biyección.

4.2.1. Distribuciones discretas

Para una distribución discreta con función puntual de probabilidad p la función de distribución vale

$$F(x) = \sum_{t \leq x} p(t)$$

4.2.2. Distribuciones continuas

Para una distribución absolutamente continua con función de densidad f la función de distribución vale

$$F(x) = \int_{-\infty}^x f(t) dt$$

y si x es un punto de continuidad de f se tiene para la derivada de F , $F'(x) = f(x)$.

4.3. Ejemplo de distribución singular en \mathbb{R}

La función de distribución F de una distribución singular es una función continua, pues $P(\{x\}) = F(x) - F(x-)$. Pero esto no quiere decir que la distribución sea continua.

Vamos a ver el **ejemplo clásico de distribución singular en \mathbb{R}** . Definimos la función de distribución $F : \mathbb{R} \rightarrow \mathbb{R}$ como

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases}$$

y se completará la función en $(0, 1)$ como sigue.

Para cada $a \in (0, 1)$ utilizamos el desarrollo ternario

$$a = \sum_{j=1}^{\infty} \frac{a_j}{3^j} = \frac{a_1}{3} + \frac{a_2}{9} + \dots, \quad a_j \in \{0, 1, 2\}$$

Este desarrollo no es único cuando a admite un desarrollo finito, pues en este caso también admite un desarrollo infinito (al igual que podemos hacer $1 = 0,99999\dots$). En estos casos podemos elegir siempre el desarrollo finito, de modo que para estos números la expresión anterior se reduce a

$$a = \sum_{j=1}^n \frac{a_j}{3^j}, \quad a_j \in \{0, 1, 2\}, \quad a_n \neq 0$$

y se eliminan los desarrollos que tienen todos los numeradores $a_j = 2$ desde un lugar en adelante. Así, todo $a \in (0, 1)$ tiene un desarrollo único.

Así, completamos la definición de F :

- **(a)** si algún $a_j = 1$ y a_n es el primero que vale 1 pondremos

$$F(a) = a' = \sum_{j=1}^{n-1} \frac{a'_j}{2^j} + \frac{1}{2^n}, \quad a'_j = \frac{a_j}{2}$$

- **(b)** si todo $a_j \neq 1$ pondremos

$$F(a) = a' = \sum_{j=1}^{\infty} \frac{a'_j}{2^j}, \quad a'_j = \frac{a_j}{2}$$

Queda, por tanto, completada la definición de F . Que posee las propiedades:

1. Es monótona creciente

Sea $a < b$, $a = \sum_{j=1}^{\infty} \frac{a_j}{3^j}$, $b = \sum_{j=1}^{\infty} \frac{b_j}{3^j}$, si $a_j = b_j \neq 1$, $\forall j = 1, 2, \dots, n-1$ y $a_n = b_n = 1$, entonces $F(a) = F(b)$ por (a).

Si $a_j = b_j$ para $j = 1, 2, \dots, n-1$ y la primera diferencia está en $a_n < b_n$, se distinguen tres casos:

a) $a_n = 0$, $b_n = 1 \implies$

$$F(a) = \sum_{j=1}^{\infty} \frac{a_j}{2^{j+1}} = \sum_{j=1}^{n-1} \frac{a_j}{2^{j+1}} + \sum_{j=n}^{\infty} \frac{a_j}{2^{j+1}} = \sum_{j=1}^{n-1} \frac{b_j}{2^{j+1}} + \sum_{j=n+1}^{\infty} \frac{a_j}{2^{j+1}} \leq \sum_{j=1}^{n-1} \frac{b_j}{2^{j+1}} + \frac{1}{2^n} = F(b)$$

b) $a_n = 0$, $b_n = 2 \implies$

$$F(a) = \sum_{j=1}^{\infty} \frac{a_j}{2^{j+1}} = \sum_{j=1}^{n-1} \frac{a_j}{2^{j+1}} + \sum_{j=n}^{\infty} \frac{a_j}{2^{j+1}} = \sum_{j=1}^{n-1} \frac{b_j}{2^{j+1}} + \sum_{j=n+1}^{\infty} \frac{a_j}{2^{j+1}} \leq \sum_{j=1}^{n-1} \frac{b_j}{2^{j+1}} + \frac{1}{2^n} + \dots = F(b)$$

c) $a_n = 1, b_n = 2 \implies$

$$F(a) = \sum_{j=1}^{n-1} \frac{a_j}{2^{j+1}} + \frac{1}{2^n} = \sum_{j=1}^{n-1} \frac{b_j}{2^{j+1}} + \frac{1}{2^n} \leq \sum_{j=1}^{n-1} \frac{b_j}{2^{j+1}} + \frac{1}{2^n} + \sum_{j=n+1}^{\infty} \frac{b_j}{2^{j+1}} = F(b)$$

2. Es continua

Para comprobar esto, basta tener en cuenta que cualquier punto $a' \in (0, 1)$ puede expresarse como un desarrollo de la forma que tiene F en (a) o en (b), por lo que existe a tal que $F(a) = a'$, por lo que el intervalo $(0, 1)$ se transforma monótonamente en el $(0, 1)$ por F , lo que prueba que F es continua.

3. F es constante en los intervalos

$$J_{mi} = \left(\sum_{j=1}^{m-1} \frac{a_j}{3^j} + \frac{1}{3^m}, \sum_{j=1}^{m-1} \frac{a_j}{3^j} + \frac{2}{3^m} \right), \quad a_j \in \{0, 2\}$$

y la suma de las longitudes de estos intervalos disjuntos es

$$\frac{1}{3} + 2 \frac{1}{3^2} + 2^2 \frac{1}{3^3} + \dots + 2^{m-1} \frac{1}{3^m} + \dots = \frac{1}{3} \frac{1}{1 - \frac{2}{3}} = 1$$

luego toda la probabilidad está concentrada en el conjunto

$$C = (0, 1) \setminus \bigcup_{m,i} J_{mi}$$

que tiene medida de Lebesgue cero, por lo que

$$\lambda_1(C) = 0, \quad P(C) = 1$$

y tenemos que F es una función de distribución singular.

4.4. Incremento mixto y continuidad lateral de funciones de dos variables

4.4.1. Operador diferencia

Dada una función real G y los dos valores a, b se puede definir el **incremento** de $G(x)$ respecto de la variable x como $G(b) - G(a)$ y se escribe

$$\Delta_{x=a}^b G(x) = G(b) - G(a)$$

El símbolo $\Delta_{x=a}^b$ se denomina **operador diferencia** o **incremento** relativo a la variable x .

4.4.2. Incremento mixto

Si $G : \mathbb{R}^2 \rightarrow \mathbb{R}$, esta operación se puede aplicar a cada variable:

$$\Delta_{x=a_1}^{b_1} G(x, y) = G(b_1, y) - G(a_1, y)$$

y como esta es función de la variable y se puede volver a aplicar el operador diferencia a esta variable, y se tiene

$$\begin{aligned}\Delta_{y=a_2}^{b_2} \left(\Delta_{x=a_1}^{b_1} G(x, y) \right) &= \Delta_{y=a_2}^{b_2} \Delta_{x=a_1}^{b_1} G(x, y) = \Delta_{y=a_2}^{b_2} (G(b_1, y) - G(a_1, y)) = \\ &= G(b_1, b_2) - G(a_1, b_2) - G(b_1, a_2) + G(a_1, a_2)\end{aligned}$$

Este es el **incremento reiterado** respecto a cada una de estas dos variables. Esta operación se suele representar escribiendo

$$\Delta_{(a_1, a_2)}^{(b_1, b_2)}$$

Nótese que permutando el orden en que se toman los incrementos el resultado no varía.

El resultado del incremento reiterado viene dado por la fórmula

$$\Delta_{(a_1, a_2)}^{(b_1, b_2)} G(x, y) = \sum_{z_i \in \{a_i, b_i\}} (-1)^{\sigma(z_1, z_2)} G(z_1, z_2)$$

donde $\sigma(z_1, z_2)$ indica el número de igualdades $z_i = a_i$ que son ciertas al tomar $i = 1, 2$.

4.4.3. Límites laterales en el plano

Dados $a, b \in \mathbb{R}^2$ se denota $a \leq b$ para indicar que $a_1 \leq b_1$ y $a_2 \leq b_2$.

El **límite lateral por la derecha** de $G : \mathbb{R}^2 \rightarrow \mathbb{R}$ en sus variables lo representamos por

$$G(x+) = G(x_1+, x_2+) = \lim_{h_1, h_2 \downarrow 0} G(x_1 + h_1, x_2 + h_2)$$

y para el **límite lateral por la izquierda** se utiliza la notación análoga $G(x-)$.

G será **continua por la derecha en un punto** x si $G(x+) = G(x)$.

G se dirá **continua** si es continua en x , $\forall x \in \mathbb{R}^2$.

Los **límites por la derecha en cada variable** son

$$G(x_1+, x_2), G(x_1, x_2+)$$

y su obtención es obvia.

La **continuidad lateral para una variable** se da cuando $G(x_1+, x_2) = G(x_1, x_2)$.

Proposición 4.1. *Si $G : \mathbb{R}^2 \rightarrow \mathbb{R}$ es continua por la derecha en sus dos variables en \mathbb{R}^2 entonces G es continua por la derecha en cada una de las variables.*

Como en una dimensión, podemos tomar límites hacia el infinito y abusar un poco de notación cuando estos existen.

4.5. Función de distribución en \mathbb{R}^2 . Función de distribución de una distribución de probabilidad en \mathbb{R}^2

Una **función de distribución** de dos dimensiones es una función $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ que cumple:

1. $\Delta_a^b F(x_1, x_2) \geq 0$, $\forall a, b \in \mathbb{R}^2$, $a < b$

2. $F(x_1+, x_2+) = F(x_1, x_2)$, $\forall x = (x_1, x_2) \in \mathbb{R}^2$
3. $F(-\infty, x_2) = F(x_1, -\infty) = F(-\infty, -\infty) = 0$, $F(\infty, \infty) = 1$

Si $a < b$ el intervalo $(a, b]$ determina los dos puntos a y b y recíprocamente. Por ello el incremento mixto de la condición primera también se puede escribir

$$\Delta_{(a,b]}F \geq 0, \forall (a, b] \subset \mathbb{R}^2$$

Teorema 4.6. Dada la función de probabilidad en \mathbb{R}^2 , $P : \mathcal{B}_2 \rightarrow \mathbb{R}$ la función $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ definida por

$$F(x, y) = P((-\infty, x] \times (-\infty, y])$$

es una función de distribución de dos dimensiones.

Demostración. Debemos comprobar las tres propiedades que definen una función de distribución.

1. Sean $x < y$ con $x, y \in \mathbb{R}^2$,

$$\begin{aligned} \Delta_a^b F(x, y) &= \Delta_{y=a_2}^{b_2} \left(\Delta_{x=a_1}^{b_1} F(x, y) \right) = \Delta_{y=a_2}^{b_2} (F(b_1, y) - F(a_1, y)) = * \\ &= F(b_1, y) - F(a_1, y) = P((-\infty, b_1] \times (-\infty, y]) - P((-\infty, a_1] \times (-\infty, y]) = \\ &= P((-\infty, b_1] \times (-\infty, y] - (-\infty, a_1] \times (-\infty, y]) = P((a_1, b_1] \times (-\infty, y]) \\ &= \Delta_{y=a_2}^{b_2} (P((a_1, b_1] \times (-\infty, y])) = P((a_1, b_1] \times (-\infty, b_2]) - P((a_1, b_1] \times (-\infty, a_2]) = \\ &= P((a_1, b_1] \times (-\infty, b_2] - (a_1, b_1] \times (-\infty, a_2]) = P((a_1, b_1] \times (a_2, b_2]) \geq 0 \end{aligned}$$

2. Sea $(x_1, x_2) \in \mathbb{R}^2$, $h, k \in \mathbb{R}^+$ y $t = \max\{h, k\}$. Por ser $(x_1, x_2) < (x_1 + h, x_2 + k) < (x_1 + t, x_2 + t)$, se tiene por inclusión de los correspondientes sucesos,

$$F(x_1, x_2) \leq F(x_1 + h, x_2 + k) \leq F(x_1 + t, x_2 + t).$$

Tomando límites y teniendo en cuenta, por 1.15, que

$$\lim_{t \rightarrow 0} P((-\infty, x_1 + t] \times (-\infty, x_2 + t]) = P\left(\bigcap_{t > 0} (-\infty, x_1 + t] \times (-\infty, x_2 + t]\right) = P((-\infty, x_1] \times (-\infty, x_2])$$

obtenemos

$$F(x_1, x_2) \leq F(x_1+, x_2+) \leq F(x_1, x_2)$$

y por lo tanto F es continua por la derecha.

3. Utilizando de nuevo 1.15, obtenemos

$$\begin{aligned} &F(-\infty, x_2) &&= F(x_1, -\infty) &&= \\ &= P\left(\bigcap_{x_1 \in \mathbb{R}} (-\infty, x_1] \times (-\infty, x_2]\right) &&= P\left(\bigcap_{x_2 \in \mathbb{R}} (-\infty, x_1] \times (-\infty, x_2]\right) &&= P(\emptyset) = 0 \end{aligned}$$

y haciendo un argumento similar al del punto anterior tenemos, llamando $t = \max\{x, y\}$ con $x, y \rightarrow -\infty$,

$$F(-\infty, -\infty) \leq \lim_{t \rightarrow -\infty} F(t, t) = P\left(\bigcap_{t \in \mathbb{R}} (-\infty, t] \times (-\infty, t]\right) = P(\emptyset) = 0;$$

análogamente, tomando $t = \min\{x, y\}$,

$$F(+\infty, +\infty) \geq \lim_{t \rightarrow \infty} F(t, t) = P\left(\bigcup_{t \in \mathbb{R}} (-\infty, t] \times (-\infty, t]\right) = P(\mathbb{R}^2) = 0;$$

□

Teorema 4.7. *Existe una biyección entre las funciones de probabilidad P en \mathbb{R}^2 y las funciones de distribución F de dos dimensiones.*

4.5.1. Distribuciones discretas

Para una distribución discreta con función puntual de probabilidad $p : \mathbb{R}^2 \rightarrow \mathbb{R}$ la función de distribución vale

$$F(x, y) = \sum_{(u, v) \leq (x, y)} p(u, v)$$

4.5.2. Distribuciones continuas

Para una distribución continua con función de densidad $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ la función de distribución vale

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, dudv$$

En este caso, si (x, y) es un punto de continuidad de f se tiene para la derivada de la función de distribución

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

4.6. Funciones con varias variables. Incremento mixto. Continuidad lateral

Los resultados de la sección 4 se generalizan para una función $G(x_1, \dots, x_k)$ en \mathbb{R}^k .

Así, se llama **incremento mixto** respecto de todas las variables de G al incremento reiterado respecto de cada una de ellas:

$$\Delta_a^b G(x_1, \dots, x_k) = \Delta_{x_1=a_1}^{b_1} \Delta_{x_2=a_2}^{b_2} \cdots \Delta_{x_k=a_k}^{b_k} G(x_1, \dots, x_k)$$

Igual que antes, denotaremos $a \leq b$ para señalar que $a_i \leq b_i$, $\forall i = 1, \dots, k$. Los límites laterales simbolizan lo que esperaríamos naturalmente que simbolicen.

La función G será **continua por la derecha en un punto** $x \in \mathbb{R}^k$ si $G(x+) = G(x)$. Si G es continua por la derecha en todo $x \in \mathbb{R}^k$ diremos que G es **continua por la derecha**.

Proposición 4.2. *Si $G : \mathbb{R}^k \rightarrow \mathbb{R}$ es continua por la derecha en sus k variables entonces G es continua por la derecha en cada grupo parcial de sus variables*

4.7. Función de distribución en \mathbb{R}^k . Función de distribución de una distribución de probabilidad en \mathbb{R}^k

Una **función de distribución de k dimensiones** es una función $F : \mathbb{R}^k \rightarrow \mathbb{R}$ que cumple:

- $\Delta_a^b F(x_1, \dots, x_k) \geq 0, \forall a, b \in \mathbb{R}^k, a < b$
- $F(x_1+, \dots, x_k+) = F(x_1, \dots, x_k), \forall x = (x_1, \dots, x_k) \in \mathbb{R}^k$
- Para cada conjunto de una o varias variables $\{x_i, \dots, x_r\}$ se tiene

$$\lim_{x_i \rightarrow -\infty, \dots, x_r \rightarrow -\infty} F(x_1, \dots, x_k) = 0;$$

por otra parte,

$$\lim_{x_1 \rightarrow \infty, x_2 \rightarrow \infty, \dots, x_k \rightarrow \infty} F(x_1, \dots, x_k) = 1.$$

Si $a < b$ el intervalo $(a, b]$ determina los dos puntos a, b y recíprocamente. Por ello, el incremento mixto de la primera condición se puede escribir como

$$\Delta_{(a,b]} F(x) \geq 0, \forall (a, b] \subset \mathbb{R}^k$$

Proposición 4.3. *La función de distribución F es monótona en sus variables.*

Demostración. Basta comprobarlo para una variable. Si $a < b \leq b'$ con $b = (b_1, \dots, b_k), b' = (b'_1, b_2, \dots, b_k)$ y $b_1 < b'_1$ se tiene que

$$\begin{aligned} \Delta_a^{b'} F(x) &= \Delta_{x_1=a_1}^{b'_1} \Delta_{x_2=a_2}^{b_2} \cdots \Delta_{x_k=a_k}^{b_k} G(x_1, \dots, x_k) = \\ &= \Delta_{x_2=a_2}^{b_2} \cdots \Delta_{x_k=a_k}^{b_k} G(b'_1, \dots, x_k) - \Delta_{x_2=a_2}^{b_2} \cdots \Delta_{x_k=a_k}^{b_k} G(a_1, \dots, x_k) = \\ &= \Delta_{x_2=a_2}^{b_2} \cdots \Delta_{x_k=a_k}^{b_k} G(b'_1, \dots, x_k) - \Delta_{x_2=a_2}^{b_2} \cdots \Delta_{x_k=a_k}^{b_k} G(b_1, \dots, x_k) + \\ &+ \Delta_{x_2=a_2}^{b_2} \cdots \Delta_{x_k=a_k}^{b_k} G(b_1, \dots, x_k) - \Delta_{x_2=a_2}^{b_2} \cdots \Delta_{x_k=a_k}^{b_k} G(a_1, \dots, x_k) = \\ &= \Delta_{(b_1, a_2, \dots, a_k)}^{b'} F(x) + \Delta_a^b F(x) \geq \Delta_a^b F(x) \end{aligned}$$

y haciendo $a \rightarrow (-\infty, \dots, -\infty)$ obtenemos

$$F(b) \leq F(b')$$

□

Teorema 4.8. *Dada la función de probabilidad $P : \mathcal{B}_k \rightarrow \mathbb{R}$ la función $F : \mathbb{R}^k \rightarrow \mathbb{R}$ definida por*

$$F(x_1, \dots, x_k) = P((-\infty, x_1] \times \dots \times (-\infty, x_k])$$

*es una función de distribución de k dimensiones. Diremos que F es la **función de distribución de la distribución dada P .***

Teorema 4.9. *Existe una biyección entre las funciones de probabilidad P en \mathbb{R}^k y las funciones de distribución F de k dimensiones.*

4.7.1. Distribución discreta

Para una distribución discreta con función puntual de probabilidad $p : \mathbb{R}^k \rightarrow \mathbb{R}$ la función de distribución viene dada por

$$F(x) = \sum_{t \leq x} p(t)$$

4.7.2. Distribución continua

La función de distribución que corresponde a la distribución continua de función de densidad f viene dada por la integral

$$F(x_1, \dots, x_k) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_k} f(t_1, \dots, t_k) dt_1 \dots dt_k$$

Y en los puntos de continuidad de f se tiene que

$$\frac{\partial^k F(x_1, \dots, x_k)}{\partial x_1 \dots \partial x_k} = f(x_1, \dots, x_k)$$

Las observaciones realizadas para las distribuciones absolutamente continuas en \mathbb{R} son igualmente válidas para las distribuciones absolutamente continuas en \mathbb{R}^k .

4.8. Conjunto de continuidad de una función de distribución en \mathbb{R}^k

Una función de distribución de una variable tiene un conjunto de puntos de discontinuidad a lo más numerable. Esto no puede afirmarse para una función de distribución en varias variables. Pero tenemos el siguiente resultado:

Teorema 4.10. *El conjunto de puntos de discontinuidad de una función de distribución F de k variables está contenido en un conjunto a lo más numerable de planos paralelos a $k - 1$ ejes coordenados.*

Por tanto, el conjunto de continuidad de una función de distribución de k variables es denso en \mathbb{R}^k .

5. Variable Aleatoria

5.1. Función inversa. Función compuesta

En esta sección se introducen cuatro funciones obtenidas a partir de una función dada que serán utilizadas en lo que sigue.

1. Sean Ω_1 y Ω_2 dos espacios arbitrarios y f una función con dominio Ω_1 y que toma valores en Ω_2

$$f : \Omega_1 \rightarrow \Omega_2$$

Cuando f es una biyección, determina una **función inversa** $h := f^{-1} : \Omega_2 \rightarrow \Omega_1$ definida por

$$h(\omega_2) = \omega_1 \text{ si } f(\omega_1) = \omega_2.$$

Llamaremos a $h(\omega_2)$ a la **preimagen** o **imagen inversa** de ω_2 por f .

2. Daremos otra definición de función inversa de f sin la restricción de la biyectividad, por lo que tiene un uso más general. Dada una función $f : \Omega_1 \rightarrow \Omega_2$ se llama **función inversa** de f y la representamos por f^{-1} a la función de dominio $\mathcal{P}(\Omega_2)$ y que toma valores de $\mathcal{P}(\Omega_1)$

$$f^{-1} : \mathcal{P}(\Omega_2) \rightarrow \mathcal{P}(\Omega_1),$$

definida de la siguiente manera:

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

El conjunto $f^{-1}(B)$ se llama **preimagen** o **imagen inversa** de B por f .

Teorema 5.1. *Sea $f : \Omega_1 \rightarrow \Omega_2$ una función cualquiera. La función inversa f^{-1} conserva las operaciones con conjuntos. Es decir, valen las siguientes relaciones:*

- $f^{-1}\left(\bigcup_t A_t\right) = \bigcup_t f^{-1}(A_t)$
- $f^{-1}\left(\bigcap_t A_t\right) = \bigcap_t f^{-1}(A_t)$
- $f^{-1}(A^c) = f^{-1}(A)^c$
- $f^{-1}(A \setminus B) = f^{-1}(A) \setminus f^{-1}(B)$
- $A \subseteq B \rightarrow f^{-1}(A) \subseteq f^{-1}(B)$
- $f^{-1}(\Omega_2) = \Omega_1$

3. La función $f : \Omega_1 \rightarrow \Omega_2$ determina también una función \hat{f}^{-1} entre las familias de conjuntos de Ω_1 y Ω_2

$$\hat{f}^{-1} : \mathcal{P}(\mathcal{P}(\Omega_2)) \rightarrow \mathcal{P}(\mathcal{P}(\Omega_1))$$

que se define, apoyándose en $f^{-1} : \mathcal{P}(\Omega_2) \rightarrow \mathcal{P}(\Omega_1)$, de la siguiente manera:

$$\hat{f}^{-1}(\mathcal{C}) = \{f^{-1}(C) : C \in \mathcal{C}\}$$

Igual que antes, podemos usar f^{-1} para estas funciones pues el argumento nos determina cual de las 3 funciones estamos usando.

Teorema 5.2. *Sea $f : \Omega_1 \rightarrow \Omega_2$ una función cualquiera y \mathcal{C} una familia de subconjuntos de Ω_2 . Entonces vale $f^{-1}(\sigma(\mathcal{C})) = \sigma(f^{-1}(\mathcal{C}))$.*

Demostración. Utilizando el teorema anterior, es fácil demostrar que $f^{-1}(\sigma(\mathcal{C}))$ es una σ -álgebra:

- **S1:** $\Omega_1 = f^{-1}(\Omega_2) \in f^{-1}(\sigma(\mathcal{C}))$
- **S2:** $\forall C \in \sigma(\mathcal{C}), f^{-1}(C)^c = f^{-1}(C^c) \in f^{-1}(\sigma(\mathcal{C}))$
- **S3:** Sean $C_n \in \sigma(\mathcal{C}), n = 1, 2, 3, \dots$, se tiene $\bigcup_{n=1}^{\infty} f^{-1}(C_n) = f^{-1}(\bigcup_{n=1}^{\infty} C_n) \in f^{-1}(\sigma(\mathcal{C}))$

Así, como $f^{-1}(\mathcal{C}) \in f^{-1}(\sigma(\mathcal{C}))$, se tiene $f^{-1}(\sigma(\mathcal{C})) \supset \sigma(f^{-1}(\mathcal{C}))$. Para probar la otra inclusión introducimos el mayor subconjunto $\mathcal{E} \subset \sigma(\mathcal{C})$ tal que $f^{-1}(\mathcal{E}) \subset \sigma(f^{-1}(\mathcal{C}))$. Es claro que $\mathcal{C} \subset \mathcal{E}$. Si ahora probásemos que \mathcal{E} es una σ -álgebra, tendríamos $\sigma(\mathcal{C}) \subset \mathcal{E} \subset \sigma(\mathcal{C})$ y por lo tanto la igualdad y así la inclusión buscada. Efectivamente:

- **S1:** $\Omega_2 \in \mathcal{E}$ pues $\Omega_2 \in \sigma(\mathcal{C})$ y $f^{-1}(\Omega_2) = \Omega_1 \in \sigma(f^{-1}(\mathcal{C}))$.
- **S2:** $C \in \mathcal{E} \subset \sigma(\mathcal{C}) \implies C^c \in \sigma(\mathcal{C})$ y $f^{-1}(C^c) = f^{-1}(C)^c \in \sigma(f^{-1}(\mathcal{C}))$
- **S3:** $C_n \in \mathcal{E} \subset \sigma(\mathcal{C}), n = 1, 2, \dots \implies$

$$\bigcup_{n=1}^{\infty} C_n \in \sigma(\mathcal{C}) \text{ y } f^{-1}\left(\bigcup_{n=1}^{\infty} C_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(C_n) \in \sigma(f^{-1}(\mathcal{C})) \quad \square$$

4. La función $f : \Omega_1 \rightarrow \Omega_2$ también determina una función de las partes de Ω_1 en las partes de Ω_2 , $\hat{f} : \mathcal{P}(\Omega_1) \rightarrow \mathcal{P}(\Omega_2)$, definida por :

$$\hat{f}(A) = \{\omega_2 \in \Omega_2 : \exists \omega_1 \in A : \omega_2 = f(\omega_1)\} = \{f(\omega_1) : \omega_1 \in A\}$$

Por el mismo motivo que con las funciones inversas, usaremos indistintamente f para esta función y la primera.

Llamaremos **función compuesta** a la composición habitual de dos funciones $f_1 : \Omega_1 \rightarrow \Omega_2$ y $f_2 : \Omega_2 \rightarrow \Omega_3$ definida por $f(\omega) = f_2(f_1(\omega))$.

Teorema 5.3. *Dadas las funciones $f_1 : \Omega_1 \rightarrow \Omega_2$ y $f_2 : \Omega_2 \rightarrow \Omega_3$, la función compuesta $f = f_2 \circ f_1$ tiene como inversa $f^{-1} : \mathcal{P}(\Omega_3) \rightarrow \mathcal{P}(\Omega_1)$ dada por $f_1^{-1} \circ f_2^{-1}$.*

5.2. Notación. Variable aleatoria (de una dimensión)

Supondremos fijado un espacio de probabilidad (Ω, \mathcal{S}, P) con el espacio muestral Ω . Sea $C(\omega)$ una relación que es verdadera para unos puntos $\omega \in \Omega$ y falsa para otros. Podemos denotar el subconjunto de Ω formado por los elementos $\omega \in \Omega$ que cumplen la condición $C(\omega)$ como:

$$\{C\} = \{\omega \in \Omega : C(\omega)\}$$

y su probabilidad se denota $P(C)$.

Como ejemplo, para una función $X : \Omega \rightarrow \mathbb{R}$ se puede escribir

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\}$$

$$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\}$$

y las probabilidades de estos conjuntos se escribirán $P(X \in B)$ y $P(X \leq x)$ respectivamente.

Llamaremos **variable aleatoria** a una función

$$X : \Omega \rightarrow \mathbb{R}$$

si se verifica

$$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{S}, \forall x \in \mathbb{R}.$$

Esta condición puede escribirse de otras dos formas equivalentes:

- $X^{-1}((-\infty, x]) \in \mathcal{S}, \forall x \in \mathbb{R}$
- $X^{-1}(\mathcal{F}_1) \subset \mathcal{S}, \mathcal{F}_1 = \{(-\infty, x] : x \in \mathbb{R}\}$

Estas funciones X se dice que son variables aleatorias de una dimensión, pues luego introduciremos variables aleatorias de varias dimensiones.

Teorema 5.4. $X : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria si, y solo si, se tiene que

$$\{X \in B\} \in \mathcal{S}, \forall B \in \mathcal{B}_1$$

Esta condición puede escribirse como cualquiera de las formas siguientes

$$X^{-1}(B) \in \mathcal{S}, \forall B \in \mathcal{B}_1$$

$$X^{-1}(\mathcal{B}_1) \subset \mathcal{S}$$

- ' \Leftarrow ' Es evidente, pues $(-\infty, x] \in \mathcal{B}_1$

- ' \Rightarrow '
 - $X^{-1}(\mathcal{F}_1) \subset \mathcal{S} \Rightarrow \sigma(X^{-1}(\mathcal{F}_1)) \subset \mathcal{S} \Leftrightarrow X^{-1}(\sigma(\mathcal{F}_1)) \subset \mathcal{S} \Leftrightarrow X^{-1}(\mathcal{B}_1) \subset \mathcal{S}$

Nótese que para probar que X es una variable aleatoria es suficiente también comprobar que $\{X < x\} \in \mathcal{S}, \forall x \in \mathbb{R}$.

Un **espacio medible** es un par (Ω, \mathcal{S}) tal que \mathcal{S} es una σ -álgebra en Ω .

Dados dos espacios medibles $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$, se dice que una función $f : \Omega_1 \rightarrow \Omega_2$ es una **función medible** de $(\Omega_1, \mathcal{F}_1)$ en $(\Omega_2, \mathcal{F}_2)$ si

$$B \in \mathcal{F}_2 \Rightarrow f^{-1}(B) \in \mathcal{F}_1 \quad (\Leftrightarrow f^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1)$$

De modo que una variable aleatoria es una función medible de (Ω, \mathcal{S}) en $(\mathbb{R}, \mathcal{B}_1)$.

5.3. Distribución de una variable aleatoria

Una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ determina una función de distribución de probabilidad en la recta \mathbb{R} tomando como σ -álgebra la familia de Borel en \mathbb{R} y como función de probabilidad

$$P_X : \mathcal{B}_1 \rightarrow \mathbb{R}$$

definida por

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

Es inmediato comprobar que P_X es una medida de probabilidad en \mathbb{R} , es decir,

1. $P_X(\mathbb{R}) = P(\Omega) = 1$
2. $P_X(B) \geq 0$
3. $P_X(\cup_n B_n) = \sum_n P_X(B_n)$ (B_n disjuntos)

Se tiene, por tanto, un espacio de probabilidad en la recta $(\mathbb{R}, \mathcal{B}_1, P_X)$ que se llama **distribución de probabilidad de la variable aleatoria X** .

Se llama **función de distribución de una variable aleatoria X** a la función de distribución de la distribución $(\mathbb{R}, \mathcal{B}_1, P_X)$ de X . Llamando F a esta, se tiene

$$F(x) = P_X((-\infty, x]) = P(X \in (-\infty, x]) = P(X \leq x)$$

La variable aleatoria X es **discreta** si existe un conjunto $D \subset \mathbb{R}$ finito o numerable con $P(X \in D) = 1$.

La variable aleatoria X es **absolutamente continua** si existe una función de densidad f con la que se pueden calcular las probabilidades

$$P(X \in (a, b)) = \int_a^b f(x) dx$$

La variable aleatoria X es **singular** si $\forall x \in \mathbb{R}, P(X = x) = 0$ y existe un conjunto $C \subset \mathbb{R}$ de medida de Lebesgue cero tal que $P(X \in C) = 1$.

Observación 5.1. Nótese que cualquier distribución de probabilidad en la recta $(\mathbb{R}, \mathcal{B}_1, Q)$ puede considerarse como la distribución de una variable aleatoria.

En efecto, si se toma como espacio de probabilidad el dado y como variable aleatoria $X : \mathbb{R} \rightarrow \mathbb{R}$ la función identidad $X(x) = x$ la función de probabilidad Q_X de X resulta ser

$$Q_X(B) = Q(X \in B) = Q(X^{-1}(B)) = Q(B), \quad B \in \mathcal{B}_1$$

o sea, la misma función Q de partida.

Por tanto, el espacio de probabilidad de X , $(\mathbb{R}, \mathcal{B}_1, Q_X)$ es el mismo $(\mathbb{R}, \mathcal{B}_1, Q)$.

De este modo, es equivalente estudiar las distribuciones de probabilidad en la recta que las distribuciones de probabilidad de las variables aleatorias de una dimensión.

5.4. Transformación o cambio de variable aleatoria. Funciones medibles Borel

Sea E un conjunto de Borel de la recta, $E \in \mathcal{B}_1$. Una función real $g : E \rightarrow \mathbb{R}$ se llama **medible Borel** si se cumple que

$$B \in \mathcal{B}_1 \implies g^{-1}(B) \in \mathcal{B}_1$$

o sea

$$g^{-1}(\mathcal{B}_1) \subset \mathcal{B}_1$$

Para comprobar que una función $g : E \rightarrow \mathbb{R}$ es medible Borel es suficiente comprobar que

$$\{t \in \mathbb{R} : g(t) \leq x\} \in \mathcal{B}_1, \forall x \in \mathbb{R} \text{ o equivalentemente } \{t \in \mathbb{R} : g(t) < x\} \in \mathcal{B}_1, \forall x \in \mathbb{R}$$

y la demostración es análoga a la del Teorema 5.3 (caracterización de variable aleatoria).

Teorema 5.5. *Sea E un conjunto abierto de la recta, $E \in \mathcal{O}_1$. Si la función real $g : E \rightarrow \mathbb{R}$ es continua entonces es medible Borel.*

Demostración. Por ser g continua, las preimágenes de abiertos son abiertos en E . Y como E es abierto, también lo serán en \mathbb{R} , $g^{-1}(\mathcal{O}_1) \subset \mathcal{O}_1$. Por tanto

$$\sigma(g^{-1}(\mathcal{O}_1)) \subset \sigma(\mathcal{O}_1) \implies g^{-1}(\mathcal{B}_1) = g^{-1}(\sigma(\mathcal{O}_1)) \subset \sigma(\mathcal{O}_1) = \mathcal{B}_1.$$

□

Teorema 5.6. *Sea $X : \Omega \rightarrow \mathbb{R}$ una variable aleatoria tal que $X(\Omega)$ está contenida en E , un abierto, y $g : E \rightarrow \mathbb{R}$ una función medible Borel. Entonces la función compuesta $Y = g(X)$ es una variable aleatoria.*

Demostración. En efecto

$$\{Y \leq y\} = \{g(X) \in (-\infty, y]\} = \{X \in g^{-1}((-\infty, y])\} \in \mathcal{S}$$

ya que $g^{-1}((-\infty, y]) \in \mathcal{B}_1$. □

Observación 5.2. Generalización: supongamos que $g : E \rightarrow \mathbb{R}$ es una función medible Borel y $X : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria tal que $X(\Omega) \setminus E \neq \emptyset$. Entonces $g(X)$ no queda bien definida.

Si se cumple $P(X^{-1}(E)) = 1$ la variable aleatoria X' definida por $X' = X \cdot I_{\{X \in E\}} + c \cdot I_{\{X \notin E\}}$, con $c \in E$, o sea

$$X'(w) = \begin{cases} X(w) & \text{si } X(w) \in E \\ c & \text{si } X(w) \notin E \end{cases}$$

tiene la misma distribución que X , $P(X = X') = 1$. Esta distribución común a X y X' es independiente de c . Desde el punto de vista del cálculo de probabilidades las variables aleatorias X y X' son equivalentes. Convendremos por tanto en considerar $g(X)$ como la variable $g(X')$ que queda bien definida.

5.5. Cambio de variable aleatoria en distribuciones continuas

Vamos a ver el cambio de variable aleatoria en el caso en que la variable aleatoria X es continua con función de densidad f_X conocida y se pide la función de densidad f_Y de $Y = g(X)$ donde g es continua y derivable.

Teorema 5.7. Sean E y F dos intervalos de la recta y $g : E \rightarrow F$ una biyección estrictamente monótona que admite la función inversa $h : F \rightarrow E$ con derivada h' continua. Además, sea X una variable aleatoria con función de densidad f_X tal que $f_X(x) = 0, \forall x \notin E$. Entonces la variable aleatoria $Y = g(X)$ tiene la función de densidad

$$f_Y(y) = \begin{cases} f_X(h(y)) |h'(y)| & y \in F \\ 0 & y \notin F \end{cases}$$

Si no se tiene $X(\Omega) \subset E$ aplicaremos la generalización del apartado anterior.

Teorema 5.8. Sean E y F dos conjuntos de \mathbb{R} que son uniones de intervalos abiertos

$$E = \cup_{r=1}^m E_r, \quad F = \cup_{r=1}^m F_r, \quad E_i \cap E_j = \emptyset, \quad \forall i \neq j$$

y $g : E \rightarrow F$ una función que se descompone en un número finito de restricciones biyectivas estrictamente monótonas $g_r : E_r \rightarrow F_r$ con inversas $h_r : F_r \rightarrow E_r$ que tienen derivadas h'_r continuas. Además, X será una variable aleatoria con función de densidad f_X tal que $f_X(x) = 0, \forall x \notin E$. Entonces la variable aleatoria $Y = g(X)$ tiene la función de densidad

$$f_Y(y) = \sum_r f_X(h_r(y)) |h'_r(y)|$$

siendo en esta suma $f_X(h_r(y)) |h'_r(y)| = 0$ si $y \notin F_r$.

Demostración. Como $P(Y \in F^c) = 0$ se tiene que $f_Y(y) = 0$ para $y \in F^c$. Falta determinar $f_Y(y)$ para $y \in F$.

Para cualquier punto $y' \in F$ (excluidos los extremos de F_r) se puede tomar un intervalo (a, b) tal que $a < y' < b$ y que para cada F_r valga una de estas dos posibilidades

$$(a, b) \subset F_r, \quad \text{ó} \quad (a, b) \cap F_r = \emptyset$$

Por tanto, se tendrá

$$\begin{aligned} P(Y \in (a, b)) &= P(g(X) \in (a, b)) = P\left(g(X) \in (a, b), X \in \bigcup_r E_r\right) = P\left(g(X) \in (a, b), \bigcup_r \{X \in E_r\}\right) = \\ &= \sum_r P(g(X) \in (a, b), X \in E_r) = \sum_r P(g_r(X) \in (a, b), X \in E_r) = \sum_r P(X \in h_r((a, b)), X \in E_r) = \\ &= \sum_r P(X \in h_r((a, b))) \quad (\text{para } h_r((a, b)) \neq \emptyset) \end{aligned}$$

Las funciones g_r y h_r son ambas crecientes o ambas decrecientes, y para el intervalo $h_r(a, b)$ transformado de (a, b) se tiene

$$h_r((a, b)) = \begin{cases} (h_r(a), h_r(b)) & h_r \text{ creciente} \\ (h_r(b), h_r(a)) & h_r \text{ decreciente} \end{cases}$$

por lo que los términos de la última suma, si h_r es creciente, valen

$$\begin{aligned} P(X \in h_r((a, b))) &= P(X \in (h_r(a), h_r(b))) = P(h_r(a) < X < h_r(b)) = \int_{h_r(a)}^{h_r(b)} f_X(x) dx \stackrel{*}{=} \\ &= \int_a^b f_X(h_r(y)) h'_r(y) dy = \int_a^b f_X(h_r(y)) |h'_r(y)| dy \end{aligned}$$

donde en * se ha hecho el cambio $x = h_r(y)$.

Si h_r es decreciente, entonces, de igual forma:

$$\begin{aligned} P(X \in h_r((a, b))) &= P(X \in (h_r(b), h_r(a))) = P(h_r(b) < X < h_r(a)) = \int_{h_r(b)}^{h_r(a)} f_X(x) dx = \\ &= \int_b^a f_X(h_r(y)) h'_r(y) dy = \int_a^b f_X(h_r(y)) (-h'_r(y)) dy = \int_a^b f_X(h_r(y)) |h'_r(y)| dy \end{aligned}$$

por lo que todos los términos de la suma tienen la misma expresión y la suma queda como

$$P(Y \in (a, b)) = \sum_r \int_a^b f_X(h_r(y)) |h'_r(y)| dy = \int_a^b \sum_r f_X(h_r(y)) |h'_r(y)| dy$$

lo que demuestra el enunciado. □

5.6. Variables aleatorias con valores infinitos

A veces, para evitar excepciones en las operaciones de límites de números reales o en otras situaciones, conviene agregar a los números reales los signos de límites infinitos.

Llamamos **recta ampliada** al conjunto que se obtiene agregando al conjunto \mathbb{R} de los números reales los dos símbolos $\{-\infty, \infty\}$. Se denota por

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$$

En este conjunto se pueden completar las estructuras de orden, algebraica y topológica que existen en \mathbb{R} . Así, además de las relaciones ordinarias entre elementos de \mathbb{R} es necesario añadir

$$-\infty < a < \infty, \quad \forall a \in \mathbb{R}$$

$$a + \infty = \infty + a = \infty + \infty = \infty, \quad \forall a \in \mathbb{R}$$

$$a + -\infty = -\infty + a = -\infty + -\infty = -\infty, \quad \forall a \in \mathbb{R}$$

$$a(\infty) = (\infty)a = \infty, \quad a(-\infty) = (-\infty)a = -\infty, \quad a > 0$$

$$a(\infty) = (\infty)a = -\infty, \quad a(-\infty) = (-\infty)a = \infty, \quad a < 0$$

$$0(\infty) = (\infty)0 = 0(-\infty) = (-\infty)0 = 0$$

Pero no se pueden definir las operaciones

$$(\infty) + (-\infty), (-\infty) + (\infty), (\infty) - (\infty), (-\infty) - (-\infty)$$

Es inmediato que $\overline{\mathbb{R}}$ no es un cuerpo.

En este conjunto se introduce la familia de Borel como la σ -álgebra en $\overline{\mathbb{R}}$

$$\overline{\mathcal{B}}_1 = \{C : C = B \cup D, B \in \mathcal{B}_1, D \subset \{-\infty, \infty\}\}$$

Una variable aleatoria que puede tomar valores reales ampliados es una función $X : \Omega \rightarrow \overline{\mathbb{R}}$ tal que $\{X \leq x\} \in \mathcal{S}, \forall x \in \overline{\mathbb{R}}$.

Una variable aleatoria con valores reales ampliados es **casi seguramente finita** si vale $P(X \in \mathbb{R}) = 1$. Rara vez se utilizan variables aleatorias con valores reales ampliados que no sean casi seguramente finitas.

6. Variables aleatorias de k dimensiones

6.1. Variable aleatoria de k dimensiones

Dadas las variables aleatorias unidimensionales X_1, \dots, X_k , la función $X : \Omega \rightarrow \mathbb{R}^k$ definida por

$$X(\omega) = (X_1(\omega), \dots, X_k(\omega))$$

(para cada $\omega \in \Omega$) se llama **variable aleatoria de k dimensiones** o **vector aleatorio de k dimensiones**. Cada una de las X_i se llama **componente** de X .

Teorema 6.1. *La función $X = (X_1, \dots, X_k) : \Omega \rightarrow \mathbb{R}^k$ es una variable aleatoria de k dimensiones si, y sólo si, se verifica que*

$$\{X \in B\} \in \mathcal{S}$$

para todo $B \in \mathcal{B}_k$.

Demostración. [\implies] Puesto que cada componente es una variable aleatoria unidimensional, $X_i^{-1}((-\infty, x_i]) \in \mathcal{S}$ para cada i , es decir, $\{X_i \in (-\infty, x_i]\} \in \mathcal{S}$ para cada i . Dado que todos los conjuntos descritos de esta manera están en \mathcal{S} , su intersección también y así

$$\bigcap_{i=1}^k \{X_i \in (-\infty, x_i]\} \in \mathcal{S} \implies \{X \in \prod_{i=1}^k (-\infty, x_i]\} \in \mathcal{S} \implies X^{-1}\left(\prod_{i=1}^k (-\infty, x_i]\right) \in \mathcal{S}$$

Tomando

$$\mathcal{I}_k = \left\{ C \subset \mathbb{R}^k : C = \prod_{i=1}^k (-\infty, x_i], (x_1, \dots, x_k) \in \mathbb{R}^k \right\}$$

tenemos que

$$X^{-1}(\mathcal{I}_k) \subset \mathcal{S} \implies \sigma(X^{-1}(\mathcal{I}_k)) \subset \mathcal{S} \iff X^{-1}(\sigma(\mathcal{I}_k)) \subset \mathcal{S} \iff X^{-1}(\mathcal{B}_k) \subset \mathcal{S}$$

como queríamos ver.

[\impliedby] Veamos que X_1, \dots, X_k son variables aleatorias. Tomamos el semiespacio $S_{i,t} = \{(t_1, \dots, t_k) \in \mathbb{R}^k : t_i \leq t\} \in \mathcal{B}_k$ (para $i = 1, \dots, k$ y $t \in \mathbb{R}$). Entonces

$$\{X \in S_{i,t}\} \in \mathcal{S} \implies \{X_i \leq t\} \in \mathcal{S}$$

luego X_i es una variable aleatoria unidimensional para cada i y por tanto X es una variable aleatoria k -dimensional. \square

6.2. Distribución de una variable aleatoria de k dimensiones

Se llama **distribución de la variable aleatoria X de k dimensiones** a la distribución $(\mathbb{R}^k, \mathcal{B}_k, P_X)$ tal que la función de probabilidad $P_X : \mathcal{B}_k \rightarrow \mathbb{R}$ está definida por

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

para cada $B \in \mathcal{B}_k$. Es inmediato que P_X es una función de probabilidad en \mathbb{R}^k (llamada **función de probabilidad** de la variable aleatoria X). Esta distribución también se denomina **distribución conjunta** de las variables X_1, \dots, X_k .

La **función de distribución** $F : \mathbb{R}^k \rightarrow \mathbb{R}$ de P_X viene determinada por

$$F(x_1, \dots, x_k) = P_X \left(\prod_{i=1}^k (-\infty, x_i] \right) = P(X_1 \leq x_1, \dots, X_k \leq x_k)$$

para cada $(x_1, \dots, x_k) \in \mathbb{R}^k$. Según el tipo de distribución que sea (discreta, absolutamente continua o singular) la variable aleatoria X y su función de densidad también serán del mismo tipo.

6.3. Transformación de variables aleatorias. Funciones medibles Borel

Sean n, k enteros positivos y $E \in \mathcal{B}_k$. Una función $g : E \rightarrow \mathbb{R}^n$ se dice **medible Borel**, si se cumple que

$$B \in \mathcal{B}_n \implies g^{-1}(B) \in \mathcal{B}_k$$

(notemos que la familia $E \cap \mathcal{B}_k = \{C \subset \mathbb{R}^k : C = E \cap B, B \in \mathcal{B}_k\}$ es una σ -álgebra en E).

Teorema 6.2. *Sea $E \subset \mathbb{R}^k$ un abierto y $g : E \rightarrow \mathbb{R}^n$ una función continua. Entonces g es medible Borel.*

Demostración. Por ser g continua, se tiene $g^{-1}(\mathcal{O}_n) \subset \mathcal{O}_k$, pues ser continua es que las preimágenes de los abiertos son abiertos. Por tanto

$$\sigma(g^{-1}(\mathcal{O}_n)) \subset \sigma(\mathcal{O}_k) \implies g^{-1}(\sigma(\mathcal{O}_n)) \subset \sigma(\mathcal{O}_k) \implies g^{-1}(\mathcal{B}_n) \subset \mathcal{B}_k$$

como queríamos. □

Teorema 6.3. *Si $X : \Omega \rightarrow \mathbb{R}^k$ es una variable aleatoria de k dimensiones, $X(\Omega) \subset E \in \mathcal{B}_k$ y $g : E \rightarrow \mathbb{R}^n$ es medible Borel, entonces $Y = g(X)$ es una variable aleatoria de n dimensiones.*

Demostración. Por ser X variable aleatoria, se tiene que $X^{-1}(\mathcal{B}_k) \subset \mathcal{S}$, luego

$$Y^{-1}(\mathcal{B}_n) = (g(X))^{-1}(\mathcal{B}_n) = X^{-1}(g^{-1}(\mathcal{B}_n)) \stackrel{g \text{ medible Borel}}{\subset} X^{-1}(\mathcal{B}_k) \subset \mathcal{S}$$

como queríamos. □

Observación 6.1. Como hacíamos en el caso unidimensional, cuando no se tiene que $X(\Omega) \subset E$, pero aún así se tiene que $P(X^{-1}(E)) = 1$. Tomaremos entonces $X' = X \cdot I_{\{X \in E\}} + c \cdot I_{\{X \notin E\}}$ en lugar de X cuando vayamos a definir $g(X)$.

Como consecuencia de este Teorema, la suma y el producto de variables aleatorias también son variables aleatorias. Del mismo modo, la composición con las funciones habituales (polinomios, trigonométricas, etc.) también lo serán. Por último, si $Y = g(X_1, \dots, X_k)$ es una función racional en las variables X_1, \dots, X_k y el conjunto donde se anula el denominador tiene probabilidad nula, entonces Y es una variable aleatoria con distribución bien determinada.

6.4. Cambio de variable aleatoria en el caso continuo

Teorema 6.4. Sean $E, F \subset \mathbb{R}^2$ abiertos. Sea (X, Y) una variable aleatoria bidimensional con distribución de tipo continuo y función de densidad $f_{(X,Y)}(x, y)$, que es nula en E^c . Sea $g : E \rightarrow F$ una biyección continua cuya inversa $h = g^{-1}$ tiene derivadas continuas y jacobiano $J(h(u, v))$. Entonces la variable aleatoria $(U, V) = g(X, Y)$ tiene la función de densidad

$$f_{(U,V)}(u, v) = \begin{cases} f_{(X,Y)}(h(u, v))|J(h(u, v))| & \text{si } (u, v) \in F \\ 0 & \text{si } (u, v) \notin F \end{cases}$$

Demostración. Como $P((U, V) \in F^c) = P(g(X, Y) \in F^c) = P((x, y) \in h(F^c)) = P((x, y) \in E^c) = 0$, tenemos así la última expresión del enunciado.

Para $B \subset F$ abierto se tiene que

$$P((U, V) \in B) = P(g(X, Y) \in B) = P((X, Y) \in h(B)) = \int_{h(B)} f_{(X,Y)}(x, y) dx dy$$

y resolviendo esta integral mediante el cambio de variable $h(u, v) = (x, y)$ resulta

$$P((U, V) \in B) = \int_B f_{(X,Y)}(h(u, v))|J(h(u, v))| du dv$$

que termina de probar la fórmula del enunciado. □

Teorema 6.5. Sea (X, Y) una variable aleatoria bidimensional con distribución de tipo continuo y función de densidad $f_{(X,Y)}(x, y)$, que es nula en E^c . Si $g : E \rightarrow F$ (abiertos de \mathbb{R}^2) tiene las restricciones biyectivas

$$g_r : E_r \rightarrow F_r, \quad (1 \leq r \leq m), \quad \bigcup_r E_r = E, \quad \bigcup_r F_r = F$$

donde E_r son conjuntos abiertos disjuntos y F_r abiertos con fronteras de área nula, y las inversas $h_r = g_r^{-1}$ tienen derivadas continuas y jacobianos

$$J(h_r(u, v)) = \frac{\partial(x_r, y_r)}{\partial(u, v)}$$

entonces la variable aleatoria $(U, V) = g(X, Y)$ tiene función de densidad

$$f_{(U,V)}(u, v) = \sum_r f_{(X,Y)}(x_r(u, v), y_r(u, v)) |J(h_r)|$$

siendo $f_{(X,Y)}(x_r(u, v), y_r(u, v)) |J(h_r)| = 0$ cuando $(u, v) \notin F_r$

Demostración. Se puede tomar un conjunto B tal que para cada r se cumpla que $B \subset F_r$ o $B \cap F_r = \emptyset$. Para este conjunto se tiene

$$P((U, V) \in B) = P(g(X, Y) \in B, (X, Y) \in \bigcup_r E_r) = \sum_r P(g_r(X, Y) \in B, (X, Y) \in E_r) =$$

$$= \sum_r P((X, Y) \in h_r(B), (X, Y) \in E_r) = \sum_r P((X, Y) \in h_r(B)) = \sum_r \int_{h_r(B)} f_{(X,Y)}(x, y) dx dy$$

y haciendo en la integral r -ésima el cambio de variable $x = x_r(u, v)$, $y = y_r(u, v)$, se obtiene

$$\sum_r \int_B f_{(X,Y)}(x_r(u, v), y_r(u, v)) |J(h_r)| du dv = \int_B \sum_r f_{(X,Y)}(x_r(u, v), y_r(u, v)) |J(h_r)| du dv$$

de donde sigue la fórmula del enunciado. □

Teorema 6.6. Sean $E, F \subset \mathbb{R}^k$ abiertos y $X = (X_1, \dots, X_k)$ una variable aleatoria de k dimensiones con distribución de tipo continuo y función de densidad $f_X(x)$ que es nula en E^c . Si $g : E \rightarrow F$ tiene las restricciones biyectivas $g_r : E_r \rightarrow F_r$ (donde los E_r son disjuntos, $\bigcup_r E_r = E$ y F_r son abiertos de frontera de medida nula) y las inversas $h_r = g_r^{-1}$ tienen derivadas primeras continuas y jacobianos $J(h_r)$ entonces la variable aleatoria $Y = (Y_1, \dots, Y_k) = g(X)$ tiene la función de densidad

$$f_Y(y) = \sum_r f_X(h_r(y)) |J(h_r(y))|$$

siendo $f_X(h_r(y)) |J(h_r(y))| = 0$ cuando $y \notin F_r$.

6.5. Independencia de variables aleatorias

Las variables aleatorias X_1, \dots, X_m son **independientes** si las familias $\mathcal{F}_1, \dots, \mathcal{F}_m$ definidas por

$$\mathcal{F}_r = \{X_r \leq t : t \in \mathbb{R}\} = X_r^{-1}(\mathcal{I}_1), \quad \mathcal{I}_1 = \{(-\infty, t] : t \in \mathbb{R}\}$$

son familias independientes.

Teorema 6.7. Si las variables aleatorias X_1, \dots, X_m unidimensionales son independientes, entonces las familias

$$\mathcal{J}_r = \{\{X_r \in B\} : B \in \mathcal{B}_1\} = X_r^{-1}(\mathcal{B}_1)$$

también lo son.

Demostración. Las familias independientes $\mathcal{F}_1, \dots, \mathcal{F}_m$ pueden escribirse como $X_1^{-1}(\mathcal{I}_1), \dots, X_m^{-1}(\mathcal{I}_1)$ (donde \mathcal{I}_1 es un π -sistema en \mathbb{R}). Como las preimágenes de π -sistemas también son π -sistemas, las σ -álgebras que engendran estas familias son independientes y para $r = 1, \dots, m$ se tiene que

$$\sigma(X_r^{-1}(\mathcal{I}_1)) = X_r^{-1}(\sigma(\mathcal{I}_1)) = X_r^{-1}(\mathcal{B}_1)$$

que es lo afirmado en el enunciado. □

Teorema 6.8. Si X_1, \dots, X_m son variables aleatorias de una dimensión independientes y $g_r : \mathbb{R} \rightarrow \mathbb{R}$ para $r = 1, 2, \dots, m$ son funciones medibles Borel, entonces las variables aleatorias Y_1, \dots, Y_m definidas por $Y_r = g_r(X_r)$, $r = 1, \dots, m$ son independientes.

Demostración. Por las hipótesis son independientes las familias de sucesos

$$X_1^{-1}(\mathcal{B}_1), \dots, X_m^{-1}(\mathcal{B}_1)$$

Como

$$Y_r^{-1}(\mathcal{B}_1) = (g_r(X_r))^{-1}(\mathcal{B}_1) = X_r^{-1}(g^{-1}(\mathcal{B}_1)) \subset X_r^{-1}(\mathcal{B}_1)$$

por lo que las familias $Y_r^{-1}(\mathcal{B}_1)$ son independientes por ser subfamilias de familias independientes. Así, el resultado queda demostrado. \square

Teorema 6.9. Si X_1, \dots, X_m son variables aleatorias independientes de una dimensión con las que se forma la variable aleatoria $X = (X_1, \dots, X_m)$ de dimensión m resulta

$$F(x_1, \dots, x_m) = F(x_1) \cdot F(x_2) \cdots F(x_m)$$

donde F es la función de distribución de X y F_r la de X_r .

Recíprocamente, si las funciones de distribución F de X y F_r de X_r cumplen esta igualdad, las variables aleatorias X_1, \dots, X_m son independientes.

Demostración. Inmediato de la definición de función de distribución y de la de independencia \square

Si X es una variable aleatoria de n dimensiones y $x \in \mathbb{R}^n$, se escribe $X \leq x$ para denotar la desigualdad coordenada a coordenada.

La definición dada para variables aleatorias de una dimensión se generaliza para variables aleatorias de una o varias dimensiones del siguiente modo.

Dadas las variables aleatorias X_1, \dots, X_m de dimensiones respectivas n_1, \dots, n_m , o sea $X_r = (X_{r1}, \dots, X_{rn_r})$ se dice que son independientes si lo son las familias $\mathcal{I}_1, \dots, \mathcal{I}_m$ definidas por

$$\mathcal{I}_r = \left\{ \prod_{1 \leq i \leq n_r} (-\infty, y_i] : y_i \in \mathbb{R}, i = 1, 2, \dots, n_r \right\} = X_r^{-1}(\mathcal{I}_{n_r})$$

Teorema 6.10. Si las variables aleatorias X_1, \dots, X_m de dimensiones n_1, \dots, n_m respectivamente son independientes, también son independientes las familias $\mathcal{G}_1, \dots, \mathcal{G}_m$ definidas por

$$\mathcal{G}_r = \{\{X_r \in B\} : B \in \mathcal{B}_{n_r}\} = X_r^{-1}(\mathcal{B}_{n_r})$$

Demostración. Las familias \mathcal{I}_r pueden escribirse como $X_r^{-1}(\mathcal{I}_{n_r})$, donde \mathcal{I}_{n_r} es un π -sistema en \mathbb{R}^{n_r} . También son π -sistemas las familias $X_r^{-1}(\mathcal{I}_{n_r})$ y son independientes las σ -álgebras que engendran

$$\sigma(X_r^{-1}(\mathcal{I}_{n_r})) = X_r^{-1}(\sigma(\mathcal{I}_{n_r})) = X_r^{-1}(\mathcal{B}_{n_r})$$

como queríamos ver. \square

Teorema 6.11. Si X_1, \dots, X_m son variables aleatorias de dimensiones n_1, \dots, n_m independientes y $g_r : \mathbb{R}^{n_r} \rightarrow \mathbb{R}^{h_r}$ para $r = 1, \dots, m$ son funciones medibles Borel, entonces las variables aleatorias Y_1, \dots, Y_m definidas por

$$Y_r = g_r(X_r), \quad r = 1, \dots, m$$

son independientes.

Demostración. La demostración es análoga a la del teorema 6.9, arreglando las dimensiones de los espacios, como hemos hecho en el teorema 6.11. \square

Teorema 6.12. *Si X_1, \dots, X_m son variables aleatorias independientes de dimensiones n_1, \dots, n_m respectivamente con las que se forma la variable aleatoria $X = (X_1, \dots, X_m)$ de dimensión $\sum_r n_r$ formada tomando ordenadamente las componentes de X_1 seguidas de las de X_2 y así. Entonces, resulta*

$$F(x_1, \dots, x_m) = F_1(x_1) \cdot F_2(x_2) \cdot \dots \cdot F_m(x_m)$$

donde F es la función de distribución de X y F_r la de X_r .

Y recíprocamente, si las funciones de distribución F de X y F_r de X_r cumplen esta igualdad, entonces las variables aleatorias X_1, \dots, X_m son independientes.

Demostración. De nuevo, inmediato de las definiciones de variable aleatoria e independencia. \square

La definición anterior de independencia se aplica a un conjunto finito de variables aleatorias de una o varias dimensiones. Para la independencia de un conjunto de infinitas variables aleatorias se tiene la siguiente definición.

Sea $\{X_t : t \in T\}$ una familia de variables aleatorias donde cada variable aleatoria tiene una o varias dimensiones. Las variables aleatorias de esta familia son independientes si para cada conjunto finito

$$\{t_1, \dots, t_n\} \subset T, \quad n = 2, 3, \dots$$

las variables aleatorias

$$X_1, \dots, X_n$$

son independientes.

En particular, las variables aleatorias de una sucesión X_1, \dots, X_n, \dots son independientes si las variables de cada conjunto finito X_1, \dots, X_k son independientes.

6.5.1. Caso de independencia

Proposición 6.1. Sea $X = (X_1, \dots, X_k)$ una variable aleatoria de k dimensiones y sean

$$I = \{i_1, \dots, i_h\}, \quad J = \{j_1, \dots, j_m\} = \{1, 2, \dots, k\} \setminus I,$$

$$U = (X_{i_1}, \dots, X_{i_h}) \quad \text{y} \quad V = (X_{j_1}, \dots, X_{j_m}),$$

$$u = (x_{i_1}, \dots, x_{i_h}) \quad \text{y} \quad v = (x_{j_1}, \dots, x_{j_m}).$$

Las funciones de distribución de X, U, V son

$$F(x) = F(x_1, \dots, x_k) = P(X_1 \leq x_1, \dots, X_k \leq x_k),$$

$$F_U(u) = P(X_{i_1} \leq x_{i_1}, \dots, X_{i_h} \leq x_{i_h}),$$

$$F_V(v) = P(X_{j_1} \leq x_{j_1}, \dots, X_{j_m} \leq x_{j_m}).$$

Si U y V son independientes entonces

$$F(x) = F(u, v) = F_U(u)F_V(v).$$

7. Distribuciones Marginales Y Condicionadas

7.1. Distribuciones marginales. Casos discreto y continuo

Si $X = (X_1, X_2, \dots, X_k)$ es una variable aleatoria de k dimensiones, las distribuciones de las variables aleatorias de dimensión h cuyas h componentes son componentes de X se llaman **distribuciones marginales** de la distribución X . Es decir,

$$\forall I = \{i_1, \dots, i_h\} \subset \{1, 2, \dots, k\},$$

la distribución de la variable aleatoria

$$U = (X_{i_1}, \dots, X_{i_h})$$

se dice que es la **distribución marginal** de la distribución de X relativa a las componentes de índices i_1, \dots, i_h de X .

En esta sección vamos a mantener fijos los dos conjuntos de índices siguientes:

$$I = \{i_1, \dots, i_h\},$$

$$J = \{j_1, \dots, j_m\} = \{1, \dots, k\} \setminus \{i_1, \dots, i_h\}.$$

Y para mayor sencillez y sin perder generalidad podemos suponer que los índices en I están ordenados de manera creciente, así como los índices de J . Estos dos conjuntos I, J determinan las variables aleatorias

$$U = (X_{i_1}, \dots, X_{i_h})$$

$$V = (X_{j_1}, \dots, X_{j_m})$$

Para cada punto $x = (x_1, \dots, x_k) \in \mathbb{R}^k$ escribimos

$$u = (x_{i_1}, \dots, x_{i_h}) \in \mathbb{R}^h, v = (x_{j_1}, \dots, x_{j_m}) \in \mathbb{R}^m$$

Si $X = (X_1, \dots, X_k)$ tiene la función de distribución

$$F(x) = F(x_1, \dots, x_k) = P(X_1 \leq x_1, \dots, X_k \leq x_k)$$

la función de distribución de U se puede denotar por

$$F_{i_1, \dots, i_h}(x_{i_1}, \dots, x_{i_h}) = P(X_{i_1} \leq x_{i_1}, \dots, X_{i_h} \leq x_{i_h})$$

y se obtiene haciendo tender a $+\infty$ las variables x_j cuyos subíndices j pertenecen al conjunto J . Es obvio que estas igualdades no depende de las variables x_{i_1}, \dots, x_{i_h} elegidas.

7.1.1. Distribuciones marginales en el caso discreto

Sea $X = (X_1, \dots, X_k)$ una variable aleatoria discreta con función puntual de probabilidad $p : \mathbb{R}^k \rightarrow \mathbb{R}$ dada por

$$p(x) = p(x_1, \dots, x_k) = P(X_1 = x_1, \dots, X_k = x_k) \text{ con } p(x) \geq 0, \sum_x p(x) = 1$$

La distribución marginal de X relativa a las componentes de i_1, \dots, i_h , es decir, la distribución de $U = (X_{i_1}, \dots, X_{i_h})$, es una distribución discreta con la función puntual de probabilidad

$$\begin{aligned} p_{i_1, \dots, i_h}(x_{i_1}, \dots, x_{i_h}) &= P(X_{i_1} = x_{i_1}, \dots, X_{i_h} = x_{i_h}) = \\ &= \sum_{x_{j_1}, \dots, x_{j_m}} P(X_1 = x_1, \dots, X_k = x_k) = \sum_{x_{j_1}, \dots, x_{j_m}} p(x_1, \dots, x_k) \end{aligned}$$

7.1.2. Distribuciones marginales en el caso continuo

Sea $X = (X_1, \dots, X_k)$ una variable aleatoria continua con función de densidad $f : \mathbb{R}^k \rightarrow \mathbb{R}$

$$f(x) = f(x_1, \dots, x_k) \geq 0 \int_{\mathbb{R}^k} f(x) dx_1 \dots dx_k = 1$$

La distribución marginal de X relativa a las componentes i_1, \dots, i_h es la distribución de U y tiene la función densidad de probabilidad $f_{i_1, \dots, i_h} : \mathbb{R}^h \rightarrow \mathbb{R}$ definida por

$$f_{i_1, \dots, i_h}(x_{i_1}, \dots, x_{i_h}) = \int_{\mathbb{R}^m} f(x_1, \dots, x_k) dx_{j_1} \dots dx_{j_m}$$

Para comprobar que esta es efectivamente su función de probabilidad, tomemos el intervalo $(a, b) \subseteq$

$$\mathbb{R}^h, \prod_{i=1}^h (a_i, b_i) = (a, b)$$

Se tiene entonces que

$$\begin{aligned} P(U \in (a, b)) &= P(a_1 < X_{i_1} < b_1, \dots, a_h < X_{i_h} < b_h) = \\ &= \int_{u \in (a, b), v \in \mathbb{R}^k} f(x_1, \dots, x_k) dx_1 \dots dx_k = \int_{u \in (a, b)} \left(\int_{v \in \mathbb{R}^m} f(x_1, \dots, x_k) dx_{j_1} \dots dx_{j_m} \right) dx_{i_1} \dots dx_{i_h} \end{aligned}$$

Luego si hacemos esta última integral con $u \in \mathbb{R}^m$, tendríamos que esta es la función de densidad de U .

7.2. Distribuciones condicionadas en el caso discreto y en el caso continuo

7.2.1. Caso discreto

Sea $X = (X_1, \dots, X_k)$ una variable aleatoria discreta con función puntual de probabilidad $p : \mathbb{R}^k \rightarrow \mathbb{R}$ dada por

$$\begin{aligned} p(x) &= p(x_1, \dots, x_k) = P(X_1 = x_1, \dots, X_k = x_k) \\ &\text{con } p(x) \geq 0 \text{ y } \sum p(x) = 1 \end{aligned}$$

Como anteriormente, fijamos los dos conjuntos de índices

$$I = \{i_1, \dots, i_h\} \quad i_1 < i_2 < \dots < i_h$$

$$J = \{j_1, \dots, j_h\} = \{1, \dots, k\} \setminus \{i_1, \dots, i_h\} \quad j_1 < j_2 < \dots < j_h$$

Formando las variables aleatorias

$$U = (X_{i_1}, \dots, X_{i_h})$$

$$V = (X_{j_1}, \dots, X_{j_m})$$

Y para cada punto $x = (x_1, \dots, x_k) \in \mathbb{R}^k$ podemos formar los puntos $u = (x_{i_1}, \dots, x_{i_h}) \in \mathbb{R}^h$ y los puntos $v = (x_{j_1}, \dots, x_{j_m}) \in \mathbb{R}^m$. Con esta notación podemos dar la siguiente definición:

Se llama **distribución de V condicionada por $U = u$** a la distribución que se obtiene con la función puntual de probabilidad en la variable v

$$(1) P(V = v|U = u) = \frac{P(U = u, V = v)}{P(U = u)} = \frac{P(X = x)}{P(U = u)} = \frac{p(x)}{p_{i_1, \dots, i_h}(u)}$$

válida para $p_{i_1, \dots, i_h}(x_{i_1}, \dots, x_{i_h}) > 0$. En esta distribución u es un parámetro, es decir, para cada valor de u se obtiene una distribución distinta; por tanto (1) representa una familia de distribuciones con el parámetro u .

La función de probabilidad (1) la denotaremos por

$$p_{j_1, \dots, j_m | i_1, \dots, i_h}(x_{j_1}, \dots, x_{j_m} | u) = \frac{p(x)}{p_{i_1, \dots, i_h}(u)}, \quad u = (x_{i_1}, \dots, x_{i_h})$$

7.2.2. Distribuciones condicionadas en el caso continuo

Sea $X = (X_1, \dots, X_k)$ una variable aleatoria continua con función de densidad $f : \mathbb{R}^k \rightarrow \mathbb{R}$

$$f(x) = f(x_1, \dots, x_k) \geq 0 \int_{\mathbb{R}^k} f(x) dx_1 \dots dx_k = 1$$

Con las mismas notaciones que antes para U y para V , la **distribución de V condicionada por $U = u$** tiene la función de densidad en la variable v

$$f_{j_1, \dots, j_m}(x_{j_1}, \dots, x_{j_m} | u) = \frac{f(x)}{f_{i_1, \dots, i_h}(u)}, \quad u = (x_{i_1}, \dots, x_{i_h})$$

donde se supone $f_{i_1, \dots, i_h}(x_{i_1}, \dots, x_{i_h}) > 0$.

Igual que antes, u es un parámetro, luego para las distribuciones continuas también tenemos una familia de distribuciones, en este caso continuas.

7.3. Caso de independencia

7.3.1. Caso discreto

Siguiendo la notación anterior, si la variable X tiene una distribución discreta y las variables aleatorias U y V son independientes se verifica

$$P(X = x) = P(U = u, V = v) = P(U = u)P(V = v)$$

Luego para las funciones de punto de sus distribuciones se tiene

$$(1) p(x_1, \dots, x_k) = p_{i_1, \dots, i_h}(x_{i_1}, \dots, x_{i_h}) \cdot p_{j_1, \dots, j_m}(x_{j_1}, \dots, x_{j_m})$$

De esta igualdad, resulta que la distribución condicionada de V por $U = u$ en el caso de independencia de U y V coincide con la marginal de V , es decir:

$$p_{j_1, \dots, j_m | i_1, \dots, i_h}(x_{j_1}, \dots, x_{j_m} | x_{i_1}, \dots, x_{i_h}) = p_{j_1, \dots, j_m}(x_{j_1}, \dots, x_{j_m})$$

7.3.2. Caso continuo

Si la variable X tiene una distribución continua y las variables aleatorias U y V son independientes se verifica para las funciones de densidad lo siguiente

$$(2) f(x_1, \dots, x_k) = f_{j_1, \dots, j_m}(x_{j_1}, \dots, x_{j_m}) \cdot f_{i_1, \dots, i_h}(x_{i_1}, \dots, x_{i_h})$$

Si se toman dos intervalos $(a, b) \subset \mathbb{R}^h$ y $(c, d) \subset \mathbb{R}^m$ la independencia de U y de V nos da la igualdad

$$P(U \in (a, b), V \in (c, d)) = P(U \in (a, b)) \cdot P(V \in (c, d))$$

que en forma de integral es

$$\begin{aligned} \int_{u \in (a,b), v \in (c,d)} f(x_1, \dots, x_k) dx_1 \dots dx_k &= \int_{u \in (a,b)} f_{i_1, \dots, i_h}(u) dx_{i_1} \dots dx_{i_h} \cdot \int_{v \in (c,d)} f_{j_1, \dots, j_m}(v) dx_{j_1} \dots dx_{j_m} = \\ &= \int_{u \in (a,b), v \in (c,d)} f_{i_1, \dots, i_h}(u) \cdot f_{j_1, \dots, j_m}(v) dx_1 \dots dx_k \end{aligned}$$

Por esta última igualdad tenemos efectivamente (2), excepto en un conjunto de medida cero (la integral en un conjunto de medida cero da 0), pero en este conjunto podemos modificar una de las densidades para que valga (2) sin importar el punto $x \in \mathbb{R}^k$.

También podemos obtener (2) si usamos la igualdad proporcionada por la proposición 6.14 y la derivamos respecto de las k variables x_1, \dots, x_k en los puntos de continuidad de f .

De nuevo, la condición de independencia (2) implica que la distribución condicionada de V por $U = u$ es la misma que la marginal de V si U y V son independientes, o sea

$$f_{j_1 \dots j_m | i_1 \dots i_h}(x_{j_1}, \dots, x_{j_m} | x_{i_1}, \dots, x_{i_h}) = f_{j_1 \dots j_m}(x_{j_1}, \dots, x_{j_m})$$

8. Distribuciones condicionadas

8.1. Definición general de las distribuciones condicionadas

El problema de la construcción de las distribuciones condicionadas consiste en lo siguiente. Supongamos que (X, Y) tiene una función de distribución $F(x, y)$ conocida y $F_1(x)$ es la función de distribución marginal de X . Se trata de encontrar, para cada x , una función de distribución en la variable y , $F_{2|1}(y|x)$, que represente la distribución de Y condicionada por $X = x$.

Para encontrar al condición que debe satisfacer la función $F_{2|1}(y|x)$, fijémonos en el caso discreto. Se puede escribir según la fórmula de la probabilidad total

$$P((X, Y) \in C) = \sum_{(x,y) \in C} P(X = x, Y = y) = \sum_{(x,y) \in C} P(X = x) P(Y = y|X = x)$$

donde C es un conjunto rectangular de Borel de \mathbb{R}^2 , y reagrupando los sumandos en el último miembro, la igualdad anterior se escribe como

$$\sum_{(x,y) \in C} P(X = x, Y = y) = \sum_{x \in pr_1 C} \left(\sum_{y \in C_x} P(Y = y|X = x) \right) P(X = x) \quad (1)$$

donde $C_x = \{y \in \mathbb{R} | (x, y) \in C\}$, $pr_1 C = \{x \in \mathbb{R} | \exists y : (x, y) \in C\}$.

Estas dos fórmulas pueden aplicarse a cualquier conjunto $C \subset \mathbb{R}^2$ y nos dan las dos definiciones siguientes.

El conjunto C_x es la **sección** de C por la recta de primera coordenada x . El conjunto $pr_1 C$ se llama **proyección** de C sobre el primer eje coordenado.

En forma de funciones puntuales de probabilidad, la igualdad (1) es

$$P((X, Y) \in C) = \sum_{(x,y) \in C} p(x, y) = \sum_{x \in pr_1 C} \left(\sum_{y \in C_x} p(y|x) \right) p_1(x)$$

Traduciendo esta relación mediante las funciones de distribución la anterior se convierte en

$$\int_C dF(x, y) = \int_{pr_1 C} \left(\int_{C_x} d_y F_{2|1}(y|x) \right) dF_1(x) \quad (2)$$

d_y significa que la diferencial actúa sobre la variable y , y se ha llamado $F_{2|1}(y|x)$ a la función de distribución que corresponde a la función puntual de probabilidad $p_{2|1}(y|x)$.

Este método heurístico de obtener la relación (2) entre F, F_1 y $F_{2|1}$ para distribuciones discretas no demuestra en el caso general la existencia de las funciones de distribución condicionadas $F_{2|1}$ sino solo una relación deseable para estas funciones de distribución.

Si en (2) se sustituye $C = B \times (-\infty, v]$, $B \in \mathcal{B}_1, v \in \mathbb{R}$, se obtiene

$$\int_{B \times (-\infty, v]} dF(x, y) = \int_B \left(\int_{(-\infty, v]} d_y F_{2|1}(y|x) \right) dF_1(x), \quad B \in \mathcal{B}_1, v \in \mathbb{R} \quad (3)$$

Esta condición es una ecuación suficiente para determinar $F_{2|1}$ a partir de F y F_1 y permite dar la siguiente definición.

Dada cualquier variable aleatoria bidimensional (X, Y) con función de distribución $F(x, y)$, llamamos **distribución de Y condicionada por $X = x$** a la distribución con función de distribución en la variable y $F_{2|1}(y|x)$ que satisface (3) para todo $B \in \mathcal{B}_1, v \in \mathbb{R}$ donde $F_1(x) = F(x, \infty)$. En $F_{2|1}(y|x)$ cada valor de x determina una distribución condicionada y , como muestra (3), estas distribuciones condicionadas deben ser consideradas simultáneamente.

Como ocurre con cualquier definición descriptiva, esta definición precisa de un Teorema de existencia del concepto definido, y, en su caso, un Teorema de unicidad.

La demostración de la existencia se verá más adelante.

Respecto a la unicidad, conviene advertir que las funciones de distribución $F_{2|1}$ quedan determinadas por (3) salvo en un conjunto de valores de x de medida F_1 nula.

Observación 8.1. Si en (2) se introduce la función indicador de un conjunto, se obtiene

$$\int_{\mathbb{R}^2} I_C(x, y) dF(x, y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} I_{C_x}(y) d_y F_{2|1}(y|x) \right) dF_1(x)$$

para los rectángulos medibles $C \in \mathcal{B}_2$.

La (3) conduce a la generalización siguiente

$$\int_{\mathbb{R}^2} g(x, y) dF(x, y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(x, y) d_y F_{2|1}(y|x) \right) dF_1(x)$$

donde g es medible Borel. Conocido el concepto de esperanza matemática, esta última igualdad relaciona la esperanza matemática de $g(X, Y)$ con la esperanza de $g(X, Y)$ condicionada por $X = x$.

Si llamamos P, P_2 y $P_{(x)}$ a las funciones de probabilidad de las funciones de distribución F, F_1 y $F_{2|1}$ esta igualdad se escribe

$$\int_{\mathbb{R}^2} g(x, y) dP = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(x, y) dP_{(x)}(\{y\}) \right) dP_1(\{x\})$$

8.2. Casos particulares. Caso de independencia

Debemos comprobar que la definición general dada anteriormente que se apoya en la condición (3) está de acuerdo con las definiciones dadas anteriormente para las distribuciones condicionadas.

8.2.1. Distribuciones discretas

Si x_i fuese un punto con $F_1(x_i) - F_1(x_i-) > 0$ entonces $F_{2|1}(y|x_i)$ quedaría bien determinada, pues tomando en (3) $B = \{x_i\}$ resulta

$$P(X = x_i, Y \leq v) = \int_{\{x_i\}} \left(\int_{(-\infty, v]} d_y F_{2|1}(y|x) \right) dF_1(x) = \int_{\{x_i\}} F_{2|1}(v|x) dF_1(x) = F_{2|1}(v|x_i) (F(x_i) - F(x_i-))$$

lo que nos da

$$F_{2|1}(y|x_i) = \frac{P(X = x_i, Y \leq y)}{F(x_i) - F(x_i-)} \quad (4)$$

que prueba que para estos valores $X = x_i$ la función $F_{2|1}(y|x_i)$ queda determinada.

En el caso de tener (X, Y) una distribución de tipo discreto, de (4) resulta

$$F_{2|1}(y_j|x_i) - F_{2|1}(y_j - |x_i) = \frac{P(X = x_i, Y = y_j) - P(X = x_i, Y = y_j -)}{P(X = x_i) - P(X = x_i -)} = \frac{p(x_i, x_j)}{p_1(x_i)} = p_{2|1}(y_j|x_i)$$

que es la fórmula que se dio como definición de distribución condicionada anteriormente para las distribuciones bidimensionales discretas.

8.2.2. Distribuciones continuas

En el caso continuo utilizaremos la definición de $f_{2|1}$ dada para las distribuciones continuas y la (3), suponiendo que $f_1(x) > 0$ en el conjunto de borel B , nos da

$$\begin{aligned} P((X, Y) \in B \times (-\infty, v]) &= \int_{B \times (-\infty, v]} dF(x, y) = \int_{B \times (-\infty, v]} f(x, y) dx dy = \int_B \left(\int_{(-\infty, v]} \frac{f(x, y)}{f_1(x)} dy \right) f_1(x) dx = \\ &= \int_B \left(\int_{(-\infty, v]} f_{2|1}(y|x) dy \right) f_1(x) dx \end{aligned}$$

introduciendo la función de distribución con función de densidad $f_{2|1}$

$$F_{2|1}(y|x) = \int_{-\infty}^y f_{2|1}(t|x) dt \quad (5)$$

y la relación anterior se escribe

$$\int_{B \times (-\infty, v]} dF(x, y) = \int_B \left(\int_{(-\infty, v]} d_y F_{2|1}(y|x) dy \right) f_1(x) dx = \int_B \left(\int_{(-\infty, v]} d_y F_{2|1}(y|x) dy \right) dF_1(x)$$

que prueba que las funciones de distribución (5) satisfacen la condición (3) y queda justificada la definición dada anteriormente de la densidad $f_{2|1}$.

Proposición 8.1. Caso de independencia.

Si X, Y son variables aleatorias independientes, la función de distribución de la variable Y condicionada con $X = x$ es la misma distribución marginal de Y

Demostración. Siguiendo las notaciones anteriores, sea F la función de distribución de (X, Y) y F_1, F_2 las funciones de distribución de las variables aleatorias X e Y , respectivamente.

Comprobaremos que como familia de funciones de distribución condicionadas $F_{2|1}(y|x)$ se puede tomar la familia $F_2(y)$ que resulta de suprimir la condición $X = x$, es decir

$$F_{2|1}(y|x) = F_2(y) \quad (*)$$

Para ello partimos del segundo miembro de (3) y con la sustitución (*) se tiene

$$\begin{aligned} \int_B \left(\int_{(-\infty, v]} d_y F_{2|1}(y|x) \right) dF_1(x) &= \int_B \left(\int_{(-\infty, v]} d_y F_2(y) \right) dF_1(x) = \int_B F_2(v) dF_1(x) = \\ &= F_2(v) \int_B dF_1(x) = P(X \in B) P(Y \leq v) \stackrel{X, Y \text{ indep}}{=} P((X, Y) \in B \times (-\infty, v]) = \int_{B \times (-\infty, v]} dF(x, y) \end{aligned}$$

que prueba que con (*) se satisface (3) cuando X, Y son independientes. \square

8.3. Convolución

Sean X, Y dos variables aleatorias independientes con F_1 y F_2 como sus funciones de distribución, la función de distribución de la variable aleatoria 2-dimensional (X, Y) será $F(x, y) = F_1(x)F_2(y)$. Si se desea calcular la función de distribución F_Z de $Z = X + Y$ de las dos variables aleatorias se tiene

$$F_Z(z) = P(X + Y \leq z) = P((x, y) \in C) = \int_C dF(x, y) \quad C = \{(x, y) \in \mathbb{R}^2 : x + y \leq z\}$$

Por la independencia de X e Y tenemos que $F_{2|1}(y|x) = F_2(y)$, luego

$$F_Z(z) = \int_C F(x, y) = \int_{\mathbb{R}} \int_{C_x} d_y F_{2|1}(y|x) dF_1(x) = \int_{\mathbb{R}} \left(\int_{C_x} d_y F_2(y) \right) dF_1(x) = \int_{\mathbb{R}} \left(\int_{C_x} dF_2(y) \right) dF_1(x)$$

$$C_x = \{y : (x, y) \in C\} = \{y : x + y \leq z\} = (-\infty, z - x]$$

luego

$$F_Z(z) = \int_{\mathbb{R}} \left(\int_{(-\infty, z-x]} dF_2(y) \right) dF_1(x) = \int_{\mathbb{R}} F_2(z - x) dF_1(x)$$

es la solución buscada.

Dado este desarrollo, llamaremos **convolución** de dos funciones de distribución F_1 y F_2 a la función F_Z obtenida por la fórmula: $F_Z(z) = \int_{\mathbb{R}} F_2(z - x) dF_1(x)$. Además esta operación es conmutativa $\int_{\mathbb{R}} F_2(z - x) dF_1(x) = \int_{\mathbb{R}} F_1(z - x) dF_2(x)$.

Teorema 8.1. *Si una de las distribuciones F_1 o F_2 es absolutamente continua, entonces F_Z también es absolutamente continua.*

Demostración. Veámoslo con F_2 , pues si fuera con F_1 sería análogo. $F_2(y) = \int_{-\infty}^y f_2(u) du$ con f_2 función de densidad de F_2 será entonces $F_2(z - x) = \int_{-\infty}^{z-x} f_2(u) du$ y haciendo el cambio $u = v - x$ tenemos $F_2(z - x) = \int_{-\infty}^z f_2(v - x) dv$, luego la fórmula de la convolución queda así:

$$F_Z(z) = \int_{\mathbb{R}} F_2(z - x) dF_1(x) = \int_{\mathbb{R}} \left(\int_{-\infty}^z f_2(v - x) dv \right) dF_1(x) = \int_{-\infty}^z \left(\int_{\mathbb{R}} f_2(v - x) dF_1(x) \right) dv$$

Luego la función F_Z tiene como función de densidad $f_Z(z) = \int_{\mathbb{R}} f_2(z - x) dF_1(x)$. Si además F_1 tiene función de densidad f_1 , obtenemos $f_Z = \int_{\mathbb{R}} f_2(v - x) f_1(x) dx$ que es la convolución de las funciones de densidad f_1 y f_2 . □

9. Funciones Beta y Gamma

9.1. Función Gamma

Llamaremos función gamma (Γ) a la función definida en el intervalo $(0, +\infty)$ por

$$\Gamma(p) = \int_0^{+\infty} e^{-t} t^{p-1} dt.$$

Teorema 9.1. $\forall p > 0$ vale: $\Gamma(p+1) = p \cdot \Gamma(p)$. En particular, si $n \in \mathbb{N}$ tenemos $\Gamma(n+1) = n!$

Demostración. Haciendo integración por partes, tenemos lo siguiente:

$$\Gamma(p+1) = \int_0^{+\infty} e^{-t} t^p dt \stackrel{*}{=} [-e^{-t} t^p]_0^{+\infty} + p \cdot \int_0^{+\infty} e^{-t} t^{p-1} dt$$

Haciendo en partes $*$: $\begin{cases} u = t^p & du = p t^{p-1} dt \\ dv = e^{-t} dt & v = -e^{-t} \end{cases}$. Además $[-e^{-t} t^p]_0^{+\infty} = 0$ pues $p > 0$, luego tenemos que $\Gamma(p+1) = p \cdot \int_0^{+\infty} e^{-t} t^{p-1} dt = p \cdot \Gamma(p)$.

Si en la fórmula anterior tenemos $\Gamma(1) = \int_0^{+\infty} e^{-t} dt = [-e^{-t}]_0^{+\infty} = 1$ por lo que $\Gamma(n+1) = n \cdot \Gamma(n) = \dots = n! \cdot \Gamma(1) = n!$ teniendo así el segundo resultado. \square

Teorema 9.2. Si $a \in \mathbb{C}$, tal que $Re(a) > 0$, y $p > 0$, entonces:

$$\int_0^{+\infty} e^{-at} t^{p-1} dt = \frac{\Gamma(p)}{a^p}$$

Demostración. Realizando el cambio de variable $\begin{cases} t = a \cdot u & dt = a du \end{cases}$ en la integral de la función Γ , tenemos:

$$\Gamma(p) = \int_0^{+\infty} e^{-t} t^{p-1} dt = \int_0^{+\infty} e^{-au} (au)^{p-1} \cdot a du = a^p \int_0^{+\infty} e^{-au} u^{p-1} du$$

Luego despejando tenemos

$$\frac{\Gamma(p)}{a^p} = \int_0^{+\infty} e^{-at} t^{p-1} dt$$

Esta igualdad se tiene $\forall a \in \mathbb{R}^+$, ahora bien, si $a \in \mathbb{C}$ con $Re(a) > 0$, tenemos que las funciones complejas a cada lado de la igualdad son analíticas en a y coinciden en el semieje real positivo, por lo que coinciden en todos los puntos en los que sean holomorfas, luego al menos en el semiplano real positivo. \square

9.2. Función Beta

Llamaremos **función beta** (B) a la función definida $\forall p > 0, \forall q > 0$ por la integral:

$$B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt$$

Teorema 9.3. La función beta es simétrica en sus dos variables, es decir $B(p, q) = B(q, p)$

Demostración. Haciendo el cambio de variable $\begin{cases} t = 1 - u \\ dt = -1 du \end{cases}$ tenemos:

$$B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt = - \int_1^0 (1-u)^{p-1}u^{q-1} du = \int_0^1 u^{q-1}(1-u)^{p-1} du = B(q, p)$$

□

Teorema 9.4. La función beta puede escribirse de la forma

$$B(p, q) = \int_0^{+\infty} \frac{x^{p-1}}{(1+x)^{p+q}} dx \quad p > 0, q > 0$$

Demostración. Haciendo el cambio de variable $\begin{cases} t = \frac{x}{1+x} \\ dt = \frac{1}{(1+x)^2} dx \end{cases}$ tenemos:

$$\begin{aligned} B(p, q) &= \int_0^1 t^{p-1}(1-t)^{q-1} dt = \int_0^{+\infty} \frac{\left(\frac{x}{1+x}\right)^{p-1} \left(1 - \frac{x}{1+x}\right)^{q-1}}{(1+x)^2} dx = \\ &= \int_0^{+\infty} \frac{x^{p-1} \left(\frac{1+x-x}{1+x}\right)^{q-1}}{(1+x)^{p+1}} dx = \int_0^{+\infty} \frac{x^{p-1}}{(1+x)^{p+q}} dx \end{aligned}$$

□

9.3. Relación entre las funciones gamma y beta

Teorema 9.5. $\forall p > 0, \forall q > 0$ tenemos

$$B(p, q) = \frac{\Gamma(p) \cdot \Gamma(q)}{\Gamma(p+q)}$$

Teorema 9.6. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Demostración. $B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt$ y haciendo $p = q = \frac{1}{2}$ obtenemos

$$\Gamma\left(\frac{1}{2}\right)^2 = \frac{\Gamma\left(\frac{1}{2}\right)^2}{\Gamma(1)} = B\left(\frac{1}{2}, \frac{1}{2}\right) = \int_0^1 t^{\frac{1}{2}-1}(1-t)^{\frac{1}{2}-1} dt = \int_0^1 \frac{1}{\sqrt{t(1-t)}} dt$$

y haciendo $t = \frac{1+x}{2}$, entonces $1-t = \frac{1-x}{2}$ y $dt = \frac{1}{2} dx$ obtenemos

$$\Gamma\left(\frac{1}{2}\right)^2 = \int_{-1}^1 \frac{1}{\sqrt{1-x^2}} dx = 2 \int_0^1 \frac{1}{\sqrt{1-x^2}} dx = 2 [\arcsin x]_0^1 = 2 \frac{\pi}{2} = \pi.$$

□

De la propiedad anterior y de $\Gamma(p+1) = p \cdot \Gamma(p)$ tenemos que

$$\Gamma\left(n + \frac{1}{2}\right) = \sqrt{\pi} \prod_{i=1}^{n-1} \left(i + \frac{1}{2}\right) = \sqrt{\pi} \prod_{i=1}^{n-1} \left(\frac{2i+1}{2}\right) = \frac{1 \cdot 3 \cdots (2n-1)}{2^n} \sqrt{\pi}$$

Usando la notación

$$1 \cdot 3 \cdots (2n-1) = (2n-1)!!$$

$$2 \cdot 4 \cdots (2n) = (2n)!!$$

entonces se escribe

$$\Gamma\left(n + \frac{1}{2}\right) = \frac{(2n-1)!!}{2^n} \sqrt{\pi}$$

Una expresión de la función Γ mediante su límite es $\Gamma(p) = \lim_n \frac{n^n n!}{p(p+1) \cdots (p+n)}$

10. Esperanza Matemática

10.1. Variables aleatorias simples

Sea un espacio medible (Ω, \mathcal{S}) y $\{A_i\}_{i \in \{1, 2, \dots, n\}} \subset \mathcal{S}$ una partición finita de Ω por conjuntos medibles, las **funciones simples** $f : \Omega \rightarrow \mathbb{C}$ son aquellas de la forma

$$f(x) = \sum_{i=1}^n a_i I_{A_i}(x),$$

donde $a_1, a_2, \dots, a_n \in \mathbb{C}$ son constantes complejas (o reales). Nótese que las funciones simples son medibles pues, sea $C \subset \mathbb{C}$ un conjunto cualquiera, se tiene que

$$f^{-1}(C) = \{\omega \in \Omega : f(\omega) = a_i \in C\} = \bigcup_{a_i \in C} A_i \in \mathcal{S}.$$

Una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ se llama **variable aleatoria simple** si puede expresarse como una suma finita de la forma

$$X = \sum_{i=1}^m x_i I_{A_i}$$

donde $x_i \in \mathbb{R}$, $A_i \in \mathcal{S}$, $i = 1, \dots, m$ y $\bigcup_{i=1}^m A_i = \Omega$ con $A_i \cap A_j = \emptyset, \forall i \neq j$, es decir, que $\{A_i\}_{i=1, \dots, m}$ es una partición de Ω .

Proposición 10.1. Si $g : \mathbb{R} \rightarrow \mathbb{R}$ es una función cualquiera y X es una variable aleatoria simple, entonces la función compuesta $g(X)$ es una variable aleatoria simple.

Demostración. En efecto, si X es una variable aleatoria simple, entonces

$$g(X) = g\left(\sum_{i=1}^m x_i I_{A_i}\right) \stackrel{*}{=} \sum_{i=1}^m g(x_i) I_{A_i}$$

donde $*$ se debe a que los A_i son una partición de Ω , o sea, que solo uno de los sumandos será no nulo, y a ese será al que apliquemos g . □

Proposición 10.2. Si $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ es cualquier función y X, Y son variables aleatorias simples, entonces la función $g(X, Y)$ es una variable aleatoria simple.

Demostración. Tenemos que

$$X = \sum_{i=1}^m x_i I_{A_i}, \quad Y = \sum_{j=1}^n y_j I_{B_j}$$

donde se verifican en ambos casos todas las condiciones de las variables aleatorias simples.

Teniendo en cuenta que

$$A_i = \bigcup_j A_i \cap B_j \implies I_{A_i} = \sum_j I_{A_i \cap B_j}$$

(y de igual forma $I_{B_j} = \sum_i I_{A_i \cap B_j}$), se puede escribir

$$X = \sum_i \sum_j x_i I_{A_i \cap B_j}$$

$$Y = \sum_j \sum_i y_j I_{A_i \cap B_j}$$

De forma que, si $w \in A_i \cap B_j$, entonces

$$g(X(w), Y(w)) = g(x_i, y_j)$$

por lo que

$$g(X, Y) = \sum_i \sum_j g(x_i, y_j) I_{A_i \cap B_j}$$

ya que $\forall w \in \Omega, \exists i, j : w \in A_i \cap B_j$.

Si llamamos, para cada par i, j ,

$$C_r = A_i \cap B_j, \quad z_r = g(x_i, y_j), \quad r = n(i-1) + j = 1, 2, \dots, nm$$

se puede escribir

$$g(X, Y) = \sum_{r=1}^{nm} z_r I_{C_r}$$

donde $\{C_r\}_{r=1, \dots, nm}$ es una partición de Ω , por lo que $g(X, Y)$ es una variable aleatoria simple. \square

Como consecuencia particular obtenemos que si X, Y son variables aleatorias simples también lo son la suma, el producto, el máximo y el mínimo de estas variables. Además, aplicando inducción³, resulta que, dadas n variables aleatorias simples X_1, \dots, X_n , la suma, producto, máximo y mínimo de estas n variables nos dan variables aleatorias simples.

10.2. Esperanza matemática de variables aleatorias simples

Llamaremos **esperanza matemática de la variable aleatoria simple** $X = \sum_{i=1}^m x_i I_{A_i}$ a la suma

$$E(X) = \sum_{i=1}^m x_i P(A_i)$$

El siguiente teorema permite asegurar la unicidad de la esperanza matemática de una variable aleatoria simple:

Teorema 10.1. *La esperanza matemática $E(X)$ de una variable aleatoria simple depende solo de X como función de Ω en \mathbb{R} , no de la expresión con la que definamos la variable aleatoria X .*

³Si tenemos una familia de funciones $\{g_n : \mathbb{R}^n \rightarrow \mathbb{R}\}_{n \geq 2}$ tal que $\forall n > 2 \forall i \in \mathbb{N}, g_n(x_1, \dots, x_n) = g_2(x_i, g_{n-1}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n))$, entonces una trivial inducción revela que si

$$[X_i, X_j \text{ cumplen } C] \implies [g_2(X_i, X_j) \text{ cumple } C],$$

para C una condición cualquiera, entonces para todo subconjunto finito $\{i_1, \dots, i_n\} = I \subset \mathbb{N}$

$$[\forall i \in I, X_i \text{ cumple } C] \implies [g_n(X_{i_1}, \dots, X_{i_n}) \text{ cumple } C].$$

Demostración. Supongamos que X puede escribirse, además de como hemos venido haciendo, como

$$X = \sum_{j=1}^n y_j I_{B_j},$$

donde $y_j \in \mathbb{R}$ y los $B_j \in \mathcal{S}$ forman una partición de Ω . Entonces, tenemos dos valores para la esperanza de X :

$$E_1 = \sum_{i=1}^m x_i P(A_i)$$

$$E_2 = \sum_{j=1}^n y_j P(B_j)$$

ahora bien, se tiene $P(A_i) = \sum_j P(A_i \cap B_j)$ y $P(B_j) = \sum_i P(A_i \cap B_j)$, por lo que es

$$E_1 = \sum_i x_i \sum_j P(A_i \cap B_j) = \sum_i \sum_j x_i P(A_i \cap B_j)$$

$$E_2 = \sum_j y_j \sum_i P(A_i \cap B_j) = \sum_j \sum_i y_j P(A_i \cap B_j)$$

Y para cada par i, j se tiene

$$x_i P(A_i \cap B_j) = y_j P(A_i \cap B_j)$$

ya que si $A_i \cap B_j = \emptyset$, entonces ambos miembros de la igualdad son nulos; y si $A_i \cap B_j \neq \emptyset$, entonces $\{x_i\} = X(A_i \cap B_j) = \{y_j\}$.

Por lo tanto, $E_1 = E_2$ y la expresión es única. □

10.3. Propiedades de la esperanza matemática de variables aleatorias simples

Las propiedades fundamentales del cálculo con esperanzas matemáticas están contenidas en el siguiente enunciado:

Teorema 10.2. *Se verifica:*

1. Si X es una variable aleatoria simple y c una constante se tiene para la variable aleatoria simple cX

$$E(cX) = cE(X)$$

2. Si X_1, \dots, X_n son variables aleatorias simples, vale

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + \dots + E(X_n)$$

3. Si X, Y son variables aleatorias simples, entonces

$$X \leq Y \implies E(X) \leq E(Y)$$

4. Si X_1, \dots, X_n son variables aleatorias simples independientes, entonces se tiene

$$E(X_1 X_2 \dots X_n) = E(X_1) E(X_2) \dots E(X_n)$$

Demostración. Vamos a ver todas las afirmaciones

1. $cX = \sum_{i=1}^m cx_i I_{A_i}$ y, por tanto

$$E(cX) = \sum_{i=1}^m cx_i P(A_i) = c \sum_{i=1}^m x_i P(A_i) = cE(X)$$

2. Basta hacerlo para dos variables aleatorias simples X e Y , y aplicarlo un número finito de veces⁴. Se tiene, entonces

$$\begin{aligned} E(X) + E(Y) &= \sum_{i=1}^m x_i P(A_i) + \sum_j y_j P(B_j) = \\ &= \sum_i \sum_j x_i P(A_i \cap B_j) + \sum_j \sum_i y_j P(A_i \cap B_j) = \sum_{i,j} (x_i + y_j) P(A_i \cap B_j) = E(X + Y) \end{aligned}$$

3. Para $X = 0$ es inmediato. Es decir, si $Y \geq 0$, entonces $E(X) = \sum_i 0P(A_i) = 0 \leq \sum_j y_j P(B_j) = E(Y)$. Para ver el caso general, tenemos que $X \leq Y$, pero entonces $Y - X \geq 0$ y, por el caso anterior, $E(Y - X) \geq 0$, además, por 2), tenemos que $0 \leq E(Y - X) = E(Y) - E(X) \implies E(Y) \geq E(X)$.
4. Comenzamos haciéndolo para dos variables aleatorias independientes X, Y y hacemos inducción como en el caso de la suma. Además, sin pérdida de generalidad, vamos a suponer que $i \neq j \implies x_i \neq x_j$, e igual para Y . Y tomamos los A_i y B_j los mayores posibles:

$$A_i = \{X = x_i\}, B_j = \{Y = y_j\}$$

y así se tiene

$$X = \sum_i x_i I_{A_i}, Y = \sum_j y_j I_{B_j} \implies XY = \sum_{i,j} x_i y_j I_{A_i \cap B_j}, \text{ y}$$

$$P(A_i \cap B_j) = P(\{X = x_i\} \cap \{Y = y_j\}) \stackrel{X,Y \text{ indep.}}{=} P(X = x_i) P(Y = y_j) = P(A_i) P(B_j);$$

por tanto,

$$\begin{aligned} E(X)E(Y) &= \left(\sum_i x_i P(A_i) \right) \left(\sum_j y_j P(B_j) \right) = \sum_{i,j} x_i y_j P(A_i) P(B_j) = \\ &= \sum_{i,j} x_i y_j P(A_i \cap B_j) = E(XY) \end{aligned}$$

⁴Utilizando de nuevo la familia de funciones $\{g_n : \mathbb{R}^n \rightarrow \mathbb{R}\}_{n \geq 2}$ tal que $\forall n > 2 \forall i \in \mathbb{N}, g_n(x_1, \dots, x_n) = g_2(x_i, g_{n-1}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n))$, tenemos que si para todo subconjunto finito $\{i_1, \dots, i_n\} = I \subset \mathbb{N}$ se tiene

$$\begin{aligned} [X_i, X_j \text{ cumplen } C] &\implies [E(g_2(X_i, X_j)) = g_2(E(X_i), E(X_j))] \text{ y} \\ [\forall i \in I, X_i \text{ cumplen } C] &\implies [g_{n-1}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n), X_i \text{ cumplen } C], \end{aligned}$$

para C una condición cualquiera, entonces una trivial inducción nos lleva a

$$[\forall i \in I, X_i \text{ cumple } C] \implies [E(g_n(X_{i_1}, \dots, X_{i_n})) = g_2(E(X_{i_1}), \dots, E(X_{i_n}))].$$

Así, queda probado para $n = 2$. Para ver el caso general, supongamos que se verifica para $n - 1$ variables aleatorias simples independientes

$$E(X_1 X_2 \cdots X_{n-1}) = E(X_1) \cdots E(X_{n-1})$$

Entonces, si tenemos n variables aleatorias simples independientes, también son independientes las variables aleatorias

$$X_1 X_2 \cdots X_{n-1}, \text{ y } X_n$$

(6.11). Así, tenemos que

$$E(X_1 X_2 \cdots X_n) \stackrel{\text{indep. + caso } n=2}{=} E(X_1 X_2 \cdots X_{n-1}) E(X_n) \stackrel{\text{inducción}}{=} E(X_1) E(X_2) \cdots E(X_{n-1}) E(X_n)$$

□

La siguiente propiedad será útil en la próxima sección:

Teorema 10.3. *Sea Y una variable aleatoria simple no negativa y (X_n) una sucesión monótona creciente de variables aleatorias simples no negativas*

$$0 \leq X_n \leq X_{n+1}, \quad \forall n$$

y tales que

$$\lim_n X_n \geq Y$$

Entonces

$$\lim_n E(X_n) \geq E(Y)$$

Nótese que el límite $\lim_n X_n$ es una variable aleatoria que puede no ser simple.

Demostración. Por la definición de esperanza y la propiedad 2) del teorema anterior, se tiene

$$E(Y) = \sum_{j=1}^m y_j P(B_j)$$

$$E(X_n) = E\left(X_n \sum_{j=1}^m I_{B_j}\right) = E\left(\sum_{j=1}^m X_n I_{B_j}\right) = \sum_{j=1}^m E(X_n I_{B_j})$$

luego si demostramos que para cada $j = 1, \dots, m$ vale

$$\lim_n E(X_n I_{B_j}) \geq y_j P(B_j)$$

tendremos el resultado.

Así, fijado j , tomamos $\varepsilon > 0$. Sea

$$A_n := \{X_n > y_j - \varepsilon\} \cap B_j \subset B_j$$

Como X_n es monótona creciente, $A_n \subset A_{n+1}$; y como $\lim X_n \geq Y = y_j$ (en B_j) y $\exists n_0 : \forall n \geq n_0, A_n \neq \emptyset^5$, $\lim_n A_n = B_j$. Así, tenemos que

$$\lim_n P(A_n) = P(B_j)$$

Por otra parte, tenemos

$$E(X_n I_{B_j}) = E(X_n I_{A_n}) + E(X_n I_{B_j \setminus A_n}) \geq E(X_n I_{A_n}) \geq E((y_j - \varepsilon) I_{A_n}) = (y_j - \varepsilon) P(A_n)$$

y, al tomar límites, resulta que

$$\lim_n E(X_n I_{B_j}) \geq (y_j - \varepsilon) P(B_j)$$

y como el $\varepsilon > 0$ es arbitrario, tenemos el resultado. \square

10.4. Variables aleatorias no negativas

Sea $X : \Omega \rightarrow \overline{\mathbb{R}}$ una variable aleatoria no negativa, es decir, $X(\omega) \geq 0, \forall \omega \in \Omega$. Llamamos **esperanza matemática** de X al límite (finito o infinito)

$$\lim_n E(X_n)$$

donde (X_n) es una sucesión de variables aleatorias simples tales que

$$0 \leq X_n \leq X_{n+1}, n = 1, \dots$$

$$\lim_n X_n = X$$

Es necesario comprobar la existencia y unicidad de este concepto, además de su compatibilidad con la definición dada anteriormente para las variables aleatorias simples.

Para la existencia basta asegurarnos de que siempre existe una sucesión (X_n) de variables aleatorias simples que cumplen con las propiedades anteriores, ya que, en tal caso, el límite que define la esperanza de X existe en $\overline{\mathbb{R}}$ por ser $E(X_n)$ una sucesión numérica monótona creciente.

Para verlo, sean $\{g_n : \overline{\mathbb{R}} \rightarrow \overline{\mathbb{R}}\}_{n \in \mathbb{N}}, g : \overline{\mathbb{R}} \rightarrow \overline{\mathbb{R}}$ funciones simples definidas como

$$g_n(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{h}{2^n} & \text{si } \frac{h}{2^n} \leq x \leq \frac{h+1}{2^n} \text{ para } h \in \{0, 1, \dots, n2^n - 1\} \\ n & \text{si } x \geq n \end{cases}$$

$$g(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$$

Que presentan las siguientes propiedades:

1. $g_n(x) \geq 0$: obviamente
2. $g_n(x) \leq g_{n+1}(x)$: fijado x

⁵Si no tomásemos ε , podíamos pensar en definir $A_n := \{X_n \geq y_j\} \cap B_j$, pero ninguno de estos conjuntos tiene por qué ser no vacío y por tanto su límite podría no converger a B_j si este es no vacío. Ejemplo de este caso es la sucesión $X_n := y_j - 1/n$.

Por el contrario, sí podemos afirmar $\exists n_0 : \forall n \geq n_0, \exists \omega \in B_j : X_n(\omega) > y_j - \varepsilon$, pues esta es condición necesaria para que se de $\lim X_n \geq Y = y_j$.

- a) Si $x < 0 \implies g_n(x) = g_{n+1}(x) = 0$
 b) Si $0 \leq x < n$ y se tiene $\frac{h}{2^n} \leq x < \frac{h+1}{2^n}$, podemos escribir

$$\left[\frac{h}{2^n}, \frac{h+1}{2^n} \right) = \left[\frac{2h}{2^{n+1}}, \frac{2h+1}{2^{n+1}} \right) \cup \left[\frac{2h+1}{2^{n+1}}, \frac{2h+2}{2^{n+1}} \right)$$

por tanto

$$g_{n+1}(x) = \begin{cases} \frac{2h}{2^{n+1}} = \frac{h}{2^n} = g_n(x) & \text{si } \frac{2h}{2^{n+1}} \leq x < \frac{2h+1}{2^{n+1}} \\ \frac{2h+1}{2^{n+1}} \geq \frac{2h}{2^{n+1}} = \frac{h}{2^n} = g_n(x) & \text{si } \frac{2h+1}{2^{n+1}} \leq x < \frac{2h+2}{2^{n+1}} \end{cases}$$

- c) Si $n \leq x < n+1$ y $\frac{h}{2^{n+1}} \leq x < \frac{h+1}{2^{n+1}}$ resulta

$$n \leq x < \frac{h+1}{2^{n+1}} \implies n2^{n+1} < h+1 \stackrel{h,n \text{ enteros}}{\implies} n2^{n+1} \leq h \implies g_n(x) = n \leq \frac{h}{2^{n+1}} = g_{n+1}(x)$$

- d) Si $x \geq n+1 \implies g_n(x) = n < n+1 = g_{n+1}(x)$, y la propiedad 2) queda probada.

3. $g_n(x) \rightarrow g(x)$: sea $x > 0$, se tiene

$$h := [x2^n] \leq x2^n < h+1 \implies \frac{h}{2^n} \leq x < \frac{h+1}{2^n}$$

y entonces, para los enteros $n > x$,

$$g_n(x) = \frac{h}{2^n} \leq x < \frac{h+1}{2^n} = g_n(x) + \frac{1}{2^n}$$

haciendo $n \rightarrow \infty$, queda $\lim_n g_n(x) = x = g(x)$.

Ahora vamos a ver que para una variable aleatoria no negativa, X , podemos encontrar una sucesión de variables aleatorias simples con límite X .

Proposición 10.3. Dada una variable aleatoria X no negativa, la sucesión (X_n) de variables aleatorias simples definidas por

$$X_n = g_n(X),$$

donde g_n son las funciones anteriormente definidas, es monótona creciente, de términos no negativos y converge a X . Es decir,

$$0 \leq X_n \leq X_{n+1}$$

$$\lim X_n = X$$

Demostración. Por la propiedad 1, $X_n = g_n(X) \geq 0$.

Por la propiedad 2, $X_n = g_n(X) \leq g_{n+1}(X) = X_{n+1}$.

Por la propiedad 3, $\lim_n X_n = \lim_n g_n(X) = g(X) = X$ □

Por tanto, la esperanza matemática de una variable aleatoria no negativa existe siempre, finita o infinita.

Ahora bien, si en la definición de las g_n se cambian las potencias 2^n por m^n , con $m \in \mathbb{N}$, se obtienen para X_n sucesiones de variables aleatorias simples que verifican las propiedades requeridas. Por lo tanto, debemos estudiar la unicidad del límite con el que definimos la esperanza matemática.

Proposición 10.4. Dada una variable aleatoria no negativa X , el límite

$$\lim_n E(X_n)$$

es independiente de la sucesión (X_n) siempre que esta verifique

$$0 \leq X_n \leq X_{n+1}$$

$$\lim_n X_n = X$$

Demostración. Supongamos que tenemos dos sucesiones (X_n) e (Y_n) que satisfacen estas propiedades. Entonces

$$E_1 = \lim_n E(X_n), \quad E_2 = \lim_n E(Y_n)$$

De las condiciones que verifican X_n, Y_n , se tiene que, para todo m

$$\lim_n X_n \geq Y_m$$

entonces, por el teorema 10.5 (10.3), se tiene

$$\lim_n E(X_n) \geq E(Y_m) \implies E_1 = \lim_n E(X_n) \geq \lim_n E(Y_m) = E_2$$

y de forma análoga $E_2 \geq E_1$, por lo que ambos valores deben ser iguales y queda demostrada la unicidad de este valor. \square

Por último, debemos comprobar que esta nueva definición de esperanza matemática es compatible con la definición dada para variables aleatorias simples.

Proposición 10.5. Si X es una variable aleatoria simple no negativa el valor de $E(X)$ obtenido con arreglo a la definición de esperanza matemática para variables aleatorias simples coincide con el valor obtenido con la definición dada para variables aleatorias no negativas.

Demostración. Si X es una variable aleatoria simple no negativa, para calcular la esperanza matemática de X como variable aleatoria no negativa según la nueva definición, tomamos la sucesión constante

$$X_n = X$$

que claramente verifica las propiedades requeridas y se tiene

$$\lim_n E(X_n) = \lim_n E(X) = E(X)$$

por lo que ambas definiciones arrojan el mismo valor a la esperanza matemática de las variables aleatorias simples no negativas. \square

A partir de ahora, para cualquier variable aleatoria no negativa utilizaremos la notación $E(X)$ para denotar su esperanza matemática, lo que no es ambiguo, como hemos comprobado.

10.5. Propiedades de la esperanza matemática para variables aleatorias no negativas

Teorema 10.4. *Se verifican las siguientes propiedades:*

1. Si X es una variable aleatoria no negativa y c una constante no negativa se tiene para la variable aleatoria cX

$$E(cX) = cE(X)$$

2. Si X_1, \dots, X_n son variables aleatorias no negativas se tiene

$$E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n)$$

3. Si X, Y son variables aleatorias no negativas,

$$X \leq Y \implies E(X) \leq E(Y)$$

4. Si X_1, \dots, X_n son variables aleatorias no negativas independientemente, entonces para el producto $X_1 X_2 \cdots X_n$ se tiene

$$E(X_1 X_2 \cdots X_n) = E(X_1) E(X_2) \cdots E(X_n)$$

Demostración. Vamos a ver todas las propiedades:

1. Tomamos X_n variables aleatorias simples tales que

$$0 \leq X_n \leq X_{n+1}, \forall n$$

$$\lim_n X_n = X$$

entonces

$$E(cX_n) = cE(X_n) \implies E(cX) = \lim_n E(cX_n) = \lim_n cE(X_n) = c \lim_n E(X_n) = cE(X)$$

2. Para dos variables aleatorias X, Y tomamos las respectivas sucesiones X_n e Y_n de variables aleatorias simples que verifican las propiedades requeridas, y se tiene

$$E(X + Y) = \lim_n E(X_n + Y_n) = \lim_n [E(X_n) + E(Y_n)] = \lim_n E(X_n) + \lim_n E(Y_n) = E(X) + E(Y)$$

y esto se generaliza para cualquier número finito de sumandos.

3. Si $Y \geq 0$, es claro que $E(Y) \geq 0$. Ahora bien, si $X \leq Y$, entonces

$$E(Y) = E(Y - X + X) = E(Y - X) + E(X) \stackrel{Y-X \geq 0}{\geq} E(X)$$

4. Consideremos dos variables aleatorias independientes no negativas X, Y . Definimos, entonces, usando las g_n definidas en la sección anterior:

$$X_n = g_n(X), Y_n = g_n(Y)$$

que verifican, de nuevo, las propiedades requeridas (10.3). Como X e Y son independientes, también lo son cualquier función medible de X con cualquier función medible de Y (6.8), en particular X_n e Y_n son independientes (las g_n son medibles por ser simples), por lo que, por las propiedades de la esperanza matemática para variables aleatorias simples, tenemos

$$E(X_n Y_n) = E(X_n) E(Y_n)$$

y, tomando límites:

$$E(XY) = E(X) E(Y)$$

Y esto se generaliza para cualquier número finito de factores.

□

Teorema 10.5. *Si X es una variable aleatoria no negativa entonces*

$$P(X > 0) = 0 \iff E(X) = 0$$

$$P(X > 0) > 0 \iff E(X) > 0$$

Demostración. Basta ver que

$$P(X > 0) = 0 \implies E(X) = 0$$

$$P(X > 0) > 0 \implies E(X) > 0$$

ya que las otras implicaciones son los contrarrecíprocos de la otra afirmación.

Supongamos que $P(X > 0) = 0$. Sea

$$X_1 = \sum_i x_i I_{A_i}$$

una variable aleatoria simple donde suponemos sin pérdida de generalidad $A_i \neq \emptyset$, $\forall i$, ya que los sumandos con $A_i = \emptyset$ pueden suprimirse.

Si vale $0 \leq X_1 \leq X$ se cumple que $x_i \geq 0$, $\forall i$. Además, si para algún i se tiene $x_i > 0$, entonces $P(A_i) = 0$, pues de lo contrario se tendría

$$A_i \subset \{X > 0\} \implies 0 < P(A_i) \leq P(X > 0) = 0 \#$$

Por tanto, tenemos que

$$E(X_1) = \sum x_i P(A_i) = 0$$

Por tanto, si ciertos X_n forman una sucesión de variables aleatorias simples no negativas que convergen a X , todas tienen esperanza nula y el límite de las esperanzas es $E(X) = 0 \checkmark$.

Si $P(X > 0) > 0$, como

$$\lim_{\varepsilon \downarrow 0} P(X > \varepsilon) = P\left(\bigcup_{\varepsilon > 0} \{X > \varepsilon\}\right) = P(X > 0) > 0,$$

entonces existe un valor $\varepsilon > 0$ tal que $P(X > \varepsilon) > 0$.

Definimos ahora la variable

$$U = \varepsilon I_{\{X > \varepsilon\}} \leq X$$

que verifica

$$E(U) = \varepsilon P(X > \varepsilon) > 0$$

y como

$$X \geq U \implies E(X) \geq E(U) > 0 \checkmark$$

□

Teorema 10.6. *Si la variable aleatoria X no negativa tiene esperanza finita entonces X es casi seguramente finita, o sea*

$$E(X) < \infty \implies P(X = \infty) = 0$$

Demostración. Sea $A = \{X = \infty\}$. Entonces, para cualquier entero positivo n , se tiene

$$nI_A \leq X \implies nP(A) \leq E(X)$$

entonces, si $E(X)$ es finita

$$P(A) \leq \frac{E(X)}{n} \xrightarrow{n \rightarrow \infty} 0$$

y debe ser $P(A) = 0$

□

10.6. Esperanza matemática de variables aleatorias con valores reales

Para cualquier variable aleatoria X definiremos su parte positiva

$$X^+ := \max\{0, X\},$$

y su parte negativa

$$X^- := \max\{0, -X\}.$$

Con estas definiciones se tiene

$$X = X^+ - X^-, \quad |X| = X^+ + X^-, \quad (-X)^+ = X^-, \quad (-X)^- = X^+.$$

Dada una variable aleatoria cualquiera $X : \Omega \rightarrow \overline{\mathbb{R}}$ llamaremos **esperanza matemática** de X a la expresión

$$E(X) = E(X^+) - E(X^-)$$

excluido únicamente el caso en que $E(X^+)$ y $E(X^-)$ tomen a la vez el valor infinito, en cuyo caso la esperanza de X no existe. Es inmediato comprobar que esta definición da un valor único en $\overline{\mathbb{R}}$ y es compatible con las definiciones de esperanza anteriores, por lo que se empleará el mismo símbolo $E(X)$.

Veamos el siguiente Teorema auxiliar

Teorema 10.7. Sea X una variable aleatoria cualquiera. Si existen las variables aleatorias no negativas U, V tales que

$$X = U - V,$$

entonces

$$X^+ \leq U, \quad X^- \leq V.$$

Además, si alguna de las variables U, V tiene esperanza finita entonces la esperanza de X existe y vale

$$E(X) = E(U) - E(V)$$

Demostración. Sea $A = \{X > 0\}, B = \{X < 0\}$. Entonces vale

$$I_A X^- = 0, \quad I_B X^+ = 0.$$

Multiplicando la igualdad

$$X = X^+ - X^- = U - V$$

por I_A y $-I_B$ obtenemos las dos relaciones

$$\begin{aligned} X^+ &= I_A X = I_A U - I_A V \leq I_A U \leq U \\ X^- &= -I_B X = -I_B U + I_B V \leq I_B V \leq V. \end{aligned}$$

De estas relaciones se sigue que si $E(U)$ es finita también lo es $E(X^+)$ y análogamente si $E(V)$ es finita, entonces $E(X^-)$ lo es.

De $X = X^+ - X^- = U - V$ también tenemos $X^+ + V = U + X^-$ y como todas son no negativas, tomando esperanzas a cada lado de la igualdad podemos afirmar

$$E(X^+) + E(V) = E(U) + E(X^-).$$

Si $E(U)$ o $E(V)$ son finitas, entonces (respectivamente) $E(X^+)$ o $E(X^-)$ lo son y podemos llegar (restando las componentes finitas a cada lado y cambiando el signo a la igualdad final si hiciese falta) a

$$E(X) := E(X^+) - E(X^-) = E(U) - E(V).$$

□

10.7. Propiedades de la esperanza matemática

Completada la definición de esperanza matemática para variables reales el siguiente Teorema generaliza las *propiedades fundamentales* vistas anteriormente para el cálculo con esperanzas matemáticas.

Teorema 10.8. *Propiedades de la esperanza matemática para variables reales:*

1. Si X posee esperanza y $c \in \mathbb{R}$ es una constante, entonces cX posee esperanza y vale

$$E(cX) = cE(X).$$

2. Si X_1, X_2, \dots, X_n son variables aleatorias que poseen esperanzas y tienen sentido las sumas

$$\begin{aligned} E(X_1) + E(X_2) + \dots + E(X_n) \\ X_1 + X_2 + \dots + X_n, \end{aligned}$$

es decir, siempre que en las mismas no aparezcan dos sumandos infinitos de signos opuestos, entonces

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n).$$

3. Si X, Y poseen esperanzas y vale $X \leq Y$ entonces

$$E(X) \leq E(Y).$$

4. Si X_1, X_2, \dots, X_n son variables aleatorias reales independientes con esperanzas finitas entonces el producto $X_1 X_2 \dots X_n$ tiene esperanza finita y vale

$$E(X_1 X_2 \dots X_n) = E(X_1) E(X_2) \dots E(X_n).$$

Demostración. Veamos cada propiedad:

1. X posee esperanza $\iff E(X^+)$ o $E(X^-)$ finitos $\iff |c| E(X^+) \stackrel{|c|, X^+ \geq 0}{=} E(|c| X^+)$ o $|c| E(X^-) \stackrel{|c|, X^- \geq 0}{=} E(|c| X^-)$ finitos.

Consideremos primero $c \geq 0$. Tenemos $cX = cX^+ - cX^-$, $cX^+ \geq 0$, $cX^- \geq 0$ y por el teorema 10.7 tenemos

$$E(cX) = E(cX^+) - E(cX^-) \stackrel{c, X^+, X^- \geq 0}{=} c(E(X^+) - E(X^-)) = cE(X).$$

Si por el contrario tenemos $c < 0$, entonces tendremos $cX = |c| X^- - |c| X^+$, $|c| X^- \geq 0$, $|c| X^+ \geq 0$ y de nuevo por el teorema 10.7 tenemos

$$E(cX) = E(|c| X^-) - E(|c| X^+) \stackrel{|c|, X^+, X^- \geq 0}{=} |c|(E(X^-) - E(X^+)) = -|c| E(X) = cE(X).$$

2. Tomando $U := (X_1^+ + X_2^+ + \dots + X_n^+) \geq 0$ y $V := (X_1^- + X_2^- + \dots + X_n^-) \geq 0$ es claro que se dan las igualdades

$$X_1 + X_2 + \dots + X_n = U - V$$

$$E(U) = \sum_{i=1}^n E(X_i^+)$$

$$E(V) = \sum_{i=1}^n E(X_i^-).$$

Además, por cómo hemos exigido que sean las sumas de las esperanzas $E(X_i)$ (que no aparezan dos sumandos infinitos de signos opuestos), podemos asegurar que o bien $E(U)$ es finito o bien $E(V)$ lo es, y así podemos aplicar de nuevo el teorema 10.7 obteniendo

$$\begin{aligned} E(X_1 + X_2 + \cdots + X_n) &= E(U) - E(V) = \sum_{i=1}^n E(X_i^+) - \sum_{i=1}^n E(X_i^-) = \\ &= \sum_{i=1}^n (E(X_i^+) - E(X_i^-)) = E(X_1) + E(X_2) + \cdots + E(X_n). \end{aligned}$$

3. Sea $X \leq Y$. Si $E(X) = -\infty$ es trivial. Si es finita o $+\infty$, entonces

$$E(Y) = E(X) + E(Y - X) \stackrel{Y-X \geq 0}{\geq} E(X).$$

4. Comencemos considerando el caso de dos variables reales X, Y independientes y con esperanzas finitas. Por tanto tienen esperanzas finitas X^+, X^-, Y^+, Y^- .

Si definimos las funciones monótonas $f: \mathbb{R} \rightarrow \mathbb{R}$ y $g: \mathbb{R} \rightarrow \mathbb{R}$ dadas por

$$f(x) := \max\{0, x\}, \quad g(x) = \max\{0, -x\},$$

que son medibles por ser intervalos las preimágenes de intervalos. Vale $f(X) = X^+, g(X) = X^-, f(Y) = Y^+, g(Y) = Y^-$. Como X, Y se suponen independientes, X^+, X^- son ambos independientes tanto de Y^+ como de Y^- . Así, el producto XY se descompone en

$$XY = (X^+ - X^-)(Y^+ - Y^-) = (X^+Y^+ + X^-Y^-) - (X^-Y^+ + X^+Y^-).$$

Así, tomando $U := (X^+Y^+ + X^-Y^-) \geq 0$ y $V := (X^-Y^+ + X^+Y^-) \geq 0$ y teniendo en cuenta que

$$\begin{aligned} E(U) &= E(X^+)E(Y^+) + E(X^-)E(Y^-) < \infty \\ E(V) &= E(X^-)E(Y^+) + E(X^+)E(Y^-) < \infty, \end{aligned}$$

podemos utilizar de nuevo el teorema 10.7 para asegurar que la esperanza de XY vale

$$\begin{aligned} E(XY) &= E(U) - E(V) = \\ &= E(X^+)E(Y^+) + E(X^-)E(Y^-) - E(X^-)E(Y^+) - E(X^+)E(Y^-) = \\ &= E(X^+)(E(Y^+) - E(Y^-)) + E(X^-)(E(Y^-) - E(Y^+)) = \\ &= (E(X^+) - E(X^-))(E(Y^+) - E(Y^-)) = E(X)E(Y). \end{aligned}$$

Por inducción (aprovechando la asociatividad del producto) podemos generalizar a n variables independientes con esperanzas finitas. □

Teorema 10.9. *Si la variable aleatoria X posee esperanza y A es un suceso cualquiera, también XI_A posee esperanza.*

Demostración. Sea $X = X^+ - X^-$. Es inmediato que

$$XI_A = X^+I_A - X^-I_A, \quad X^+I_A \leq X^+, \quad X^-I_A \leq X^-,$$

por lo que o bien $E(X^+I_A)$ es finita (si lo es $E(X^+)$) o bien lo es $E(X^-I_A)$ (si lo es $E(X^-)$), de forma que XI_A tiene esperanza.⁶ \square

10.8. Variables casi seguramente (c. s.) iguales

Recordemos que denotamos que una cierta condición C es casi segura (c.s.) cuando $P(C) = 1$. Veamos algunas *propiedades complementarias* de la esperanza matemática relacionadas con este concepto aplicado a igualdades.

Teorema 10.10. *Propiedades complementarias de la esperanza matemática:*

1. Si la variable aleatoria $X = 0$ c.s., entonces $E(X) = 0$.
2. Si $X = Y$ c.s. y una de las variables X, Y posee esperanza entonces las dos poseen esperanza y vale $E(X) = E(Y)$.

Demostración. Veamos.

1. Como $0 = P(X \neq 0) = P(X > 0) + P(X < 0)$ resulta

$$\begin{aligned} 0 &= P(X > 0) = P(X^+ > 0) \\ 0 &= P(X < 0) = P(X^- > 0), \end{aligned}$$

de donde obtenemos, por el teorema 10.5,

$$E(X) = E(X^+) - E(X^-) = 0 - 0 = 0.$$

2. Como vale $X - Y = 0$ c.s., por el apartado anterior es $E(X - Y) = 0$ y si X (podría ser Y) posee esperanza, entonces $Y = X + (Y - X)$ también posee esperanza y vale

$$E(Y) = E(X) + E(Y - X) = E(X).$$

\square

Por el segundo apartado de este teorema, si X es una variable aleatoria con esperanza $E(X)$ y se cambia el valor de la variable en un suceso A de probabilidad cero, la esperanza matemática de la nueva variable no cambia. Por esta razón no nos preocupa que X pueda no estar definida en este tipo de sucesos a la hora de calcular su esperanza (por convenio le asignamos un valor arbitrario).

Teorema 10.11. *Si la variable aleatoria X tiene esperanza finita entonces X es casi seguramente finita.*

$$E(X) \in \mathbb{R} \implies P_X(-\infty) = P_X(\infty) = 0$$

⁶Si no calcula la esperanza, será por algo ;).

Demostración. Usando el teorema 10.6, tenemos que $E(X) \in \mathbb{R} \iff E(X^+) < \infty \wedge E(X^-) < \infty \implies P(X^+ = \infty) = P(X^- = \infty) = 0$. Por otro lado, es inmediato que $\{X = \infty\} = \{X^+ = \infty\}$ y $\{X = -\infty\} = \{X^- = \infty\}$. Juntando ambas afirmaciones obtenemos el resultado. \square

Solo excepcionalmente se utilizan variables aleatorias que toman valor infinito con probabilidad positiva. Por ello supondremos mientras no se diga lo contrario que las variables aleatorias que se utilicen son c.s. finitas. Estas se pueden sustituir por variables aleatorias finitas sin que sus esperanzas matemáticas cambien.

11. Esperanza Matemática (continuación)

11.1. Variables aleatorias con valores complejos

Si X, Y son dos variables aleatorias con valores reales, la expresión

$$Z = X + iY$$

define una **variable aleatoria con valores complejos**

$$Z : \Omega \rightarrow \mathbb{C}$$

por la relación

$$Z(w) = X(w) + iY(w), \forall w \in \Omega$$

Para esta variable aleatoria Z se define la **esperanza matemática** por

$$E(Z) = E(X) + iE(Y)$$

si las dos esperanzas $E(X), E(Y)$ existen con valor finito.

Decimos que las n variables aleatorias con valores complejos

$$Z_r = X_r + iY_r, \quad r = 1, \dots, n$$

son **independientes** si lo son las n variables aleatorias bidimensionales (X_r, Y_r) .

Teorema 11.1. *Se verifican:*

1. Si c es un número complejo y la variable aleatoria Z posee esperanza, vale

$$E(cZ) = cE(Z)$$

2. Si las variables aleatorias de valores complejos Z_1, \dots, Z_n poseen esperanza, se tiene

$$E(Z_1 + \dots + Z_n) = E(Z_1) + \dots + E(Z_n)$$

3. Para cualquier variable aleatoria con valores complejos que posea esperanza se tiene

$$|E(Z)| \leq E(|Z|)$$

4. Si las variables aleatorias de valores complejos Z_1, \dots, Z_n son independientes y poseen esperanza se tiene

$$E(Z_1 Z_2 \cdots Z_n) = E(Z_1) E(Z_2) \cdots E(Z_n)$$

Demostración. Veamos cada afirmación:

1. Sea $c = c_1 + c_2 i$

$$\begin{aligned} E(cZ) &= E((c_1 + c_2 i)(X + iY)) = E((c_1 X - c_2 Y) + i(c_2 X + c_1 Y)) = \\ &= E(c_1 X - c_2 Y) + iE(c_2 X + c_1 Y) = E(c_1 X) - E(c_2 Y) + iE(c_2 X) + iE(c_1 Y) = \\ &= c_1 E(X) - c_2 E(Y) + c_2 E(X) i + c_1 E(Y) i = (c_1 + c_2 i)(E(X) + iE(Y)) = cE(Z) \end{aligned}$$

2. Para el caso de dos variables aleatorias $Z_1 = X_1 + iY_1$, $Z_2 = X_2 + iY_2$:

$$\begin{aligned} E(Z_1 + Z_2) &= E((X_1 + iY_1) + (X_2 + iY_2)) = E((X_1 + X_2) + i(Y_1 + Y_2)) = \\ &= E(X_1 + X_2) + iE(Y_1 + Y_2) = E(X_1) + E(X_2) + iE(Y_1) + iE(Y_2) = \\ &= (E(X_1) + iE(Y_1)) + (E(X_2) + iE(Y_2)) = E(Z_1) + E(Z_2) \end{aligned}$$

y podemos generalizar para n sumandos (o hacerlo pa' n sumandos directamente).

3. El caso en que

$$E(Z) = E(X) + iE(Y) = 0$$

es evidente. Excluido este caso, sean $A := E(X)$, $B := E(Y)$, entonces

$$AX + BY \leq |A||X| + |B||Y| \stackrel{*}{\leq} \sqrt{A^2 + B^2} \sqrt{X^2 + Y^2}$$

para ver *:

$$\begin{aligned} |A||X| + |B||Y| \leq \sqrt{A^2 + B^2} \sqrt{X^2 + Y^2} &\iff (|A||X| + |B||Y|)^2 \leq (A^2 + B^2)(X^2 + Y^2) \iff \\ &\iff A^2X^2 + B^2Y^2 + 2|AXB Y| \leq A^2X^2 + A^2Y^2 + B^2X^2 + B^2Y^2 \iff \\ &\iff 2|AXB Y| \leq A^2Y^2 + B^2X^2 \iff 0 \leq A^2Y^2 - 2|AYBX| + B^2X^2 \iff \\ &\iff 0 \leq (|A||Y| - |B||X|)^2 \checkmark \end{aligned}$$

Si tomamos esperanzas en la desigualdad de arriba, obtenemos

$$A^2 + B^2 \leq \sqrt{A^2 + B^2} E(\sqrt{X^2 + Y^2})$$

y, como

$$\begin{aligned} |E(Z)| &= |A + iB| = \sqrt{A^2 + B^2} \\ |Z| &= |X + iY| = \sqrt{X^2 + Y^2}, \end{aligned}$$

entonces queda

$$|E(Z)|^2 \leq |E(Z)| E(|Z|) \implies |E(Z)| \leq E(|Z|)$$

4. Primero lo hacemos para dos variables aleatorias independientes Z_1, Z_2 . Entonces, dado que la independencia de Z_1 y Z_2 implica⁷ la independencia de X_1 e Y_1 con X_2 e Y_2 , se tiene

$$\begin{aligned} E(Z_1 Z_2) &= E((X_1 + iY_1)(X_2 + iY_2)) = E(X_1 X_2 + iX_1 Y_2 + iX_2 Y_1 - Y_1 Y_2) = \\ &= E((X_1 X_2 - Y_1 Y_2) + i(X_1 Y_2 + X_2 Y_1)) = E(X_1 X_2 - Y_1 Y_2) + iE(X_1 Y_2 + X_2 Y_1) = \\ &= E(X_1 X_2) - E(Y_1 Y_2) + iE(X_1 Y_2) + iE(X_2 Y_1) = \\ &= E(X_1) E(X_2) - E(Y_1) E(Y_2) + iE(X_1) E(Y_2) + iE(X_2) E(Y_1) = \\ &= (E(X_1) + iE(Y_1))(E(X_2) + iE(Y_2)) = E(Z_1) E(Z_2) \end{aligned}$$

y para n variables se hace como hicimos en 4.

□

⁷Dadas dos variables aleatorias n -dimensionales independientes $X := (X_1, X_2, \dots, X_n)$ e $Y := (Y_1, Y_2, \dots, Y_n)$, cualquier X_i es independiente de cualquier Y_j , puesto que $X_i = \pi_i(X)$, $Y_j = \pi_j(Y)$ son composiciones de funciones medibles (las proyecciones) con variables aleatorias y el teorema 6.8 nos dice que estas composiciones son independientes.

11.2. Teoremas de límite

Teorema 11.2. Teorema de la convergencia monótona

Si (X_n) es una sucesión monótona creciente de variables aleatorias no negativas, vale

$$\lim E(X_n) = E(\lim X_n)$$

Teorema 11.3. Lema de Fatou

1. Si (X_n) es una sucesión de variables aleatorias no negativas, se tiene

$$0 \leq E(\liminf X_n) \leq \liminf E(X_n)$$

2. Si (X_n) es una sucesión de variables aleatorias, Y una variable aleatoria para la que existe esperanza $E(Y) > -\infty$ y $X_n \geq Y, \forall n$, se tiene

$$E(Y) \leq E(\liminf X_n) \leq \liminf E(X_n)$$

3. Si (X_n) es una sucesión de variables aleatorias, Y una variable aleatoria para la que existe esperanza $E(Y) < +\infty$ y $X_n \leq Y, \forall n$, se tiene

$$E(Y) \geq E(\limsup X_n) \geq \limsup E(X_n)$$

Teorema 11.4. Teorema de la convergencia dominada de Lebesgue

Si (X_n) es una sucesión de variables aleatorias reales tales que

$$|X_n| \leq Y$$

donde Y es una variable aleatoria con esperanza finita, y existe el límite

$$\lim X_n = X$$

entonces

$$\lim E(X_n) = E(X)$$

Teorema 11.5. Sea (A_n) una sucesión de sucesos incompatibles, $A_i \cap A_j = \emptyset, \forall i \neq j$.

Entonces, para cualquier variable aleatoria X que posea esperanza se tiene

$$E(XI_B) = \sum_{i=1}^{\infty} E(XI_{A_i}), \quad B = \bigcup_{i=1}^{\infty} A_i$$

Demostración. Sea

$$B_n = \bigcup_{i=1}^n A_i$$

Entonces

$$E(X^+ I_{B_n}) = \sum_{i=1}^n E(X^+ I_{A_i})$$

$$E(X^- I_{B_n}) = \sum_{i=1}^n E(X^- I_{A_i})$$

y, por el Teorema de la Convergencia Monótona 11.2:

$$E(X^+ I_B) = \sum_{i=1}^{\infty} E(X^+ I_{A_i})$$

$$E(X^- I_B) = \sum_{i=1}^{\infty} E(X^- I_{A_i})$$

Luego

$$E(X I_B) = E(X^+ I_B) - E(X^- I_B) = \sum_{i=1}^{\infty} E(X^+ I_{A_i}) - \sum_{i=1}^{\infty} E(X^- I_{A_i})$$

Que es una diferencia de dos series de términos no negativos, tales que o bien las dos son convergentes, o una es convergente y la otra no (porque X posee esperanza, y no podemos tener que sea $\infty - \infty$). Entonces, podemos agrupar los sumandos:

$$E(X I_B) = \sum_{i=1}^{\infty} [E(X^+ I_{A_i}) - E(X^- I_{A_i})] = \sum_{i=1}^{\infty} E(X I_{A_i})$$

que convergerá si las dos series anteriores eran convergentes, y divergerá si una de ellas no lo era. \square

Si X es una variable aleatoria de tipo discreto, se tiene

$$X = \sum x_i I_{A_i}$$

donde el segundo miembro puede ser una suma finita o una serie. Si X posee esperanza, esta siempre viene dada por la suma o serie

$$E(X) = \sum x_i P(A_i)$$

11.3. Integral de Lebesgue abstracta (material complementario)

El concepto de esperanza matemática expuesto se construye disponiendo de un espacio de probabilidad (Ω, \mathcal{S}, P) y una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$.

Por el mismo método constructivo se obtiene el concepto de **integral de Lebesgue abstracta**, para lo cual necesitamos un **espacio de medida** $(\Omega, \mathcal{F}, \mu)$ donde \mathcal{F} es una σ -álgebra en Ω y $\mu : \mathcal{F} \rightarrow \overline{\mathbb{R}}$ es una medida en Ω y una **función medible Borel**

$$f : \Omega \rightarrow \overline{\mathbb{R}}$$

o sea, tal que

$$f^{-1}(B) \in \mathcal{F}, \forall B \in \overline{\mathcal{B}}$$

La definición de la integral se desarrolla en las tres etapas siguientes:

1. Para una función simple no negativa

$$f = \sum_{i=1}^m x_i I_{A_i}, \quad x_i \geq 0, \quad A_i \in \mathcal{F}, \quad A_i \cap A_j = \emptyset, \quad \forall i \neq j, \quad \text{y} \quad \bigcup_i A_i = \Omega$$

se define la integral de f por

$$\int f d\mu = \sum_{i=1}^m x_i \mu(A_i)$$

La razón de exigir que $f \geq 0$ se debe a que hay que evitar que puedan aparecer en la suma del segundo miembro dos términos $x_i \mu(A_i)$ con $\mu(A_i) = \infty$, uno con $x_i > 0$ y otro con $x_i < 0$.

2. Para cualquier función medible f no negativa se define la integral de f por

$$\int f d\mu = \lim \int f_n d\mu$$

donde (f_n) es una sucesión monótona creciente de funciones simples no negativas que convergen a f .

3. Si f es una función medible cualquiera y

$$f^+ = \max\{f, 0\}, \quad f^- = \max\{-f, 0\}$$

son sus partes positiva y negativa, se define la integral de f por

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu$$

siempre que estas dos últimas no sean infinitas.

De este modo se obtendrán para la integral propiedades análogas a las vistas para la esperanza matemática. Algunas notaciones que se utilizan para esta integral son

$$\int f d\mu = \int f(w) d\mu = \int f(w) d\mu(\{w\}) = \int f(w) \mu(\{dw\}) = \int_{\Omega} f d\mu$$

La siguiente relación muestra las principales **propiedades de la integral**, en las que se supone que existen las integrales y tienen sentido las operaciones que contienen:

▪ **Linealidad**

$$\int (c_1 f_1 + c_2 f_2) d\mu = c_1 \int f_1 d\mu + c_2 \int f_2 d\mu$$

▪ **Monotonía**

$$f \leq g \implies \int f d\mu \leq \int g d\mu$$

▪ **Teorema de la convergencia monótona**

$$0 \leq f_n \leq f_{n+1}, \quad \lim f_n = f \implies \lim \int f_n d\mu = \int f d\mu$$

- **Fatou**

$$0 \leq f_n \implies \int \liminf f_n d\mu \leq \liminf \int f_n d\mu$$

- **Teorema de la convergencia dominada**

$$|f_n| \leq g, \quad \int g d\mu < \infty, \quad \lim f_n = f \implies \lim \int f_n d\mu = \int f d\mu$$

-

$$f = g, \text{ casi por todo} \implies \int f d\mu = \int g d\mu$$

-

$$\int f d\mu \text{ finita} \implies f \text{ finita casi por todo}$$

11.4. Casos particulares

11.4.1. Esperanza matemática

Si el espacio de medida es un espacio de probabilidad (Ω, \mathcal{S}, P) la integral de Lebesgue de una variable aleatoria es la esperanza matemática, por lo que se puede utilizar la notación de integral

$$E(X) = \int X dP = \int X(w) dP(\{w\}) = \int_{\Omega} X(w) dP(\{w\})$$

Si el espacio de probabilidad fuese en \mathbb{R} , $(\mathbb{R}, \mathcal{B}_1, P)$ y esta distribución tuviese la función de distribución $F : \mathbb{R} \rightarrow \mathbb{R}$, para la variable aleatoria $X(x) = x$ se tendría

$$E(X) = \int x dP = \int x dF(x) = \int_{\mathbb{R}} x dF(x)$$

y para la variable aleatoria $X : \mathbb{R} \rightarrow \overline{\mathbb{R}}$, donde X es medible Borel

$$E(X) = \int X(x) dP = \int X(x) dF(x) = \int_{\mathbb{R}} X(x) dF(x)$$

Fórmulas análogas valen si se tiene una distribución en $(\mathbb{R}^k, \mathcal{B}_k, P)$ con función de distribución $F : \mathbb{R}^k \rightarrow \mathbb{R}$. Así, si $X : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$ es medible Borel, se tiene

$$E(X) = \int X(x) dP = \int X(x) dF(x) = \int_{\mathbb{R}^k} X(x_1, \dots, x_k) dF(x_1, \dots, x_k)$$

11.5. Medida inducida por una transformación del espacio

El proceso seguido para definir la integral de Lebesgue suministra un método de comprobar la igualdad de ciertas integrales, como vamos a ver.

Sean $(\Omega, \mathcal{F}, \mu)$ un espacio de medida, $(\Omega_1, \mathcal{F}_1)$ un espacio medible y $f : \Omega \rightarrow \Omega_1$ una función medible ($B \in \mathcal{F}_1 \implies f^{-1}(B) \in \mathcal{F}$). Esta función genera una medida μ_1 en Ω_1 por la relación

$$\mu_1(B) = \mu(f^{-1}(B)), \quad \forall B \in \mathcal{F}_1$$

con lo que obtenemos el espacio de medida $(\Omega_1, \mathcal{F}_1, \mu_1)$ que llamaremos **espacio de medida inducida en Ω_1 por la función f** .

Teorema 11.6. Teorema A

Sea $(\Omega, \mathcal{F}, \mu)$ un espacio de medida y $(\Omega_1, \mathcal{F}_1, \mu_1)$ el espacio de medida inducida por una función $f : \Omega \rightarrow \Omega_1$ medible de (Ω, \mathcal{F}) en $(\Omega_1, \mathcal{F}_1)$.

Si $g : \Omega_1 \rightarrow \overline{\mathbb{R}}$ es medible de $(\Omega_1, \mathcal{F}_1)$ en $(\overline{\mathbb{R}}, \overline{\mathcal{B}_1})$ entonces

$$\int_{\Omega} g(f(w)) d\mu = \int_{\Omega_1} g(w') d\mu_1$$

siempre que al menos una de estas integrales exista.

Demostración. Para el caso en que g es un indicador $g = I_A$, $A \in \mathcal{F}_1$, los valores de estas dos integrales son

$$\int_{\Omega} I_A(f(w)) d\mu = \int_{\Omega} I_{f^{-1}(A)}(w) d\mu = \mu(f^{-1}(A)) = \mu_1(A) = \int_{\Omega_1} I_A(w') d\mu_1.$$

Si $g = \sum x_i I_{A_i}$ es una función simple no negativa se tiene, aplicando lo anterior

$$\int_{\Omega} x_i I_{A_i}(f(w)) d\mu = \int_{\Omega_1} x_i I_{A_i}(w') d\mu_1 \implies \int_{\Omega} g(f(w)) d\mu = \int_{\Omega_1} g(w') d\mu_1$$

Si g es medible no negativa y (g_n) es una sucesión monótona creciente de funciones simples no negativas que converge a g resulta

$$\int_{\Omega} g_n(f(w)) d\mu = \int_{\Omega_1} g_n(w') d\mu_1, \forall n \implies \int_{\Omega} g(f(w)) d\mu = \int_{\Omega_1} g(w') d\mu_1$$

Finalmente, si g es medible, $g = g^+ - g^-$, aplicando el resultado anterior a g^+ y g^- y restándolos, obtenemos la igualdad buscada. \square

Ejemplos

- Sea (Ω, \mathcal{S}, P) un espacio de probabilidad, $X : \Omega \rightarrow \mathbb{R}^k$ medible de (Ω, \mathcal{S}) en $(\mathbb{R}^k, \mathcal{B}_k)$, o sea, X es una variable aleatoria de k dimensiones, y $(\mathbb{R}^k, \mathcal{B}_k, P_X)$ el espacio de probabilidad inducido por X .

Llamaremos F_X a la función de distribución de la función de probabilidad P_X . Entonces, para cada función medible Borel $g : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$ se tiene

$$E(g(X)) = \int_{\Omega} g(X) dP = \int_{\mathbb{R}^k} g(x_1, \dots, x_k) dF_X(x_1, \dots, x_k)$$

si esta esperanza existe.

Esta igualdad resulta tomando en el teorema A como espacio $(\Omega, \mathcal{F}, \mu)$ el espacio de probabilidad (Ω, \mathcal{S}, P) , como espacio medible $(\Omega_1, \mathcal{F}_1)$ el $(\mathbb{R}^k, \mathcal{B}_k)$ y como función f la función X .

- Sea $(\mathbb{R}^k, \mathcal{B}_k, P)$ un espacio de probabilidad de k dimensiones con función de distribución F_X , $Y : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$ una función medible Borel, es decir, una variable aleatoria de una dimensión k definida en el espacio $(\mathbb{R}^k, \mathcal{B}_k, P)$.

Llamaremos F_Y a la función de distribución de la variable aleatoria Y . Entonces, para cada función medible Borel $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ se tiene

$$E(g(Y)) = \int_{\mathbb{R}^k} g(Y(x_1, \dots, x_k)) dF_X(x_1, \dots, x_k) = \int_{\mathbb{R}} g(y) dF_Y(y)$$

si esta esperanza existe.

11.6. Medidas en un mismo espacio

Teorema 11.7. Teorema B

Sean $(\Omega, \mathcal{F}, \lambda)$ y $(\Omega, \mathcal{F}, \mu)$ dos espacios de medida tales que existe una función $f : \Omega \rightarrow \mathbb{R}$ medible de (Ω, \mathcal{F}) en $(\mathbb{R}, \mathcal{B}_1)$ tales que

$$\mu(A) = \int f I_A d\lambda, \quad \forall A \in \mathcal{F}$$

Entonces, para cualquier función medible $g : \Omega \rightarrow \overline{\mathbb{R}}$ se tiene

$$\int g d\mu = \int g f d\lambda$$

Nota: la función f se llama **derivada de Radon-Nikodym de μ respecto de λ** , lo que se escribe con la notación

$$f = \frac{d\mu}{d\lambda}$$

Esta nomenclatura es adecuada porque estas derivadas poseen propiedades similares a las derivadas de las funciones reales. Recordemos también que si se dan dos medidas μ, λ en el mismo espacio medible (Ω, \mathcal{F}) que cumplen la condición

$$\lambda(A) = 0 \implies \mu(A) = 0$$

entonces se dice que μ es absolutamente continua respecto de λ . El Teorema de Radon-Nikodym **3.4** asegura que en tal caso existe la derivada f .

Demostración. Tomando $g = I_A$ con $A \in \mathcal{F}$ tenemos

$$\int g d\mu = \int I_A d\mu = \mu(A)$$

$$\int g f d\lambda = \int I_A f d\lambda = \mu(A)$$

y la igualdad se verifica.

Escalando por un escalar sigue valiendo la igualdad.

Sumando una cantidad finita, sigue valiendo la igualdad, y tenemos la igualdad para funciones simples.

Tomando el límite obtenemos la igualdad para funciones no negativas.

Tomando la parte positiva y negativa, obtenemos la igualdad para funciones reales. \square

11.6.1. Ejemplo

Sea $(\mathbb{R}^k, \mathcal{B}_k, P)$ una distribución de probabilidad con la función de densidad $f : \mathbb{R}^k \rightarrow \mathbb{R}$, es decir,

$$P(B) = \int f I_B dx, \forall B \in \mathcal{B}_k$$

Entonces, para cada variable aleatoria $X : \mathbb{R}^k \rightarrow \overline{\mathbb{R}}$ que posea esperanza matemática se tiene

$$E(X) = \int X dP = \int_{\mathbb{R}^k} X(x_1, \dots, x_k) f(x_1, \dots, x_k) dx_1 \dots dx_k$$

Esta igualdad se obtiene tomando, en el teorema B, el espacio de medida como un espacio de probabilidad con función de densidad.

12. Momentos (1 dimensión)

12.1. Momentos

Sea X una variable aleatoria y n un entero no negativo. Se llama **momento de orden n** o momento n -ésimo de la variable aleatoria X a $E(X^n)$ si esta esperanza matemática tiene sentido. Este momento se denotará

$$\alpha_n := E(X^n)$$

El momento de orden cero siempre vale 1:

$$\alpha_0 = E(X^0) = E(1) = 1$$

El momento de orden 1 de la variable aleatoria X se llama **media** de X y se puede denotar por

$$\mu := \alpha_1 = E(X)$$

cuando esta esperanza existe. Los momentos de orden n , α_n , también se denominan momentos ordinarios o momentos respecto del origen.

Si n es un entero no negativo, y $c \in \mathbb{R}$, se llama **momento de orden n respecto de c** de la variable X a

$$\alpha_n(c) := E((X - c)^n)$$

suponiendo su existencia.

Son especialmente importantes los **momentos respecto de la media** o **momentos centrales** de orden n , definidos por

$$\mu_n := E((X - \mu)^n), \mu = E(X)$$

Se llama **varianza** de la variable aleatoria X al momento de segundo orden respecto de la media:

$$\text{Var}(X) := \mu_2 = E\left((X - E(X))^2\right)$$

Proposición 12.1. *La varianza de X es igual al momento de orden 2 respecto del origen menos la media al cuadrado:*

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \alpha_2 - \alpha_1^2 = \alpha_2 - \mu^2$$

Demostración.

$$\text{Var}(X) = E\left((X - \mu)^2\right) = E(X^2 - 2X\mu + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$$

□

La raíz cuadrada aritmética (no negativa) de la varianza de X se denomina **desviación típica** de X y se suele representar por

$$\sigma = \sqrt{\text{Var}(X)}$$

Teorema 12.1. *El momento de segundo orden respecto de c , si existe con valor finito, es mínimo para $c = \mu$*

Demostración. Partimos de la siguiente igualdad

$$(X - c)^2 = ((X - \mu) + (\mu - c))^2 = (X - \mu)^2 + 2(X - \mu)(\mu - c) + (\mu - c)^2$$

y tomamos esperanzas a ambos extremos:

$$\begin{aligned} \alpha_2(c) &:= E\left((X - c)^2\right) = E\left((X - \mu)^2\right) + 2(\mu - c)E(X - \mu) + E\left((\mu - c)^2\right) = \\ &= E\left((X - \mu)^2\right) + 2(\mu - c)(\overbrace{E(X) - \mu}^0) + E\left((\mu - c)^2\right) = \\ &= E\left((X - \mu)^2\right) + (\mu - c)^2 \geq E\left((X - \mu)^2\right). \end{aligned}$$

□

Si $r \in \mathbb{R}$, se llama **momento absoluto de orden r** de la variable aleatoria X a la esperanza

$$\beta_r = E(|X|^r)$$

De menos uso son los momentos factoriales. Para n entero positivo, $x \in \mathbb{R}$, denotaremos $x^{(n)}$ el producto de n factores

$$x^{(n)} = x(x-1)(x-2)\cdots(x-(n-1)) = \prod_{i=0}^{n-1} (x-i)$$

y, por convenio,

$$x^{(0)} = 1$$

Llamamos **momento factorial de orden n** de la variable X a

$$\alpha_{(n)} = E\left(X^{(n)}\right) = E(X(X-1)\cdots(X-(n-1)))$$

si la esperanza anterior tiene sentido. Para $n = 0$, se tiene

$$\alpha_{(0)} = E\left(X^{(0)}\right) = 1$$

12.2. Existencia de los momentos. Relación entre los momentos ordinarios y los momentos centrales

Las variables aleatorias acotadas poseen momentos finitos de todos los órdenes. Si la variable aleatoria X vale

$$|X| \leq K \implies |X|^n \leq K^n$$

entonces

$$|E(X^n)| \leq E(|X^n|) = E(|X|^n) \leq E(K^n) = K^n$$

o, escrito con otra notación

$$|\alpha_n| \leq \beta_n \leq K^n$$

Teorema 12.2. *Sea X una variable aleatoria y $f_1(x), f_2(x)$ dos polinomios de grados r y n respectivamente, con $r \leq n$. Si $E(f_2(X))$ tiene valor finito también $E(f_1(X))$ existe con valor finito.*

En particular, si existe con valor finito alguno de los momentos $\alpha_n, \alpha_n(c), \mu_n, \alpha_{(n)}$ también existen con valor finito todos los momentos $\alpha_r, \alpha_r(c), \mu_r, \alpha_{(r)}$, de órdenes $r \leq n$.

Demostración. Como f_1 es de grado r y f_2 es de grado $n \geq r$ se tiene

$$\lim_{x \rightarrow \pm\infty} \left| \frac{f_1(x)}{f_2(x)} \right| = L < \infty,$$

luego,

$$\forall H > L, \exists R > 0 : \forall |x| > R, |f_1(x)| \leq |f_2(x)| H.$$

Como $|f_1(x)|$ es una función continua en $[-R, R]$ existe el máximo finito

$$M = \max \{|f_1(x)| : |x| \leq R\}$$

Y se tiene para todo $x \in \mathbb{R}$

$$|f_1(x)| \leq M + |f_2(x)| H,$$

y sustituyendo x por la variable aleatoria X queda

$$|f_1(X)| \leq M + |f_2(X)| H$$

y, como el segundo miembro de la igualdad tiene esperanza finita, también el primer miembro tiene esperanza finita, y queda probada la primera parte del teorema.

Como los momentos de orden n de X son esperanzas matemáticas de polinomios en X de grado n , tenemos la segunda parte de la demostración, como consecuencia directa de la primera. \square

12.3. Desigualdad de Tchebychev

Teorema 12.3. Desigualdad de Tchebychev general

Sea X una variable aleatoria no negativa con esperanza matemática finita. Para cada real $a > 0$ se tiene

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Demostración. Se tiene

$$E(X) = E(XI_{\{X \geq a\}}) + E(XI_{\{X < a\}}) \stackrel{X \geq 0}{\geq} E(XI_{\{X \geq a\}}) \geq E(aI_{\{X \geq a\}}) = aP(X \geq a).$$

Por lo tanto, como $a > 0$:

$$P(X \geq a) \leq \frac{E(X)}{a}$$

\square

Esta desigualdad tiene diversas utilidades:

1. Si sustituimos X por $(X - E(X))^2$ y a por ε^2 ($\varepsilon > 0$), resulta

$$P(|X - E(X)| \geq \varepsilon) = P\left((X - E(X))^2 \geq \varepsilon^2\right) \leq \frac{E\left((X - E(X))^2\right)}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2},$$

que suele llamarse **desigualdad de Tchebychev**.

2. Para cualquier variable aleatoria X con momento absoluto de orden $k > 0$ finito y $a > 0$ finito se tiene

$$P(|X| \geq a) = P(|X|^k \geq a^k) \leq \frac{E(|X|^k)}{a^k}$$

que suele llamarse **desigualdad de Markov**.

3. Sea X una variable aleatoria que toma valores en el intervalo $E \subset \mathbb{R}$, $f : E \rightarrow \mathbb{R}$ una función monótona creciente y no negativa y la variable aleatoria $Y = f(X) \geq 0$ con esperanza finita (se le puede aplicar la desigualdad de Tchebychev general). Así, para $a \in E$ y $f(a) > 0$ se tiene

$$P(X \geq a) \stackrel{f \text{ monótona}}{\leq} P(f(X) \geq f(a)) \leq \frac{E(f(X))}{f(a)}.$$

12.4. Momentos de las distribuciones simétricas

La variable aleatoria X tiene una **distribución simétrica respecto del punto c** si las variables aleatorias $X - c$ y $c - X$ tienen la misma distribución. Para ello debe verificarse que las dos funciones de distribución de $X - c$ y de $c - X$ sean iguales. Esto equivale a que se cumpla, para todo $x \in \mathbb{R}$:

$$P(X - c \leq x) = P(c - X \leq x) = P(X \geq c - x) = 1 - P(X < c - x)$$

que, expresado con la función de distribución F de X queda

$$F(c + x) = 1 - F((c - x) -)$$

o, también,

$$F(t) = 1 - F((2c - t) -)$$

Si la función de distribución F tiene derivada, esta será la función de densidad $f(x)$ de la variable aleatoria X . Derivando la relación anterior resulta

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

luego la gráfica de la función de densidad resulta simétrica respecto de la recta $y = c$.

Si X es simétrica respecto del origen ($c = 0$) la igualdad

$$F(c + x) = 1 - F((c - x) -)$$

se convierte en

$$F(x) = 1 - F((-x) -)$$

y para la función de densidad queda

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

Teorema 12.4. *Si X tiene una distribución simétrica respecto de c y posee media finita, entonces $E(X) = c$.*

Demostración. Como las distribuciones de $X - c$ y de $c - X$ son iguales, se tiene

$$E(X - c) = E(c - X) \implies E(X) - c = c - E(X) \implies E(X) = c$$

□

Teorema 12.5. Si X es simétrica respecto de c , cualquier momento respecto de la media de orden impar que exista con valor finito es nulo.

Demostración. Sea n impar. Como las distribuciones de $X - c$ y de $c - X$ son iguales, entonces también lo son las de $(X - c)^n$ y la de $(c - X)^n$ y se tiene

$$\begin{aligned} E((X - c)^n) &= E((c - X)^n) = E((-1)^n (X - c)^n) = \\ &= (-1)^n E((X - c)^n) \stackrel{n \text{ impar}}{=} -E((X - c)^n) \stackrel{E((X - c)^n) \in \mathbb{R}}{\implies} E((X - c)^n) = 0 \end{aligned}$$

□

12.5. Otras características de las distribuciones

Una **característica de una distribución** es un número obtenido a partir de la distribución mediante un criterio determinado. Por ello, una característica de una distribución contiene alguna información acerca de la distribución de la que procede. Los momentos considerados anteriormente, y las funciones de ellos, son características obtenidas sirviéndose del concepto de esperanza matemática.

Veremos ahora algunas características de las distribuciones de una dimensión obtenidas incluyendo otros procedimientos.

Llamaremos **moda** de la distribución de una variable aleatoria X de tipo discreto a un valor m_d para el que se alcance el valor máximo de la función puntual de probabilidad p de la variable X :

$$p(m_d) = \text{máx} \{p(x) : p(x) = P(X = x), x \in \mathbb{R}\},$$

es decir, m_d es una moda si

$$P(X = m_d) \geq P(X = x), \forall x \in \mathbb{R}$$

Si X es una variable de tipo continuo con función de densidad f , se llama moda de la distribución de X al número m_d tal que

$$f(m_d) = \text{máx} \{f(x) : x \in \mathbb{R}\}$$

Llamaremos una **mediana** de la variable aleatoria X con función de distribución F a un número m_e que verifique

$$F(m_e-) \leq \frac{1}{2} \leq F(m_e)$$

Teorema 12.6. Si X tiene una distribución simétrica respecto de c este centro de simetría es una mediana.

Demostración. Se tiene

$$F(x + c) = 1 - F((c - x) -)$$

Entonces

$$F(c) = 1 - F(c-) \implies F(c) + F(c-) = 1$$

y se tiene

$$2F(c) = F(c) + 1 - F(c-) \geq 1 \implies F(c) \geq \frac{1}{2}$$

$$2F(c-) = F(c-) + 1 - F(c) \leq 1 \implies F(c-) \leq \frac{1}{2}$$

□

Dado $0 < \alpha < 1$, llamaremos **cuantil α** a un número q_α tal que

$$F(q_{\alpha-}) \leq \alpha \leq F(q_\alpha)$$

Así pues, la mediana m_e es el cuantil $\frac{1}{2}$:

$$m_e = q_{\frac{1}{2}}$$

La condición anterior siempre tiene solución en q_α , pero esta solución puede no ser única. En efecto, $F(x)$ puede ser constante e igual a α en un intervalo.

También se utilizan las denominaciones de **centil p** o de **pertcentil p** para referirse al cuantil $\alpha = \frac{p}{100}$, además de **decil p** para $\alpha = \frac{p}{10}$ y **cuartil p** para el cuantil $\alpha = \frac{p}{4}$.

Las características de una distribución pueden tener distintas finalidades o propiedades, lo que permite separar estas características en varios grupos:

- **Medidas de centralización de la variable X :** tratan de obtener un punto que, en algún sentido, sea el centro de la masa de probabilidad que X determina en la recta. Entre estas medidas están:

- la media
- la mediana
- la moda
- la media armónica

$$m_a = \frac{1}{E(X^{-1})}$$

- la media geométrica

$$m_g = \exp(E(\log X)), \text{ si } X > 0$$

- **Medidas de dispersión de la variable X :** tienen por objeto indicar si la masa de probabilidad que determina X en la recta está más concentrada o más dispersa alrededor de algún punto central. Algunas son:

- La desviación típica

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{E((X - E(X))^2)}$$

- La desviación media

$$E(|X - m_e|)$$

- El recorrido intercuartílico

$$q_{\frac{3}{4}} - q_{\frac{1}{4}}$$

- **Medidas de asimetría:** pretenden distinguir si la parte de distribución que queda a la derecha del punto central es de mayor peso (asimetría positiva) o de menor peso (asimetría negativa) que la parte que queda a la izquierda. Se suele utilizar:

- coeficiente de asimetría

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

- **Medidas de curtosis o apuntamiento:** tienen por objeto medir el mayor o menor apuntamiento o altura de la distribución en la parte central suponiendo normalizada la distribución, es decir, reducida la dispersión a una unidad convencional. Para este fin se puede utilizar:

- coeficiente de apuntamiento

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

que, en términos generales, será mayor cuanto más apuntada sea la distribución en el centro, o sea, cuanto mayores sean las probabilidades en las proximidades de la media.

Parte II

Ampliación de Probabilidad y Procesos Estocásticos

13. Momentos

13.1. Momentos de distribuciones de varias dimensiones

Dada la variable de k dimensiones

$$X = (X_1, X_2, \dots, X_k) \quad (2)$$

y un conjunto ordenado de k enteros no negativos

$$r = (r_1, r_2, \dots, r_k) \quad (3)$$

Llamaremos **momento de orden** r de la variable X a la expresión

$$\alpha_{(r_1, r_2, \dots, r_k)} = E(X_1^{r_1} X_2^{r_2} \dots X_k^{r_k})$$

Los momentos de las variables aleatorias k -dimensionales X respecto de la media. factoriales y absolutos se definen de modo similar al llevado a cabo en una dimensión.

Así, el **momento central** o **respecto de la media** de la variable aleatoria X , de orden $r = (r_1, r_2, \dots, r_k)$, viene dado por

$$\mu_{r_1, r_2, \dots, r_k} = E((X_1 - E(X_1))^{r_1} (X_2 - E(X_2))^{r_2} \dots (X_k - E(X_k))^{r_k})$$

Si en (3) algunos de los enteros r_i son no nulos, resulta que el **momento de orden** r de X coincide con el momento de orden r' de la variable X' , donde r' y X' resulta de suprimir en r y en X , respectivamente, las componentes r_i y X_i cuyos índices i tienen $r_i = 0$. Por ejemplo, los momentos de órdenes $(3, 0, 0)$, $(0, 2, 0)$, $(1, 0, 4)$ de la variable aleatoria (X, Y, Z) son respectivamente $E(X^3)$, $E(Y^2)$, $E(XZ^4)$.

En el caso de independencia de las variables X_i los momentos de $X = (X_1, X_2, \dots, X_k)$ son producto de los momentos de estas variables

$$\alpha_{(r_1, r_2, \dots, r_k)} = E(X_1^{r_1} X_2^{r_2} \dots X_k^{r_k}) = E(X_1^{r_1}) \dots E(X_k^{r_k})$$

Hecho 13.1. Sean a_1, a_2, \dots, a_k números no negativos y sean r_1, r_2, \dots, r_k enteros no negativos que suman $m > 0$, vale la desigualdad

$$\frac{m!}{r_1! r_2! \dots r_k!} a_1^{r_1} a_2^{r_2} \dots a_k^{r_k} \leq k^m (a_1^m + a_2^m + \dots + a_k^m)$$

Demostración. Utilizando la fórmula de Leibniz

$$(a_1 + a_2 + \dots + a_k)^m = \sum_{r_i \geq 0: r_1 + r_2 + \dots + r_k = m} \frac{m!}{r_1! r_2! \dots r_k!} a_1^{r_1} a_2^{r_2} \dots a_k^{r_k},$$

tenemos que, por ser cada sumando no negativo, cada uno de los mismos es menor o igual que la suma, y así podemos deducir directamente la desigualdad; en efecto, tomando $a_h := \max\{a_i\}$, se tiene

$$\frac{m!}{r_1!r_2!\cdots r_k!}a_1^{r_1}a_2^{r_2}\cdots a_k^{r_k} \leq (a_1 + a_2 + \cdots + a_k)^m \leq k^m a_h^m \leq k^m (a_1^m + a_2^m + \cdots + a_h^m + \cdots + a_k^m)$$

□

Hecho 13.2 (Importante). *Si para las variables aleatorias X_1, X_2, \dots, X_k existen con valor finito los k momentos de orden m , $E(X_i^m)$ y vale $r_1 + r_2 + \cdots + r_k \leq m$, entonces existen con valor finito los momentos $\alpha_{r_1, r_2, \dots, r_k}$.*

Demostración. Por el resultado anterior tenemos

$$|X_1|^{r_1} |X_2|^{r_2} \cdots |X_k|^{r_k} \leq \frac{r_1!r_2!\cdots r_k!}{m!} k^m (|X_1|^m + |X_2|^m + \cdots + |X_k|^m).$$

El segundo miembro de la desigualdad es una constante por la suma de v.v.a.a. que tienen esperanza finita⁸, por lo que todo el miembro tiene esperanza finita. Así, existe con valor finito el momento absoluto $\beta_{r_1, r_2, \dots, r_k} = E(|X_1|^{r_1} |X_2|^{r_2} \cdots |X_k|^{r_k})$ y por tanto el momento ordinario $\alpha_{r_1, r_2, \dots, r_k}$. El mismo razonamiento se aplica si $r_1 + r_2 + \cdots + r_k = n < m$. □

13.2. Covarianza y coeficientes de correlación. Desigualdad de Schwarz

Se llama **covarianza** de las variables aleatorias X, Y (de una dimensión) a

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y)))$$

La covarianza de X, Y es el momento central $\mu_{1,1}$ de la variable aleatoria bidimensional (X, Y) . Un cálculo sencillo permite obtener

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY - XE(Y) - YE(X) + E(X)E(Y)) = \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) = E(XY) - E(X)E(Y). \end{aligned} \quad (4)$$

Además, se obtienen trivialmente las propiedades:

- $\text{Var}(X) = \text{Cov}(X, X)$.
- Simetría: $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- Bilinealidad:
 - $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$.
 - $\text{Cov}(cX, Y) = c\text{Cov}(X, Y)$, $c \in \mathbb{R}$.

⁸ $E(X) = E(X_+) - E(X_-)$ finita $\iff E(X_+), E(X_-)$ ambas finitas $\iff E(|X|) = E(X_+) + E(X_-)$ finita y $|X|^m = |X^m|$.

Las variables X, Y se llaman **incorreladas** o **incorrelacionadas** si $Cov(X, Y) = 0$. Si X, Y son independientes entonces son incorreladas pues $Cov(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$.

Dos variables X, Y se llaman **ortogonales** si $E(XY) = 0$. Por ejemplo, $X - E(X)$ es ortogonal a cualquier constante c .

Si X e Y son incorreladas, entonces $X - E(X)$ e $Y - E(Y)$ son ortogonales.

Se llama **coeficiente de correlación** de X, Y a

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}},$$

donde se supone $Var(X) > 0, Var(Y) > 0$.

Teorema 13.1 (Desigualdad de Schwarz). *Si X, Y tienen momentos de segundo orden finitos vale la siguiente desigualdad*

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)} \quad (5)$$

Demostración. Distinguiamos dos casos:

- $E(X^2) = 0 \iff X = 0$ c.s. $\implies XY = 0$ c.s. $\implies E(XY) = 0$ y ambos miembros de la desigualdad son nulos.
- Supongamos ahora $E(X^2) > 0$. Denotando $A = E(X^2), B = E(XY), C = E(Y^2)$, se tiene, para todo $u \in \mathbb{R}$,

$$0 \leq E\left((uX + Y)^2\right) = u^2E(X^2) + 2uE(XY) + E(Y^2) = u^2A + 2uB + C = \frac{1}{A} [u^2A^2 + 2uAB + AC] = \frac{1}{A} [(uA + B)^2 + AC - B^2], \quad (6)$$

y en particular para $u = -B/A$ resulta

$$E\left(\left(Y - \frac{B}{A}X\right)^2\right) = \frac{1}{A} (AC - B^2) \geq 0 \implies AC - B^2 \geq 0 \implies |E(XY)| = |B| \leq \sqrt{AC} = \sqrt{E(X^2)E(Y^2)}$$

□

La única posibilidad de que valga la igualdad, $|B| = \sqrt{AC}$, es que se verifique $E\left(\left(Y - (B/A)X\right)^2\right) = 0$, o sea, $Y = (B/A)X = (E(XY)/E(X^2))X$ c.s.

Si en la desigualdad de Schwarz (5) se cambian X, Y por $X - E(X), Y - E(Y)$ se obtiene

$$|Cov(X, Y)| \leq \sqrt{Var(X)Var(Y)},$$

de donde $|\rho| \leq 1$. Entonces, la única posibilidad de que valga la igualdad en la anterior relación es que valga

$$Y - E(Y) = \frac{Cov(X, Y)}{Var(X)} (X - E(X)) \text{ c.s.}$$

Compárese esta igualdad con la ecuación de la recta de regresión de la sección siguiente.

13.2 Rectas de regresión

Supondremos dadas las variables aleatorias X, Y con momentos de segundo orden finitos. Utilizaremos para las medias, varianzas y covarianza de estas variables la notación siguiente

- $\mu_X = E(X), \mu_Y = E(Y)$.
- $\sigma_X^2 = Var(X) = E((X - \mu_X)^2) = \alpha_{20} - \mu_X^2$.
- $\sigma_Y^2 = Var(Y) = E((Y - \mu_Y)^2) = \alpha_{02} - \mu_Y^2$.
- $\sigma_{XY} = Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = \alpha_{11} - \mu_X\mu_Y$.

Sea $Z = (X, Y)$ una variable aleatoria bidimensional, llamaremos **coeficiente de regresión de Y sobre X** al coeficiente $\beta_{Y/X} = \frac{\sigma_{XY}}{\sigma_X^2}$ con $\sigma_X > 0$, y el de X sobre Y al coeficiente $\beta_{X/Y} = \frac{\sigma_{XY}}{\sigma_Y^2}$.

Teorema 13.2. *Si la v.a. bidimensional (X, Y) tiene una distribución con momentos de segundo orden finitos y la varianza de X es positiva^a, entonces*

$$\min_{a, b \in \mathbb{R}} E((Y - a - bX)^2) = E((Y - \alpha - \beta X)^2)$$

Con α y β dados por $\alpha := \mu_Y - \beta_{Y/X}\mu_X$ y $\beta := \beta_{Y/X}$.

Además se verifica que $\alpha + \beta X = \mu_Y + \beta_{Y/X}(X - \mu_X)$ con α y β únicos.

^a X no es casi seguramente constante.

Demostración. Sea Z una v.a. con momento de segundo orden finito y $a \in \mathbb{R}$, entonces podemos escribir

$$\begin{aligned} E((Z - a)^2) &= E((Z - E(Z))^2) + E((E(Z) - a)^2) + 2E((Z - E(Z))(E(Z) - a)) = \\ &= E((Z - E(Z))^2) + (E(Z) - a)^2 + 2(E(Z) - a)E(Z - E(Z)) = E((Z - E(Z))^2) + (E(Z) - a)^2, \end{aligned}$$

pues $E(Z) - a$ es una constante y $E(Z - E(Z)) = 0$. Por otra parte, si escribimos ahora $Z = Y - bX$, que podemos pues los momentos de orden 2 de (X, Y) son finitos y $E(XY)$ también por la desigualdad de Schwarz, tenemos que $E(Z) = \mu_Y - b\mu_X$ y sustituyendo en la anterior fórmula,

$$\begin{aligned} E((Y - bX - a)^2) &= E(((Y - \mu_Y) - b(X - \mu_X))^2) + (\mu_Y - b\mu_X - a)^2 = \\ &= \sigma_Y^2 - 2b\sigma_{XY} + b^2\sigma_X^2 + (\mu_Y - b\mu_X - a)^2 = \\ &= b^2\sigma_X^2 + \frac{\sigma_{XY}^2}{\sigma_X^2} - 2b\sigma_{XY} + \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} + (\mu_Y - b\mu_X - a)^2 = \\ &= \left(b\sigma_X - \frac{\sigma_{XY}}{\sigma_X}\right)^2 + \left(1 - \frac{\sigma_{XY}^2}{\sigma_X^2\sigma_Y^2}\right)\sigma_Y^2 + (\mu_Y - b\mu_X - a)^2 = \\ &= \left(b\sigma_X - \frac{\sigma_{XY}}{\sigma_X}\right)^2 + (1 - \rho^2)\sigma_Y^2 + (\mu_Y - b\mu_X - a)^2, \end{aligned}$$

donde ρ es el coeficiente de correlación. Ahora bien, el segundo sumando no depende ni de a ni de b , por lo que debemos tratar de reducir los otros dos sumandos al menor valor posible, y como son no negativos, tenemos el sistema lineal

$$\begin{cases} b\sigma_X - \frac{\sigma_{XY}}{\sigma_X} = 0 \\ \mu_Y - b\mu_X - a = 0, \end{cases}$$

con solución única $a = \mu_Y - \frac{\sigma_{XY}}{\sigma_X^2}\mu_X = \alpha$, $b = \frac{\sigma_{XY}}{\sigma_X^2} = \beta$, como queríamos ver. \square

Llamamos **recta de regresión de Y sobre X** a la recta que pasa por (μ_X, μ_Y) y que tiene pendiente $\beta_{Y/X}$. Esta es, según su representación,

$$Y = \mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(X - \mu_X) = \mu_Y + \beta_{Y/X}(X - \mu_X) = \alpha + \beta X$$

$$\frac{Y - \mu_Y}{\sigma_Y} = \rho \frac{X - \mu_X}{\sigma_X}$$

Llamamos a la variable aleatoria $X^* := \mu_Y + \beta_{Y/X}(X - \mu_X)$ la mejor representación lineal de Y respecto de X . Usando la expresión de X^* definimos la **variable residual de Y respecto de X** como $Y - X^*$, que según la demostración anterior tiene

$$\text{Var}(Y - X^*) = E\left((Y - \alpha - \beta X)^2\right) = (1 - \rho^2)\sigma_Y^2. \quad (7)$$

Además,

$$E(Y - X^*) = E(Y) - \mu_Y - \frac{\sigma_{XY}}{\sigma_X^2}(E(X) - \mu_X) = 0 + 0 = 0.$$

Por otra parte, permutando los papeles de X e Y obtenemos la recta de regresión de X respecto de Y , así como la residual de X respecto de Y .

13.3. Curvas de regresión

Llamaremos **esperanza de Y condicionada por $X = x$** a $m_{2|1}(x) := E(Y|X = x) = \int_{\mathbb{R}} y dF_{2|1}(y|x)$ (de aquí en adelante se omitirá el dominio de integración). Supondremos que esta integral y las siguientes tienen valor finito. Llamamos **curva de regresión** de \mathbb{R}^2 a $y = m_{2|1}(x)$ (de Y sobre X).

Teorema 13.3. Dada la distribución de (X, Y) , supongamos $y = m_{2|1}(x)$ curva de regresión de Y sobre X . Entonces $\min_g E\left((Y - g(X))^2\right) = E\left((Y - m_{2|1}(X))^2\right)$ con $g : \mathbb{R} \rightarrow \mathbb{R}$ medible Borel

Observación. La demostración que van a ver a continuación ha sufrido severas modificaciones que la han podido dejar inservible.

Demostración. Como g es medible Borel, vale

$$\begin{aligned} & \int (y - g(x))^2 dF_{2|1}(y|x) = \\ & = \int (y - m_{2|1}(x))^2 dF_{2|1}(y|x) + \int (m_{2|1}(x) - g(x))^2 dF_{2|1}(y|x) + 2 \int (y - m_{2|1}(x))(m_{2|1}(x) - g(x)) dF_{2|1}(y|x) \end{aligned}$$

Pero la tercera integral vale

$$\begin{aligned} \int (y - m_{2|1}(x))(m_{2|1}(x) - g(x)) dF_{2|1}(y|x) &= (m_{2|1}(x) - g(x)) \int (y - m_{2|1}(x)) dF_{2|1}(y|x) = \\ &= (m_{2|1}(x) - g(x)) \left(\int y dF_{2|1}(y|x) - m_{2|1}(x) \right) = (m_{2|1}(x) - g(x))(m_{2|1}(x) - m_{2|1}(x)) = 0, \end{aligned}$$

donde se ha utilizado $\int dF_{2|1}(y|x) = P(Y \in \mathbb{R}|X = x) = 1$ y la definición de $m_{2|1}(x)$.

Tras esto tenemos

$$\int (y - g(x))^2 dF_{2|1}(y|x) = \int y - m_{2|1}(x) dF_{2|1}(y|x) + \int (m_{2|1}(x) - g(x))^2 dF_{2|1}(y|x),$$

e integrando respecto de $dF_1(x)$ obtenemos

$$\begin{aligned} \int_{\mathbb{R}} \left(\int (y - g(x))^2 dF_{2|1}(y|x) \right) dF_1(x) &= \int_{\mathbb{R}} \left(\int y - m_{2|1}(x) dF_{2|1}(y|x) \right) dF_1(x) + \int_{\mathbb{R}} \left(\int (m_{2|1}(x) - g(x))^2 dF_{2|1}(y|x) \right) dF_1(x) \\ \int_{\mathbb{R}^2} (y - g(x))^2 dF(x, y) &= \int_{\mathbb{R}^2} y - m_{2|1}(x) dF(x, y) + \int_{\mathbb{R}^2} (m_{2|1}(x) - g(x))^2 dF(x, y) \\ E \left((Y - g(X))^2 \right) &= E \left((Y - m_{2|1}(X))^2 \right) + E \left((m_{2|1}(X) - g(X))^2 \right) \end{aligned}$$

Los sumandos de la derecha son no negativos y el primero no depende de g , luego $E \left((Y - g(X))^2 \right)$ es mínimo cuando

$$E \left((m_{2|1}(X) - g(X))^2 \right) = 0 \iff g(x) = m_{2|1}(x) \text{ c.s.},$$

valiendo entonces el mínimo

$$E \left((Y - g(X))^2 \right) = E \left((Y - m_{2|1}(X))^2 \right).$$

□

Llamamos **razón de correlación de Y sobre X** al número $\sigma_{Y/X}$ que satisface

$$E \left((Y - m_{2|1}(x))^2 \right) = \sigma_Y^2 (1 - \sigma_{Y/X}^2).$$

Veamos ahora la relación entre ρ y $\sigma_{Y/X}$. Como hemos visto en la demostración anterior

$$E \left((Y - g(X))^2 \right) = E \left((Y - m_{2|1}(X))^2 \right) + E \left((m_{2|1}(X) - g(X))^2 \right) = \sigma_Y^2 (1 - \sigma_{Y/X}^2) + E \left((m_{2|1}(X) - g(X))^2 \right);$$

ahora, haciendo $g(X) = \alpha + \beta X$ y usando (7), tenemos

$$\sigma_Y^2 (1 - \rho^2) = \sigma_Y^2 (1 - \sigma_{Y/X}^2) + E \left((m_{2|1}(X) - g(X))^2 \right)$$

$$\sigma_{Y/X}^2 = \rho^2 + \frac{E \left((m_{2|1}(X) - g(X))^2 \right)}{\sigma_Y^2}$$

Podemos trabajar análogamente para X sobre Y .

13.4. Momentos de sumas

Supongamos X_1, \dots, X_n n variables aleatorias cualesquiera. Sea $S_n = X_1 + \dots + X_n$, sabemos que $E(S_n) = \sum_{i=1}^n E(X_i)$, y sean $Y_i = X_i - E(X_i)$ con $i = 1, \dots, n$, tenemos:

$$\begin{aligned} Var(S_n) &= E\left((S_n - E(S_n))^2\right) = E\left(\left(\sum X_i - E(\sum X_i)\right)^2\right) = E\left(\left(\sum X_i - E(X_i)\right)^2\right) = \\ &= E\left(\left(\sum Y_i\right)^2\right) = E\left(\sum Y_i^2 + 2\sum_{i<j} Y_i Y_j\right) = \sum Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j). \end{aligned}$$

Luego si las variables son incorreladas tenemos que $Var(\sum X_i) = \sum Var(X_i)$.

Veamos ahora el momento central de tercer orden de las variables X_i cuando estas son independientes.

$$\begin{aligned} E\left((S_n - E(S_n))^3\right) &= E\left(\left(\sum X_i - E(\sum X_i)\right)^3\right) = E\left(\left(\sum Y_i\right)^3\right) = \\ &\stackrel{\text{Leibniz}}{=} E\left(\sum Y_i^3 + \frac{3!}{2!1!} \sum_{i \neq j} Y_i^2 Y_j + \frac{3!}{1!1!1!} \sum_{i<j<k} Y_i Y_j Y_k\right) = \\ &= E\left(\sum Y_i^3\right) + 3E\left(\sum_{i \neq j} Y_i^2 Y_j\right) + 6E\left(\sum_{i<j<k} Y_i Y_j Y_k\right) = \\ &\stackrel{\text{indep.}}{=} \sum E(Y_i^3) + 3\sum_{i \neq j} E(Y_i^2) E(Y_j) + 6\sum_{i<j<k} E(Y_i) E(Y_j) E(Y_k) = \\ &\stackrel{E(Y_i)=0}{=} \sum E(Y_i^3) = \sum E\left((X_i - E(X_i))^3\right). \end{aligned}$$

Calculemos ahora el momento central de cuarto orden de las variables X_i independientes. De nuevo por la regla de Leibniz, llegamos a

$$\begin{aligned} E\left((S_n - E(S_n))^4\right) &= \\ &= E\left(\sum Y_i^4 + \frac{4!}{3!1!} \sum_{i \neq j} Y_i^3 Y_j + \frac{4!}{2!2!} \sum_{i<j} Y_i^2 Y_j^2 + \frac{4!}{2!1!1!} \sum_{i \neq j, k; j<k} Y_i^2 Y_j Y_k + \frac{4!}{1!1!1!1!} \sum_{i<j<k<l} Y_i Y_j Y_k Y_l\right) = \\ &= \sum E(Y_i^4) + 6\sum_{i<j} E(Y_i^2 Y_j^2) = \sum E\left((X_i - E(X_i))^4\right) + 6\sum_{i<j} E\left((X_i - E(X_i))^2 (X_j - E(X_j))^2\right) \end{aligned}$$

14. Funciones generatrices

14.1. Funciones generatrices. Introducción

Llamamos **función generatriz** de una sucesión de números a cualquier función que determina, mediante coeficientes de algún desarrollo en serie de la función, la sucesión de números dada. Algunos ejemplos de esto son:

Función generatriz de probabilidad: $f_X := E(t^X)$

Función generatriz de momentos: $g_X := E(e^{tX})$

Función característica: $\varphi_X := E(e^{itX})$

Función generatriz de momentos respecto de la media: $g_{X-\mu} := E(e^{t(X-\mu)})$

Función generatriz de momentos factoriales: $h_X := E\left((1+t)^X\right)$

14.2. Función generatriz de probabilidad

Sea X v.a. entera no negativa con probabilidades $p_n := P(X_n = n)$ con $n \in \mathbb{N} \cup \{0\}$ y $\sum p_n = 1$, definimos la **función generatriz de probabilidad** $f_X : C \rightarrow \mathbb{R}$ de la v.a. X a la función dada por:

$$f_X(t) := E(t^X) = \sum p_n t^n,$$

siendo $C \subset \mathbb{R}$ el conjunto de números reales donde $E(t^X) < \infty$.

Si el número de sumandos es finito, entonces $C = \mathbb{R}$ ya que $\forall t \in \mathbb{R}, E(t^X) < \infty$. Si son infinitos, tenemos una serie de potencias y C sería el conjunto donde converge la serie. Siempre se tiene $[-1, 1] \subset C$, pues para $t \in [-1, 1]$ $0 < |p_n t^n| \leq p_n \implies [\sum p_n = 1 \implies \sum p_n t^n \text{ converge}]$.

Por las propiedades de las series de potencias, si t es interior a C , derivando en $f_X(t) = E(t^X) = \sum p_n t^n$ tenemos que existen las series derivadas y son convergentes, e igual si integramos.

Proposición 14.1. *Si el radio de convergencia de f_X es mayor que 1, existen los momentos factoriales de X de cualquier orden con valor finito.*

Demostración. Para un t interior a C , de la expresión de $f_X(t)$ obtenemos derivando

$$f^{(r)}(t) = E(X(X-1)\cdots(X-r+1)t^{X-r}) = \sum p_n n(n-1)\cdots(n-r+1)t^{n-r}.$$

Además, si $t = 1$ es interior a C , los momentos factoriales son finitos pues

$$\alpha_{(r)} = E(X^{(r)}) = E(X(X-1)\cdots(X-r+1)) = \sum p_n n(n-1)\cdots(n-r+1) = f^{(r)}(1).$$

□

Si fuese $C = [-1, 1]$, considerando $t \in (0, 1)$ se tiene $E(X^{(r)}t^{X-r})$. Como la variable $X^{(r)}t^{X-r}$ es no negativa y es monótona creciente en $t \in (0, 1)$, por el teorema de convergencia monótona se tiene que

$$\lim_{t \rightarrow 1^-} f^{(r)}(t) = \lim_{t \rightarrow 1^-} E(X^{(r)}t^{X-r}) = E\left(\lim_{t \rightarrow 1^-} X^{(r)}t^{X-r}\right) = E(X^{(r)}) = \alpha_{(r)},$$

pero en este caso desconocemos si el límite tiene un valor finito.

Teorema 14.1. Sean X_1, X_2, \dots, X_n v.a. independientes enteras no negativos, con funciones generatrices de probabilidad $f_r(t) = E(t^{X_r})$. Entonces la función generatriz de probabilidad f de la suma $S_n = X_1 + \dots + X_n$ es el producto de las funciones generatrices f_r en el conjunto en el que todas las f_r están bien definidas (al menos en $[-1, 1]$); es decir,

$$f(t) = f_1(t)f_2(t) \cdots f_n(t)$$

Demostración. $f(t) = E(t^{(X_1+\dots+X_n)}) = E(t^{X_1} \cdots t^{X_n})$ y como $t^{X_1}, t^{X_2}, \dots, t^{X_n}$ son independientes por serlo X_1, X_2, \dots, X_n , se da

$$f(t) = E(t^{X_1})E(t^{X_2}) \cdots E(t^{X_n}) = f_1(t)f_2(t) \cdots f_n(t).$$

□

14.3. Función generatriz de momentos

Dada la v.a. X llamamos **función generatriz de momentos de X** a la función real $g_X(t) = E(e^{tX}) = \int_{\mathbb{R}} e^{tx} dF(x)$ (escribiremos simplemente $g(t)$ cuando esté claro) definida donde esta integral tenga valor finito. Se llama **función generatriz de momentos respecto c de la v.a. X** a la función:

$$g_{X-c}(t) = E(e^{t(X-c)}) = e^{-tc}E(e^{tX}) = e^{-tc}g_X(t).$$

Dadas estas dos definiciones, es claro que una existe con valor finito si y solo si existe la otra con valor finito. Es claro que estas funciones valen 1 para $t = 0$.

Teorema 14.2. Si tienen valor finito las esperanzas $E(e^{aX})$ y $E(e^{-aX})$ para $a > 0$, entonces, para $t \in \mathbb{R}$ con $|t| \leq a$, existen con valor finito:

- $g(t) = E(e^{tX})$,
- los momentos $\alpha_n = E(X^n)$,
- la serie $\sum_{n \geq 0} \frac{\alpha_n}{n!} t^n$,

y se tiene la igualdad

$$g(t) = E(e^{tX}) = \sum_{n=0}^{\infty} \frac{\alpha_n}{n!} t^n.$$

Observación. Si X es una v.a. acotada, las hipótesis se cumplen para cualquier $a > 0$.

Demostración. Como

$$a|X| = \begin{cases} aX & X \geq 0 \\ -aX & X < 0, \end{cases}$$

para $|t| \leq a$, se tiene $e^{tX} \leq e^{|t||X|} \leq e^{a|X|} \leq e^{aX} + e^{-aX}$. Además, como ambos sumandos tienen esperanza finita por hipótesis, todas las v.a. no negativas que acoten superiormente tendrán esperanza finita. En particular $E(e^{tX}) < \infty$, quedando demostrado el primer punto.

Para todo $n \in \mathbb{N}$ se tiene que

$$|X^n| = |X|^n = \frac{n! a^n |X|^n}{a^n n!} \leq \frac{n!}{a^n} \sum_{r=0}^{\infty} \frac{a^r |X|^r}{r!} = \frac{n!}{a^n} e^{a|X|},$$

que por lo anterior ($E(e^{a|X|}) < \infty$) tiene esperanza finita, por lo que $\alpha_n = E(X^n)$ también es finita. Queda demostrado el segundo punto.

Sea ahora la sucesión de v.v.a.a. $X_n := \sum_{r=0}^n \frac{X^r}{r!} t^r$ con $|t| \leq a$, tenemos $\lim_n X_n = e^{tX}$ y

$$|X_n| = \left| \sum_{r=0}^n \frac{X^r}{r!} t^r \right| \leq \sum_{r=0}^n \frac{|t|^r |X|^r}{r!} \leq \sum_{r=0}^n \frac{a^r |X|^r}{r!} \leq e^{a|X|},$$

de modo que podemos aplicar el Teorema de la convergencia dominada de Lebesgue (TCDL) obteniendo

$$g(t) = E(e^{tX}) = E\left(\lim_n \sum_{r=0}^n \frac{X^r}{r!} t^r\right) \stackrel{\text{TCDL}}{=} \lim_n E\left(\sum_{r=0}^n \frac{X^r}{r!} t^r\right) = \lim_n \sum_{r=0}^n \frac{E(X^r)}{r!} t^r = \sum_{n=0}^{\infty} \frac{E(X^n)}{n!} t^n = \sum_{n=0}^{\infty} \frac{\alpha_n}{n!} t^n$$

□

A continuación presentamos una suerte de recíproco del teorema anterior.

Teorema 14.3. *Si existen todos los momentos $\alpha_n = E(X^n)$ de la v.a. X con valor finito y la serie $\sum_{n=0}^{\infty} \frac{\alpha_n}{n!} t^n$ es convergente para $|t| \leq a$, entonces $g(t) = E(e^{tX})$ es finita para $|t| \leq a$ y vale $\sum_{n=0}^{\infty} \frac{\alpha_n}{n!} t^n$.*

14.4. Función generatriz de momentos factoriales

Se llama **función generatriz de momentos factoriales de X** a la función $h_X(t) = E((1+t)^X)$ (escribiremos simplemente $h(t)$ cuando esté claro) limitando el dominio de h a donde dicha expresión tenga valor finito.

Teorema 14.4. *Si para la v.a. X son finitas $E((1+a)^X)$ y $E((1-a)^X)$ con $0 < a < 1$ entonces todas las derivadas de la función*

$$h_X(t) = E((1+t)^X)$$

son $\frac{\partial^n}{\partial t^n} h_X(t) = h_X^{(n)}(t) = E(X^{(n)} (1+t)^{X-n})$ con $|t| \leq a$ y $n = 1, 2, \dots$

Con las hipótesis anteriores, si $h_X(t)$ es desarrollada como una serie de potencias de t en un entorno del origen, entonces viene dado por $h_X(t) = \sum_{n \geq 0} \frac{\alpha_{(n)}}{n!} t^n$ con $\alpha_{(n)} = E(X^{(n)})$ momento factorial de orden n de X .

14.5. Función generatriz de suma de variables aleatorias independientes

Teorema 14.5. Sean X_1, \dots, X_n v.a. independientes. Sean g_r y h_r respectivamente a la función generatriz de momentos y a la función generatriz de momentos factoriales de X_r . Llamando $S_n = \sum X_r$, resulta

$$g_{S_n}(t) := E(e^{tS_n}) = \prod g_r(t)$$
$$h_{S_n}(t) := E((1+t)^{S_n}) = \prod h_r(t),$$

siendo válidas las igualdades donde las correspondientes funciones estén definidas.

Demostración. Procediendo como en el teorema , obtenemos trivialmente:

$$E(e^{tS_n}) = E(e^{t \sum X_r}) = E\left(\prod e^{tX_r}\right) = \prod E(e^{tX_r}) = \prod g_r(t)$$
$$E((1+t)^{S_n}) = E((1+t)^{\sum X_r}) = E\left(\prod (1+t)^{X_r}\right) = \prod E((1+t)^{X_r}) = \prod h_r(t).$$

□

15. Función Característica

15.1. Función característica. Primeras propiedades

Dada una v.a. X con función de distribución F , se llama **función característica de la v.a. X** a $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ definida por

$$\varphi(t) = E(e^{itX}) = E(\cos(tX)) + iE(\sin(tX)) = \int_{\mathbb{R}} \cos(tx)dF(x) + i \int_{\mathbb{R}} \sin(tx)dF(x)$$

Observación. Es clara la igualdad $\varphi(t) = g(it)$ (donde g es la función generatriz de momentos). Además, como el seno y el coseno son funciones continuas y acotadas en \mathbb{R} , las anteriores esperanzas existen siempre.

Si X es de tipo discreto escribimos

$$\varphi(t) = \sum p_r e^{itx_r}, p_r = P(X = x_r)$$

Si es de tipo continuo y f es su función de densidad escribimos

$$\varphi(t) = \int f(x)e^{itx}$$

Teorema 15.1. Para toda función característica φ de una v.a. X se cumple que

1. $\varphi(0) = 1$
2. $\forall t \in \mathbb{R}, |\varphi(t)| \leq 1$
3. φ es uniformemente continua en \mathbb{R}
4. $\varphi(-t) = \overline{\varphi(t)}$. O sea, si la descomposición en parte real y parte imaginaria de φ es

$$\varphi(t) = \varphi_1(t) + i\varphi_2(t)$$

entonces φ_1 es una función par y φ_2 es impar.

5. Si $\varphi(t_0) = e^{i\alpha}$ con $t_0 \neq 0$ y $0 \leq \alpha < 2\pi$, entonces la probabilidad de la distribución de X está concentrada en $C = \left\{ \frac{\alpha + 2k\pi}{t_0}, k \in \mathbb{Z} \right\}$

Demostración. Veamos una a una:

1. $\varphi(0) = E(e^0) = E(1) = 1$.
2. $|\varphi(t)| = |E(e^{itX})| \leq E(|e^{itX}|) = E(1) = 1$.
- 3.

$$\begin{aligned} |\varphi(t+h) - \varphi(t)| &= \left| E(e^{i(t+h)X}) - E(e^{itX}) \right| = \left| E(e^{itX} e^{ihX} - e^{itX}) \right| \\ &= \left| E(e^{itX} (e^{ihX} - 1)) \right| \leq E(|e^{itX}| |e^{ihX} - 1|) = E(|e^{ihX} - 1|) \end{aligned}$$

y la cota no depende de t .

Ahora, $|e^{ihX} - 1| \leq |e^{ihX}| + 1 = 2$ y $\lim_{h \rightarrow 0} |e^{ihX} - 1| = 0$ por lo que estamos en condiciones de aplicar el Teorema de convergencia dominada de Lebesgue 11.4 obteniendo

$$0 \leq \lim_{h \rightarrow 0} |\varphi(t+h) - \varphi(t)| \leq \lim_{h \rightarrow 0} E(|e^{ihx} - 1|) \stackrel{TCDL}{=} E(\lim_{h \rightarrow 0} |e^{ihx} - 1|) = E(0) = 0.$$

Así, $\forall \varepsilon > 0, \exists \delta : [h < \delta \implies |\varphi(t+h) - \varphi(t)| \leq E(|e^{ihx} - 1|) < \varepsilon]$.

4. $\varphi(-t) = E(e^{-itX}) = E(\cos(-tX)) + iE(\sin(-tX)) = E(\cos(tX)) - iE(\sin(tX)) = \overline{\varphi(t)}$.

5. Si $\varphi(t_0) = E(e^{it_0X}) = e^{i\alpha}$, tenemos entonces que $E(e^{-i\alpha} e^{it_0X}) = e^{-i\alpha} E(e^{it_0X}) = 1$. Pero

$$\begin{aligned} E(\cos(t_0X - \alpha)) + iE(\sin(t_0X - \alpha)) &= E(e^{i(t_0X - \alpha)}) = 1 \iff \\ \iff E(\cos(t_0X - \alpha)) = 1 &\iff E(1 - \cos(t_0X - \alpha)) = 0 \stackrel{1-\cos \geq 0}{\iff} \cos(t_0X - \alpha) = 1 \text{ c.s.} \iff \\ \iff t_0X - \alpha = 2k\pi, k \in \mathbb{Z} \text{ c.s.} &\iff X = \frac{\alpha + 2k\pi}{t_0}, k \in \mathbb{Z} \text{ c.s.} \end{aligned}$$

□

Teorema 15.2. Sea X una v.a. con función característica φ_X , la función característica de la v.a. $Y = a + bX$, donde a, b son constantes, es $\varphi_Y(t) = e^{ita} \varphi_X(bt)$

Demostración. En efecto, $\varphi_Y(t) = E(e^{itY}) = E(e^{ita} e^{itbX}) = e^{ita} E(e^{itbX}) = e^{ita} \varphi_X(bt)$. □

Teorema 15.3. Sean X_1, \dots, X_n v.a. independientes, entonces la función característica de $S_n = X_1 + \dots + X_n$ es el producto de las funciones características φ_{X_r} .

Demostración. $\varphi_{S_n}(t) = E(e^{itS_n}) = E(e^{it \sum X_r}) = E(\prod e^{itX_r}) = \prod E(e^{itX_r}) = \prod \varphi_{X_r}(t)$, donde hemos usado, como tantas otras veces, que las funciones medibles mantienen la independencia. □

Ejemplo. La distribución uniforme en el intervalo $(a-h, a+h)$ tiene la función de densidad

$$f(x) = \begin{cases} \frac{1}{2h} & a-h < x < a+h \\ 0 & \text{otro caso} \end{cases}$$

Y la función característica es

$$\varphi(t) = \int_{a-h}^{a+h} e^{itx} \frac{dx}{2h} = \left[\frac{1}{2h} \frac{e^{itx}}{it} \right]_{a-h}^{a+h} = \frac{1}{2h} \frac{e^{it(a+h)} - e^{it(a-h)}}{it} = \frac{e^{ita} (e^{ith} - e^{-ith})}{2ith} = e^{iat} \frac{\sin(ht)}{ht}$$

Si $t = 0$, hay que tomar $\varphi(0) = 1$.

15.2. Teorema de Inversión

Teorema 15.4 (Teorema de Inversión). *Sea φ la función característica asociada a la función de distribución F . Entonces $\forall a, b \in \mathbb{R}$ se tiene que*

$$\frac{F(b) + F(b^-)}{2} - \frac{F(a) + F(a^-)}{2} = \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{2\pi it} \varphi(t) dt$$

Demostración. Si $a = b$ entonces ambos miembros de la expresión son nulos. Además si la igualdad es cierta para $a < b$ entonces es cierta para $b < a$ pues basta multiplicar ambos lados por -1 . Supongamos pues que $a < b$.

Empecemos por desarrollar la integral. Por la definición de φ , tenemos

$$\int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{2\pi it} \varphi(t) dt = \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{2\pi it} \left(\int_{\mathbb{R}} e^{itx} dF(x) \right) dt = \int_{-T}^T \left(\int_{\mathbb{R}} \frac{e^{-iat} - e^{-ibt}}{2\pi it} e^{itx} dF(x) \right) dt.$$

Ahora nos interesa utilizar Fubini⁹ para conmutar el orden de integración. Para comprobar que el integrando es medible es suficiente ver que es continuo y para ver que la integral de su módulo es finita nos bastará con una buena acotación.

Veamos la continuidad. Ya sabemos que φ es (uniformemente) continua. Además, $\frac{e^{-iat} - e^{-ibt}}{2\pi it}$ es claramente continua en $\mathbb{R} \setminus \{0\}$ y

$$\begin{aligned} \frac{e^{-iat} - e^{-ibt}}{2\pi it} &= e^{-ibt} \frac{e^{i(b-a)t} - 1}{2\pi it} = e^{-ibt} e^{i(b-a)t/2} \frac{e^{i(b-a)t/2} - e^{-i(b-a)t/2}}{2\pi it} = \\ &= e^{-i(a+b)t/2} \frac{\cos((b-a)t/2) + i \sin((b-a)t/2) - \cos(-(b-a)t/2) - i \sin(-(b-a)t/2)}{2\pi it} = e^{-i(a+b)t/2} \frac{2i \sin((b-a)t/2)}{2\pi it} \\ &= e^{-i(a+b)t/2} \frac{\sin((b-a)t/2)}{\pi t} = \frac{b-a}{2\pi} e^{-i(b+a)t/2} \frac{\sin((b-a)t/2)}{(b-a)t/2}, \end{aligned}$$

cuyo límite cuando $t \rightarrow 0$ es $(b-a)/2\pi$. Por tanto, puede extenderse la continuidad a todo \mathbb{R} .

Veamos ahora cómo podemos acotar el integrando. Sabemos que $|\varphi(t)| \leq 1$, por lo que es suficiente encontrar una cota para el otro factor. Pero utilizando la reescritura anterior es trivial ver

$$\left| \frac{e^{-iat} - e^{-ibt}}{2\pi it} \right| = \left| \frac{b-a}{2\pi} \right| \left| e^{-i(b+a)t/2} \right| \left| \frac{\sin((b-a)t/2)}{(b-a)t/2} \right| = \frac{b-a}{2\pi} \left| \frac{\sin((b-a)t/2)}{(b-a)t/2} \right| \leq \frac{b-a}{2\pi}.$$

⁹El teorema de Fubini establece que una función medible $g : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ cuyo valor absoluto podemos integrar a lo largo de su dominio con valor resultante finito, es decir,

$$\int_{\Omega_1 \times \Omega_2} |g| d\mu < \infty,$$

cumple

$$\int_{\Omega_1 \times \Omega_2} g d\mu = \int_{\Omega_1} \left(\int_{\Omega_2} g d\mu_2 \right) d\mu_1 = \int_{\Omega_2} \left(\int_{\Omega_1} g d\mu_1 \right) d\mu_2,$$

donde μ, μ_1, μ_2 son las medidas correspondientes. Para funciones no negativas no es necesaria la segunda hipótesis, por lo que esta puede comprobarse utilizando directamente la igualdad final.

En la demostración del teorema de inversión estamos aplicando Fubini sobre una función compleja y el teorema de Fubini expuesto aquí es sobre una función real, pero aun así supondremos que no se está cometiendo ninguna actividad fraudulenta.

En consecuencia podemos acotar la integral del módulo como sigue

$$\begin{aligned} \int_{-T}^T \left(\int_{\mathbb{R}} \left| \frac{e^{-iat} - e^{-ibt}}{2\pi it} e^{itx} \right| dF(x) \right) dt &\leq \int_{-T}^T \left(\int_{\mathbb{R}} \frac{b-a}{2\pi} dF(x) \right) dt = \\ &= \frac{b-a}{2\pi} \int_{-T}^T \left(\int_{\mathbb{R}} dF(x) \right) dt = \frac{b-a}{2\pi} \int_{-T}^T dt = \frac{b-a}{\pi} T \end{aligned}$$

Así, ya estamos en condiciones de aplicar el teorema de Fubini:

$$\int_{-T}^T \left(\int_{\mathbb{R}} \frac{e^{-iat} - e^{-ibt}}{2\pi it} e^{itx} dF(x) \right) dt = \int_{\mathbb{R}} \left(\int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{2\pi it} e^{itx} dt \right) dF(x).$$

Desarrollemos la integral interior:

$$\begin{aligned} J_T(x) &:= \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{2\pi it} e^{itx} dt = \int_{-T}^T \frac{e^{i(x-a)t} - e^{i(x-b)t}}{2\pi it} dt = \\ &= \int_{-T}^T \frac{\cos((x-a)t) - \cos((x-b)t)}{2\pi it} dt + i \int_{-T}^T \frac{\sin((x-a)t) - \sin((x-b)t)}{2\pi it} dt = \\ &= \frac{1}{i} \int_{-T}^T \frac{\cos((x-a)t) - \cos((x-b)t)}{2\pi t} dt + \int_{-T}^T \frac{\sin((x-a)t) - \sin((x-b)t)}{2\pi t} dt. \end{aligned}$$

El primer integrando es impar en t , por lo que la integral es nula. El segundo integrando, sin embargo, es par, por lo que podemos concluir:

$$J_T(x) = 2 \int_0^T \frac{\sin((x-a)t) - \sin((x-b)t)}{2\pi t} dt = \frac{1}{\pi} \left(\int_0^T \frac{\sin((x-a)t)}{t} dt - \int_0^T \frac{\sin((x-b)t)}{t} dt \right).$$

Ahora, haciendo los cambios $(x-a)t = u$ y $(x-b)t = v$, tenemos

$$J_T(x) = \frac{1}{\pi} \left(\int_0^{(x-a)T} \frac{\sin(u)}{u} du - \int_0^{(x-b)T} \frac{\sin(v)}{v} dv \right).$$

Una vez que hemos desarrollado J_T lo suficiente, nos gustaría poder aplicarle el teorema de la convergencia dominada para meter el límite (respecto a T) original del enunciado y así simplificar al máximo el cálculo de la integral respecto de $dF(x)$.

Definimos $g(v) := \int_0^v \frac{\sin(u)}{u} du$, que es una función continua y acotada definida en todo \mathbb{R} . Sea $K \in \mathbb{R}$ tal que $|g| \leq K$, tenemos que las funciones $J_T(x) = \frac{1}{\pi} (g((x-a)T) - g((x-b)T))$ son continuas (y por lo tanto integrables) y $|J_T| \leq \frac{2K}{\pi}$. Veamos ahora a qué función convergen puntualmente nuestras funciones J_T cuando $T \rightarrow \infty$; usando $\lim_{v \rightarrow \infty} g(v) = \frac{\pi}{2}$ y $g(0) = 0$ es fácil ver que

- $x < a < b \implies \lim_T J_T = \frac{1}{\pi} \left(-\frac{\pi}{2} - \left(-\frac{\pi}{2}\right) \right) = 0.$
- $x = a < b \implies \lim_T J_T = \frac{1}{\pi} \left(0 + \frac{\pi}{2} \right) = \frac{1}{2}.$
- $a < x < b \implies \lim_T J_T = \frac{1}{\pi} \left(\frac{\pi}{2} + \frac{\pi}{2} \right) = 1.$
- $a < x = b \implies \lim_T J_T = \frac{1}{\pi} \left(\frac{\pi}{2} + 0 \right) = \frac{1}{2}.$

$$\blacksquare a < b < x \implies \lim_T J_T = \frac{1}{\pi} \left(\frac{\pi}{2} - \left(\frac{\pi}{2} \right) \right) = 0.$$

En conclusión, J_T converge puntualmente a la función

$$J(x) = \begin{cases} 1 & a < x < b \\ \frac{1}{2} & x = a \text{ ó } x = b \\ 0 & \text{otro caso} \end{cases}$$

Finalmente podemos aplicar el teorema de convergencia dominada de Lebesgue, obteniendo

$$\begin{aligned} \lim_{T \rightarrow \infty} \int_T^T \frac{e^{-iat} - e^{-ibt}}{2\pi it} \varphi(t) dt &= \lim_{T \rightarrow \infty} \int_{-\infty}^{\infty} J_T(x) dF(x) = \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} J_T(x) dF(x) = \int_{-\infty}^{\infty} J(x) dF(x) = \\ &= \int_{-\infty}^a 0 dF(x) + \int_{\{a\}} \frac{1}{2} dF(x) + \int_a^b 1 dF(x) + \int_{\{b\}} \frac{1}{2} dF(x) + \int_b^{\infty} 0 dF(x) = \frac{1}{2} \int_{\{a\}} dF(x) + \int_a^b dF(x) + \frac{1}{2} \int_{\{b\}} dF(x) = \\ &= \frac{1}{2}(F(a) - F(a^-)) + F(b^-) - F(a) + \frac{1}{2}(F(b) - F(b^-)) = \frac{F(b) + F(b^-)}{2} - \frac{F(a) + F(a^-)}{2}. \end{aligned}$$

□

Teorema 15.5. Una función característica es función característica de una sola función de distribución. En otras palabras, la correspondencia entre funciones de distribución y funciones características es una biyección.

Demostración. Por la fórmula de inversión, obtenemos la igualdad

$$F(x) = \lim_{b \downarrow x} \lim_{a \rightarrow -\infty} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{2\pi it} \varphi(t) dt$$

Y como el miembro de la derecha está determinado por la función característica, resulta que la función de distribución queda determinada en todos sus puntos por la función característica. □

15.3. Función característica integrable

Teorema 15.6. Sea φ la función característica de la función de distribución F . Si existe con valor finito la integral

$$\int_{\mathbb{R}} |\varphi(t)| dt < \infty$$

entonces F es continua y tiene derivada continua, que puede calcularse por la fórmula

$$F'(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itX} \varphi(t) dt$$

Demostración. Con las hipótesis del teorema el límite que aparece en la fórmula de inversión puede suprimirse escribiendo

$$\frac{F(b) + F(b^-)}{2} - \frac{F(a) + F(a^-)}{2} = \int_{\mathbb{R}} \frac{e^{-iat} - e^{-ibt}}{2\pi it} \varphi(t) dt$$

ya que

$$\lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{2\pi it} \varphi(t) dt = \lim_{T \rightarrow \infty} \int_{\mathbb{R}} \frac{e^{-iat} - e^{-ibt}}{2\pi it} \varphi(t) \cdot I_{(-T, T)}(t) dt$$

donde $I_{(-T, T)}(t)$ es la función indicador del intervalo $(-T, T)$ y se puede aplicar el TCDL (11.4) al límite interior, ya que vale la acotación (como vimos en la demostración del Teorema de Inversión)

$$\left| \frac{e^{-iat} - e^{-ibt}}{2\pi it} I_{(-T, T)}(t) \varphi(t) \right| \leq \frac{b-a}{2\pi} |\varphi(t)|$$

y esta última función es independiente de T y su integral es finita en \mathbb{R} . Por tanto, el límite anterior nos da la igualdad que queríamos al principio.

Ahora, para ver que F es continua, supongamos x fijo, y tomemos $x+h$ (punto de continuidad de F) con $h > 0$. De este modo, la fórmula de inversión nos da

$$F(x+h) - \frac{F(x) + F(x^-)}{2} = \int_{\mathbb{R}} \frac{e^{-ixt} - e^{-i(x+h)t}}{2\pi it} \varphi(t) dt$$

y por la desigualdad mencionada antes, obtenemos

$$\left| F(x+h) - \frac{F(x) + F(x^-)}{2} \right| \leq \int_{\mathbb{R}} \left| \frac{e^{-ixt} - e^{-i(x+h)t}}{2\pi it} \varphi(t) \right| dt \leq \frac{h}{2\pi} \int_{\mathbb{R}} |\varphi(t)| dt$$

y tomando límites cuando $h \rightarrow 0$, obtenemos

$$F(x) - \frac{F(x) + F(x^-)}{2} = 0 \iff F(x) = F(x^-)$$

luego F es continua.

Para obtener su derivada, tenemos

$$\frac{F(x+h) - F(x)}{h} \stackrel{\text{inversión}}{=} \frac{1}{h} \int_{\mathbb{R}} \frac{e^{-ixt} - e^{-i(x+h)t}}{2\pi it} \varphi(t) dt$$

En la demostración del teorema de inversión, vimos que

$$\frac{e^{-iat} - e^{-ibt}}{2\pi it} = \frac{b-a}{2\pi} \exp\left(\frac{-i(a+b)t}{2}\right) \frac{\sin\left(\frac{(b-a)t}{2}\right)}{\frac{(b-a)t}{2}}$$

y esto podemos aplicarlo a nuestro caso, quedando

$$\begin{aligned} \frac{F(x+h) - F(x)}{h} &= \frac{1}{h} \int_{\mathbb{R}} \frac{h}{2\pi} \exp\left(\left(-x - \frac{h}{2}\right)it\right) \frac{\sin\left(\frac{ht}{2}\right)}{\frac{ht}{2}} \varphi(t) dt = \\ &= \int_{\mathbb{R}} \frac{1}{2\pi} \exp\left(\left(-x - \frac{h}{2}\right)it\right) \frac{\sin\left(\frac{ht}{2}\right)}{\frac{ht}{2}} \varphi(t) dt \end{aligned}$$

El módulo del integrando está acotado por $\frac{|\varphi(t)|}{2\pi}$, que es independiente de h y tiene integral finita. Por tanto, podemos tomar límites para $h \rightarrow 0$, obteniendo

$$F'(x) = \int_{\mathbb{R}} \frac{1}{2\pi} \exp(-xit) \varphi(t) dt$$

(recordemos que $\lim_{x \rightarrow 0} \frac{\sin(x)}{x} = 1$)

Y hemos obtenido el resultado buscado.

Por último, la continuidad de F' es inmediata, ya que su integrando está acotado por $|\varphi(t)|$, independientemente de x , con integral finita y e^{-ixt} es una función continua. \square

15.4. Derivabilidad de la función característica

Teorema 15.7. *Si existe con valor finito el momento de orden n de la variable aleatoria X , entonces la función característica φ de X admite derivadas hasta el orden n y las derivadas valen*

$$\varphi^{(r)}(t) = \int (ix)^r e^{itx} dF(x) = i^r \int x^r e^{itx} dF(x) \quad (1)$$

para $r = 1, 2, \dots, n$.

Resulta de esta fórmula que los momentos están ligados con la función característica por la relación

$$\varphi^{(r)}(0) = i^r \alpha_r$$

Demostración. La existencia del momento de orden n de X asegura la existencia de los momentos

$$\alpha_r = E(X^r) = \int x^r dF(x)$$

para $r = 1, \dots, n$.

Vamos a ver el resultado por inducción, donde el caso $r = 0$, ya lo conocemos.

Admitamos, entonces, que existe la derivada de φ de un orden $r < n$ y que la fórmula (1) funciona. Demostraremos que existe la derivada de orden $r + 1$ y que esta derivada se obtiene por esta misma fórmula, incrementando r en una unidad.

En efecto, se tiene

$$\begin{aligned} \frac{\varphi^{(r)}(t+h) - \varphi^{(r)}(t)}{h} &= \frac{1}{h} \int (ix)^r (e^{i(t+h)x} - e^{itx}) dF(x) = \frac{1}{h} \int (ix)^r e^{itx} e^{\frac{ihx}{2}} (e^{\frac{ihx}{2}} - e^{-\frac{ihx}{2}}) dF(x) = \\ &= \frac{1}{h} \int (ix)^r e^{itx} e^{\frac{ihx}{2}} \left(2i \sin\left(\frac{hx}{2}\right) \right) dF(x) = \int (ix)^{r+1} e^{i\left(t+\frac{h}{2}\right)x} \frac{\sin\left(\frac{hx}{2}\right)}{\frac{hx}{2}} dF(x) \end{aligned}$$

Este último integrando tiene módulo acotado por $|x^{r+1}|$, que es independiente de h y tiene integral finita, por lo que podemos tomar límites para $h \rightarrow 0$, obteniendo

$$\varphi^{(r+1)}(t) = \int (ix)^{r+1} e^{itx} dF(x)$$

que es, precisamente, (1). \square

Teorema 15.8. *Si existe la derivada en el origen $\varphi^{2n}(0)$ de orden $2n$ de la función característica φ de la variable aleatoria X , existe con valor finito el momento de orden $2n$ de X .*

Una propiedad más precisa que estos teoremas, es la siguiente:

Hecho 15.1. Condición necesaria y suficiente para que exista la derivada $\varphi^{(r)}(0)$ de la función característica φ de la función de distribución F , es que valga

$$\lim_{a \rightarrow \infty} a^r (1 - F(a) + F(-a)) = 0$$

y que existe

$$\lim_{a \rightarrow \infty} \int_{-a}^a x^r dF(x) = m$$

Y en este caso, se tiene

$$\varphi^{(r)}(0) = i^r m$$

Teorema 15.9. Sea φ la función característica de la variable aleatoria X . Si existen todos los momentos $\alpha_n = E(X^n)$, y la serie

$$\sum_{n=0}^{\infty} \frac{\alpha_n}{n!} (it)^n$$

es convergente para $|t| < R$, entonces φ es desarrollable en serie de potencias para $|t| < R$ y vale

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{\alpha_n}{n!} (it)^n, \quad |t| < R$$

Observación. Por el teorema 14.2 sabemos que, si existe $E(e^{aX})$ y $E(e^{-aX})$ para un $a > 0$, entonces existen todos los momentos de la variable aleatoria X y la serie

$$\sum_{n=0}^{\infty} \frac{\alpha_n}{n!} t^n$$

es convergente para $|t| < a$. Por tanto, por el teorema anterior, resulta que, si existen $E(e^{aX})$ y $E(e^{-aX})$ para un $a > 0$, entonces

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{\alpha_n}{n!} (it)^n, \quad |t| < a$$

15.5. Distribuciones simétricas

Teorema 15.10. Condición necesaria y suficiente para que la variable aleatoria X tenga distribución simétrica respecto del origen es que la función característica φ de X sea real.

Más general, la variable aleatoria X tiene distribución simétrica respecto de c si la función característica φ de X se descompone en el producto

$$\varphi(t) = e^{ict} \psi(t)$$

donde $\psi(t)$ es una función característica real, y recíprocamente.

Demostración. Si X tiene distribución simétrica respecto del origen, X y $-X$ tienen la misma distribución y, por tanto, la misma función característica y se tiene

$$\varphi(t) = E(e^{itX}) = E(e^{it(-X)}) = E(e^{i(-t)X}) = \varphi(-t) = \overline{\varphi(t)}$$

por lo que φ debe ser real.

Si la variable aleatoria tiene distribución simétrica respecto de c , las variables aleatorias $Y = X - c$ y $-Y = c - X$ tienen la misma distribución, luego Y es simétrica respecto del origen y se tiene, ya que $X = Y + c$:

$$\varphi(t) = E(e^{itX}) = E(e^{it(Y+c)}) = e^{ict} E(e^{itY}) = e^{ict} \psi(t)$$

donde $\psi(t)$ es una función característica real, por lo visto antes. □

16. Funciones generatrices (varias dimensiones)

16.1. Función generatriz de probabilidad de varias dimensiones

Veamos a continuación la generalización del concepto de función generatriz de probabilidad de una variable aleatoria al caso multidimensional. En lo que sigue, decimos que una variable aleatoria multidimensional es entera no negativa si sus componentes son, precisamente, variables aleatorias unidimensionales enteras no negativas.

Definición 16.1. Sea $X = (X_1, \dots, X_k)$ una variable aleatoria entera no negativa k -dimensional y sea también el conjunto $C := \left\{ t := (t_1, \dots, t_k) \in \mathbb{R}^k : E\left(t_1^{X_1} t_2^{X_2} \dots t_k^{X_k}\right) < +\infty \right\}$. La función

$$\begin{aligned} f : C &\rightarrow \mathbb{R} \\ t &\mapsto E\left(t_1^{X_1} t_2^{X_2} \dots t_k^{X_k}\right) \end{aligned}$$

se llama **función generatriz de probabilidad de la variable aleatoria X** .

Proposición 16.1. *Llamando*

$$p_{i_1, \dots, i_k} := P(X_1 = i_1, \dots, X_k = i_k),$$

con $i_j \in \{0, 1, \dots\}$ para cada $j \in \{1, 2, \dots, k\}$, podemos escribir

$$f(t) = \sum_{i_1, \dots, i_k} p_{i_1, \dots, i_k} t_1^{i_1} \dots t_k^{i_k}.$$

Es decir, f es, o bien un polinomio en las variables t_1, \dots, t_k y $C = \mathbb{R}^k$, o bien una serie de potencias en dichas variables. En cualquier caso $[-1, 1]^k \subset C$, por ser $(1, 1, \dots, 1) \in C \iff \sum_{i_1, \dots, i_k} p_{i_1, \dots, i_k} = 1$.

Evidentemente el recíproco también es cierto, es decir, si dada una función

$$\begin{aligned} g : C &\rightarrow \mathbb{R} \\ t &\mapsto \sum_{i_1, \dots, i_k} q_{i_1, \dots, i_k} t_1^{i_1} \dots t_k^{i_k}, \end{aligned}$$

los coeficientes q_{i_1, \dots, i_k} verifican las condiciones

$$q_{i_1, \dots, i_k} \geq 0, \quad \sum_{i_1, \dots, i_k} q_{i_1, \dots, i_k} = 1$$

entonces g es una función generatriz de probabilidad con probabilidades determinadas por dichos coeficientes.

Dadas las n variables aleatorias k -dimensionales

$$Z_j = (Z_{j1}, Z_{j2}, \dots, Z_{jk}) \quad j = 1, 2, \dots, n$$

denotaremos su suma con

$$S_n := (S_{n1}, S_{n2}, \dots, S_{nk}) := Z_1 + Z_2 + \dots + Z_n := \left(\sum_{j=1}^n Z_{j1}, \sum_{j=1}^n Z_{j2}, \dots, \sum_{j=1}^n Z_{jk} \right),$$

es decir,

$$S_{nr} = \sum_{j=1}^n Z_{jr} \quad r = 1, 2, \dots, k.$$

Teorema 16.1. Sean Z_1, \dots, Z_n variables aleatorias enteras no negativas k -dimensionales independientes y sea también $S_n = \sum_{j=1}^n Z_j$. La función generatriz de probabilidad f de S_n es el producto de las funciones generatrices de probabilidad f_j de las variables aleatorias Z_j , al menos en el intervalo $[-1, 1]^k$.

Demostración. Sea $t = (t_1, \dots, t_k) \in \mathbb{R}^k$. Se tiene que

$$E\left(t_1^{S_{n1}} \dots t_k^{S_{nk}}\right) = E\left(t_1^{\sum_j Z_{j1}} \dots t_k^{\sum_j Z_{jk}}\right) = E\left(\prod_{j=1}^n t_1^{Z_{j1}} \dots t_k^{Z_{jk}}\right) \stackrel{Z_j \text{ indep.}}{=} \prod_{j=1}^n E\left(t_1^{Z_{j1}} \dots t_k^{Z_{jk}}\right)$$

y como $f(t) = E\left(t_1^{S_{n1}} \dots t_k^{S_{nk}}\right)$ y $f_j(t) = E\left(t_1^{Z_{j1}} \dots t_k^{Z_{jk}}\right)$, esto es precisamente lo que queríamos probar. \square

Teorema 16.2. Sea $X = (X_1, X_2, \dots, X_k)$ una variable aleatoria entera no negativa k -dimensional con función generatriz f . Sean también las variables aleatorias

$$U = (X_1, \dots, X_h) \quad V = (X_{h+1}, \dots, X_k)$$

con $h < k$, que tienen funciones generatrices

$$f_U(t) = f(t_1, \dots, t_h, 1, \dots, 1) \quad f_V(t) = f(1, \dots, 1, t_{h+1}, \dots, t_k)$$

respectivamente. Entonces U y V son independientes si y sólo si se verifica que

$$f(t) = f(t_1, \dots, t_k) = f_U(t_1, \dots, t_h) \cdot f_V(t_{h+1}, \dots, t_k).$$

Proposición 16.2. Sea X una variable aleatoria entera no negativa k -dimensional y $f : C \rightarrow \mathbb{R}$ su función generatriz. Si $(1, \dots, 1)$ pertenece al interior de C , tomando r_1, \dots, r_k enteros no negativos cualesquiera y $r = r_1 + \dots + r_k$, se cumple que

$$\frac{\partial^r f}{\partial^{r_1} t_1 \dots \partial^{r_k} t_k} (1, \dots, 1) = E\left(X_1^{(r_1)} \dots X_k^{(r_k)}\right)$$

16.2. Función generatriz de momentos de variables aleatorias de varias dimensiones

Definición 16.2. Sea X una variable aleatoria k -dimensional y sea también el conjunto

$$C := \left\{ t \in \mathbb{R}^k : E(\exp(t_1 X_1 + \dots + t_k X_k)) < +\infty \right\}.$$

La función

$$\begin{aligned} g : C &\rightarrow \mathbb{R} \\ t &\mapsto E(\exp(t_1 X_1 + \cdots + t_k X_k)) \end{aligned}$$

se llama **función generatriz de momentos de X** .

Teorema 16.3. *Sea X una variable aleatoria k -dimensional, cuyas componentes poseen funciones generatrices de momentos definidas todas ellas, al menos, en un cierto intervalo $[-a, a]$, con $a > 0$. Entonces existen todos los momentos*

$$\alpha_{r_1, \dots, r_k} = E(X_1^{r_1} \cdots X_k^{r_k}),$$

la serie

$$\sum_{r_1 \geq 0, \dots, r_k \geq 0} \frac{\alpha_{r_1, \dots, r_k}}{r_1! \cdots r_k!} t_1^{r_1} \cdots t_k^{r_k}$$

es absolutamente convergente en un entorno del origen y en dicho entorno se verifica que

$$g(t) = E\left(\exp\left(\sum_{j=1}^k t_j X_j\right)\right) = \sum_{r_1 \geq 0, \dots, r_k \geq 0} \frac{\alpha_{r_1, \dots, r_k}}{r_1! \cdots r_k!} t_1^{r_1} \cdots t_k^{r_k}$$

16.3. Función característica

Definición 16.3. Sea X una variable aleatoria k -dimensional con función de distribución $F : \mathbb{R}^k \rightarrow \mathbb{R}$. La función $\varphi : \mathbb{R}^k \rightarrow \mathbb{C}$ definida por

$$\begin{aligned} \varphi(t) &= E\left(\exp\left(i \sum_{j=1}^k t_j X_j\right)\right) = \int \exp\left(i \sum_{j=1}^k t_j x_j\right) dF(x) = \\ &= E\left(\cos\left(\sum_{j=1}^k t_j X_j\right)\right) + i E\left(\sin\left(\sum_{j=1}^k t_j X_j\right)\right) = \\ &= \int \cos\left(\sum_{j=1}^k t_j X_j\right) dF(x) + i \int \sin\left(\sum_{j=1}^k t_j X_j\right) dF(x). \end{aligned}$$

se denomina **función característica de X** .

Utilizando la notación matricial

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}, \quad t = \begin{pmatrix} t_1 \\ \vdots \\ t_k \end{pmatrix}, \quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix},$$

la expresión de la función característica se puede escribir como

$$\varphi(t) = E(\exp(i \cdot t^t \cdot X)) = \int_{\mathbb{R}^k} \exp(i \cdot t^t \cdot X) dF(x) = \int_{\mathbb{R}^k} \cos(t^t \cdot X) dF(x) + i \int_{\mathbb{R}^k} \sin(t^t \cdot X) dF(x).$$

Teorema 16.4. Si φ es la función característica de una distribución k -dimensional, entonces se verifican las siguientes propiedades:

1. $\varphi(0) = 1$.
2. $|\varphi(t)| \leq 1$.
3. φ es uniformemente continua en \mathbb{R}^k .

Teorema 16.5. Toda función característica determina la función de distribución de la que procede por la fórmula de inversión

$$F(x) = \lim_{a \rightarrow -\infty} \lim_{b \downarrow x} J(a, b),$$

donde

$$J(a, b) = \lim_{T \rightarrow +\infty} \int_{-T}^T \cdots \int_{-T}^T \prod_{j=1}^k \frac{e^{-it_j a_j} - e^{-it_j b_j}}{2\pi i t_j} \varphi(t) dt,$$

ya que esta última expresión toma el valor

$$J(a, b) = \int_{\mathbb{R}} \prod_{j=1}^k f_j(x_j) dF(x) \quad f_j(x_j) = \begin{cases} 1 & \text{si } a_j < x_j < b_j \\ \frac{1}{2} & \text{si } x_j \in \{a_j, b_j\} \\ 0 & \text{en otro caso} \end{cases}$$

Teorema 16.6. Si φ es la función característica de $X = (X_1, \dots, X_k)$, entonces su conjugado complejo $\bar{\varphi}$ es la función característica de $-X$.

Teorema 16.7. Sean $A \in \mathcal{M}_{h \times k}(\mathbb{R})$ y $b \in \mathbb{R}^h$. Si las variables aleatorias X e Y , de k y h dimensiones respectivamente, están relacionadas por $Y = AX + b$, entonces también se verifica que

$$\varphi_Y(u) = e^{iu^t b} \varphi_X(A^t u).$$

Demostración.

$$E\left(e^{iu^t Y}\right) = E\left(e^{iu^t (AX+b)}\right) = E\left(e^{i(u^t AX + u^t b)}\right) = e^{iu^t b} E\left(e^{iu^t AX}\right) = e^{iu^t b} E\left(e^{i(A^t u)^t X}\right) = e^{iu^t b} \varphi_X(A^t u)$$

□

Teorema 16.8. Sean X e Y dos variables aleatorias k -dimensionales independientes. Entonces, tomando $Z = X + Y$, se tiene que

$$\varphi_Z(t) = \varphi_X(t) \varphi_Y(t)$$

Demostración. $\varphi_Z(t) = E\left(e^{it^t Z}\right) = E\left(e^{it^t (X+Y)}\right) = E\left(e^{it^t X} e^{it^t Y}\right) = E\left(e^{it^t X}\right) E\left(e^{it^t Y}\right) = \varphi_X(t) \varphi_Y(t).$

□

17. Algunas distribuciones notables

17.1. Distribución multinomial

Sean

$$Z_j = (Z_{j1}, \dots, Z_{jk}), \quad j = 1, \dots, n$$

n variables aleatorias independientes todas con la misma distribución, definida como sigue. Cada Z_j solo puede tomar los k valores

$$\begin{aligned} e_1 &= (1, 0, \dots, 0) \\ e_2 &= (0, 1, 0, \dots, 0) \\ &\dots \\ e_k &= (0, 0, \dots, 0, 1) \end{aligned}$$

con probabilidades

$$P(Z_j = e_r) = P(Z_j = (\delta_{r1}, \dots, \delta_{rk})) = p_r, \quad r = 1, \dots, k$$

donde

$$p_r \geq 0, \quad r = 1, \dots, k, \quad \sum_r p_r = 1$$

y δ_{ij} es la δ de Kronecker, definida por

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

La función generatriz de probabilidad de cada Z_j es

$$f_j(t) = p_1 t_1 + p_2 t_2 + \dots + p_k t_k$$

luego la función generatriz de probabilidad de la variable aleatoria suma

$$S_n = \sum_{j=1}^n Z_j = \left(\sum_{j=1}^n Z_{j1}, \dots, \sum_{j=1}^n Z_{jk} \right) = (S_{n1}, \dots, S_{nk})$$

es el producto de las $f_j(t)$ anteriores, es decir, la función generatriz de probabilidad de S_n es

$$f(t) = (p_1 t_1 + \dots + p_k t_k)^n$$

Utilizando aquí la fórmula de Leibniz, obtenemos

$$f(t) = f(t_1, \dots, t_k) = \sum_{r_1, \dots, r_k} \frac{n!}{r_1! r_2! \cdot \dots \cdot r_k!} p_1^{r_1} p_2^{r_2} \cdot \dots \cdot p_k^{r_k} t_1^{r_1} t_2^{r_2} \cdot \dots \cdot t_k^{r_k}$$

luego

$$P(S_n = (r_1, \dots, r_k)) = \frac{n!}{r_1! r_2! \cdot \dots \cdot r_k!} p_1^{r_1} p_2^{r_2} \cdot \dots \cdot p_k^{r_k}, \quad \text{con } \sum_i r_i = n$$

Y la distribución de esta variable S_n se llama **distribución multinomial**.

La **distribución marginal correspondiente a la primera componente** S_{n1} de la variable S_n queda determinada por la función generatriz de probabilidad que se obtiene haciendo en la función $f(t)$ los valores $t_2 = t_3 = \dots = t_k = 1$, con lo que obtenemos

$$f(t_1, 1, \dots, 1) = E(t_1^{S_{n1}} 1^{S_{n2}} \cdot \dots \cdot 1^{S_{nk}}) = E(t_1^{S_{n1}}) = (p_1 t_1 + p_2 + \dots + p_k)^n \stackrel{q_1 = p_2 + \dots + p_k}{=} (p_1 t_1 + q_1)^n$$

o sea, que S_{n1} es binomial de parámetros n y p_1 . Nótese que esto vale para cada una de las componentes, y se tiene

$$E(S_{nj}) = n \cdot p_j$$

$$Var(S_{nj}) = n \cdot p_j \cdot q_j, \quad q_j = 1 - p_j$$

Para la **distribución marginal de las dos primeras componentes** de S_n se tiene la función generatriz de probabilidad

$$f_{12}(t_1, t_2) = f(t_1, t_2, 1, \dots, 1) = E\left(t_1^{S_{n1}} t_2^{S_{n2}} 1^{S_{n3}} \dots 1^{S_{nk}}\right) = E\left(t_1^{S_{n1}} t_2^{S_{n2}}\right)^{q_{12} = p_3 + \dots + p_k} (p_1 t_1 + p_2 t_2 + q_{12})^n$$

y las funciones generatrices de probabilidad en estos casos son polinomios en las variables, ya que la esperanza E da sumas finitas. La derivación permuta con la suma finita, por lo que también lo hace con el símbolo esperanza, E (la derivación permuta con el operador E . Podemos 'meter la derivada dentro' de la esperanza porque esta no es más que una suma finita).

Para calcular la covarianza de S_{n1}, S_{n2} podemos calcular

$$\frac{\partial^2 f_{12}}{\partial t_1 \partial t_2} = E\left(S_{n1} t_1^{S_{n1}-1} S_{n2} t_2^{S_{n2}-1}\right) = n \cdot (n-1) \cdot p_1 \cdot p_2 \cdot (p_1 t_1 + p_2 t_2 + q_{12})^{n-2}$$

y, haciendo $t_1 = t_2 = 1$, se obtiene

$$E(S_{n1} S_{n2}) = n(n-1) p_1 p_2$$

$$Cov(S_{n1}, S_{n2}) = E(S_{n1} S_{n2}) - E(S_{n1}) E(S_{n2}) = n(n-1) p_1 p_2 - n^2 p_1 p_2 = -n p_1 p_2$$

En general, se tiene

$$Cov(S_{ni}, S_{nj}) = -n p_i p_j, \quad i \neq j$$

17.2. Distribución normal

Una variable aleatoria X de una dimensión tiene una **distribución normal con los parámetros μ y σ** , con $-\infty < \mu < \infty$ y $\sigma > 0$, si es de tipo continuo con función de densidad

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty$$

Que X es una variable aleatoria con distribución normal con la densidad mencionada lo expresaremos escribiendo « X tiene una distribución $N(\mu, \sigma^2)$ » o « X es $N(\mu, \sigma^2)$ ».

Para que esta definición tenga sentido, es preciso comprobar que esta función es, efectivamente, una función de densidad:

Demostración. Lo vamos a hacer primero para el caso en que tenemos una variable aleatoria Y que es $N(0, 1)$. En este caso, su presunta función de densidad queda

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right), \quad -\infty < y < \infty$$

Claramente, es no negativa.

Queremos, entonces, ver que su integral vale 1:

$$\int_{-\infty}^{\infty} f_Y(y) dy \stackrel{\text{par}}{=} 2 \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp\left(-\frac{y^2}{2}\right) dy \quad \begin{array}{l} y^2 = 2t \\ 2ydy = 2dt \implies dy = \frac{dt}{y} = \frac{dt}{\sqrt{2t}} \end{array} \quad 2 \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp(-t) \cdot \frac{1}{\sqrt{2}} t^{-\frac{1}{2}} dt =$$

$$\frac{\Gamma\left(\frac{1}{2}\right)}{\sqrt{\pi}} = 1$$

por lo que en el caso de ser $N(0, 1)$ sí que es una función de densidad.

Ahora, si Y es $N(0, 1)$, entonces la variable aleatoria

$$X = \sigma Y + \mu, \quad \sigma > 0$$

tiene la función de densidad

$$f_Y\left(\frac{x - \mu}{\sigma}\right) \frac{1}{\sigma} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

que es precisamente la de nuestra definición. Por tanto, es una función de densidad. \square

Además, hemos comprobado que si Y es $N(0, 1)$, la variable aleatoria $X = \sigma Y + \mu$ es $N(\mu, \sigma^2)$. Inversamente, si X tiene una distribución $N(\mu, \sigma^2)$, entonces $Y = \frac{X - \mu}{\sigma}$ tiene una distribución $N(0, 1)$.

El significado de los parámetros μ y σ lo da el siguiente teorema:

Teorema 17.1. Si X es $N(\mu, \sigma^2)$ entonces

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

Demostración. Como antes, empecemos suponiendo que Y es $N(0, 1)$. El momento de segundo orden de Y es

$$E(Y^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 \exp\left(-\frac{y^2}{2}\right) dy \stackrel{\text{par}}{=} 2 \frac{1}{\sqrt{2\pi}} \int_0^{\infty} y^2 \exp\left(-\frac{y^2}{2}\right) dy \quad \begin{array}{l} y^2 = 2t \\ dy = \frac{dt}{\sqrt{2t}} \end{array} \quad 2 \frac{1}{\sqrt{2\pi}} \int_0^{\infty} 2t \exp(-t) \frac{1}{\sqrt{2t}} dt =$$

$$= \frac{2}{\sqrt{\pi}} \int_0^{\infty} \exp(-t) t^{\frac{1}{2}} dt = \frac{2\Gamma\left(\frac{3}{2}\right)}{\sqrt{\pi}} = \frac{2\left[\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\right]}{\sqrt{\pi}} = 1$$

Y como la función de densidad de Y es simétrica respecto del origen, se tiene $E(Y) = 0$.

Por tanto

$$E(Y) = 0$$

$$Var(Y) = E(Y^2) - E(Y)^2 = 1$$

tal y como queríamos ver.

Ahora bien, si Y es $N(0, 1)$, entonces $X = \sigma Y + \mu$ es $N(\mu, \sigma^2)$ y se tiene

$$E(X) = \sigma E(Y) + \mu = \mu$$

$$Var(Y) = Var(\sigma Y + \mu) = Var(\sigma Y) = \sigma^2 Var(Y) = \sigma^2$$

□

Teorema 17.2. La función generatriz de momentos de una variable aleatoria X con distribución $N(\mu, \sigma^2)$ es

$$E(e^{tX}) = e^{\mu t + \frac{\sigma^2 t^2}{2}}, \quad t \in \mathbb{R}$$

y la función característica

$$E(e^{itX}) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}, \quad t \in \mathbb{R}$$

Demostración. Para simplificar, tomemos primero $Y = \frac{X - \mu}{\sigma}$, que es $N(0, 1)$. Su función generatriz de momentos es

$$E(e^{tY}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty} e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty - \frac{y^2}{2}} dy$$

Aquí notamos que

$$ty - \frac{y^2}{2} = -\frac{1}{2}(y^2 - 2ty) = -\frac{1}{2}(y^2 - 2ty + t^2 - t^2) = -\frac{1}{2}((y - t)^2 - t^2) = -\frac{1}{2}(y - t)^2 + \frac{t^2}{2}$$

y es

$$E(e^{tY}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(y - t)^2 + \frac{t^2}{2}\right) dy = \frac{e^{\frac{t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(y - t)^2\right) dy$$

y esta integral da el mismo resultado, independientemente de t , que es

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(y - t)^2\right) dy = \sqrt{2\pi}$$

por lo que queda

$$E(e^{tY}) = e^{\frac{t^2}{2}}, \quad t \in \mathbb{R}$$

Pero esta igualdad también es válida en el plano complejo (ambas son funciones analíticas que coinciden en toda la recta real), es decir

$$E(\exp(zY)) = \exp\left(\frac{z^2}{2}\right), \quad z \in \mathbb{C}$$

Entonces, tomando $z = it$, obtenemos la función característica de Y :

$$E(e^{itY}) = e^{-\frac{t^2}{2}}$$

Y aplicando estos resultados a la variable aleatoria

$$X = \sigma Y + \mu$$

obtenemos

$$E(e^{tX}) = E(e^{t(\mu + \sigma Y)}) = e^{t\mu} E(e^{t\sigma Y}) = e^{t\mu} e^{\frac{t^2 \sigma^2}{2}} = e^{t\mu + \frac{t^2 \sigma^2}{2}}$$

$$E(e^{itX}) = E(e^{it(\mu + \sigma Y)}) = e^{it\mu - \frac{\sigma^2 t^2}{2}}$$

□

Momentos respecto de la media: para una variable aleatoria Y con distribución $N(0, 1)$, la función generatriz de momentos es $\exp\left(\frac{t^2}{2}\right)$, cuyo desarrollo en serie es

$$E(e^{tY}) = \sum_r \frac{1}{r!} \frac{t^{2r}}{2^r} = \sum_n \frac{\alpha_n}{n!} t^n$$

donde $\alpha_n = E(Y^n)$.

De aquí podemos deducir los momentos de Y :

$$\alpha_{2r+1} = E(Y^{2r+1}) = 0$$

$$\frac{\alpha_{2r}}{(2r)!} = \frac{1}{r! 2^r} \implies \alpha_{2r} = E(Y^{2r}) = \frac{(2r)!}{2^r r!} = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2r-1)$$

* puede obtenerse fácilmente mediante inducción en r .

Una notación utilizada en estos casos es

$$(2r-1)!! = 1 \cdot 3 \cdot \dots \cdot (2r-1)$$

$$(2r)!! = 2 \cdot 4 \cdot \dots \cdot (2r)$$

Así, los momentos respecto de la media de una variable aleatoria $X = \sigma Y + \mu$ con distribución $N(\mu, \sigma^2)$ son

$$\mu_{2r+1} = E[(X - \mu)^{2r+1}] = E[(\sigma Y)^{2r+1}] = 0$$

$$\mu_{2r} = E[(X - \mu)^{2r}] = E[(\sigma Y)^{2r}] = \sigma^{2r} \cdot (2r-1)!!$$

Teorema 17.3. Si X_r son variables aleatorias independientes con distribución $N(\mu_r, \sigma_r^2)$, $r = 1, \dots, n$ la suma

$$S_n = X_1 + \dots + X_n$$

tiene una distribución $N(\mu, \sigma^2)$ donde

$$\mu = \sum_r \mu_r$$

$$\sigma^2 = \sum_r \sigma_r^2$$

Demostración. Por la independencia de las variables aleatorias X_r , la función característica de la suma es el producto de las funciones características de los sumandos, luego

$$E(e^{itS_n}) = \prod_{r=1}^n E(e^{itX_r}) = \prod_{r=1}^n \exp\left(it\mu_r - \frac{\sigma_r^2 t^2}{2}\right) = \exp\left(\sum_{r=1}^n \left(it\mu_r - \frac{\sigma_r^2 t^2}{2}\right)\right) = \exp\left(it \sum_{r=1}^n \mu_r - \frac{t^2}{2} \sum_{r=1}^n \sigma_r^2\right)$$

que coincide con la función característica de una distribución $N(\mu, \sigma^2)$ con los μ, σ^2 del enunciado. □

17.3. Distribución B (beta)

La función B (beta) se define por

$$B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx, \quad p, q > 0$$

Con esta función se puede definir la **distribución B (beta) de parámetros p, q** como la distribución de tipo continuo en \mathbb{R} con función de densidad

$$f(x) = \begin{cases} \frac{1}{B(p, q)} x^{p-1} (1-x)^{q-1} & x \in (0, 1) \\ 0 & x \notin (0, 1) \end{cases}$$

La **distribución transformada de la B** que se introduce en el siguiente teorema será de interés más adelante.

Teorema 17.4. *Si la variable aleatoria X tiene una distribución B de parámetros p, q la variable*

$$Y = \frac{X}{1-X}$$

tiene la función de densidad dada por

$$f_Y(y) = \begin{cases} \frac{1}{B(p, q)} \frac{y^{p-1}}{(1+y)^{p+q}} & y \in (0, \infty) \\ 0 & y \leq 0 \end{cases}$$

Demostración. La transformación que hace pasar de X a Y es una biyección monótona de $(0, 1)$ en $(0, \infty)$. Aplicaremos la fórmula que da la densidad en un cambio de variable aleatoria

$$f_Y(y) = f_X(h(y)) |h'(y)|$$

donde, en este caso

$$h(y) = \frac{y}{1+y}, \quad y \in (0, \infty)$$

es la inversa del cambio realizado, con lo que se obtiene la función de densidad de Y

$$\begin{aligned} f_Y(y) &= f_X\left(\frac{y}{1+y}\right) \frac{1}{(1+y)^2} = \frac{1}{B(p, q)} \left(\frac{y}{1+y}\right)^{p-1} \left(1 - \frac{y}{1+y}\right)^{q-1} \frac{1}{(1+y)^2} = \\ &= \frac{1}{B(p, q)} \left(\frac{y}{1+y}\right)^{p-1} \left(\frac{1}{1+y}\right)^{q-1} \frac{1}{(1+y)^2} = \frac{1}{B(p, q)} \frac{y^{p-1}}{(1+y)^{p-1+q-1+2}} = \\ &= \frac{1}{B(p, q)} \frac{y^{p-1}}{(1+y)^{p+q}} \end{aligned}$$

para $y \in (0, \infty)$, y fuera de este intervalo, vale 0. □

Los momentos de la distribución de la variable Y transformada de la B son

$$\alpha_r = E(Y^r) = \frac{1}{B(p, q)} \int_0^\infty y^r \frac{y^{p-1}}{(1+y)^{p+q}} dy = \frac{1}{B(p, q)} \int_0^\infty \frac{y^{p+r-1}}{(1+y)^{p+r+q-r}} dy = \frac{B(p+r, q-r)}{B(p, q)}$$

válido para $r < q$. Para $r \geq q$ se tiene

$$E(Y^r) = \infty$$

Simplificando la expresión anterior, para $r < q$, se tiene

$$\begin{aligned} \alpha_r &= \frac{B(p+r, q-r)}{B(p, q)} = \frac{\Gamma(p+r)\Gamma(q-r)}{\Gamma(p)\Gamma(q)} = \frac{p \cdot (p+1) \cdot \dots \cdot (p+r-1) \Gamma(p) \Gamma(q-r)}{\Gamma(p) \cdot (q-r) \cdot (q-r+1) \cdot \dots \cdot (q-1) \Gamma(q-r)} = \\ &= \frac{p(p+1) \cdot \dots \cdot (p+r-1)}{(q-r)(q-r+1) \cdot \dots \cdot (q-1)} = \frac{(p+r-1)^{(r)}}{(q-1)^{(r)}} \end{aligned}$$

17.4. Distribución Γ (gamma)

La función gamma se puede definir por la integral

$$\Gamma(p) = \int_0^\infty e^{-t} t^{p-1} dt, \quad p > 0$$

Se verifica

$$\int_0^\infty e^{-at} t^{p-1} dt = \frac{\Gamma(p)}{a^p}, \quad a, p > 0 \quad (1)$$

Una variable aleatoria tiene una **distribución Γ (gamma) de parámetros $a, p > 0$** si su función de densidad es de la forma

$$f(x) = \begin{cases} \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

La distribución con la función de densidad anterior será denotada por «distribución $\Gamma(a, p)$ ».

Que es una función de densidad es inmediato, ya que es no negativa y, usando (1) obtenemos que

$$\int_{\mathbb{R}} f(x) dx = \int_0^\infty \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} dx = \frac{a^p}{\Gamma(p)} \int_0^\infty e^{-ax} x^{p-1} dx \stackrel{(1)}{=} \frac{a^p}{\Gamma(p)} \frac{\Gamma(p)}{a^p} = 1$$

La **función característica de $\Gamma(a, p)$** es

$$\varphi(t) = E(e^{itX}) = \frac{a^p}{\Gamma(p)} \int_0^\infty e^{itx} x^{p-1} e^{-ax} dx = \frac{a^p}{\Gamma(p)} \int_0^\infty e^{-(a-it)x} x^{p-1} dx$$

Ahora bien, (1) se verifica también si a es un complejo con parte real positiva, por lo que es

$$\varphi(t) = \frac{a^p}{\Gamma(p)} \frac{\Gamma(p)}{(a-it)^p} = \frac{a^p}{(a-it)^p} = \frac{1}{\left(\frac{a-it}{a}\right)^p} = \left(1 - \frac{it}{a}\right)^{-p}$$

Teorema 17.5. Sean X, Y dos variables aleatorias independientes con distribuciones $\Gamma(a, p)$ y $\Gamma(a, q)$, respectivamente. Entonces, las variables

$$U = \frac{X}{Y}, \quad V = X + Y$$

son dos variables aleatorias independientes, U con función de densidad

$$f_U(u) = \begin{cases} \frac{1}{B(p, q)} \frac{u^{p-1}}{(1+u)^{p+q}} & u > 0 \\ 0 & u \leq 0 \end{cases}$$

(o sea, U es una transformada de una distribución B).

V con función de densidad

$$f_V(v) = \begin{cases} \frac{a^{p+q}}{\Gamma(p+q)} v^{p+q-1} e^{-av} & v > 0 \\ 0 & v \leq 0 \end{cases}$$

que se corresponde a una distribución $\Gamma(a, p+q)$.

Demostración. El cambio de la variable (X, Y) a la (U, V) se hace con la transformada definida por

$$\begin{cases} u = \frac{x}{y} \\ v = x + y \end{cases} \quad \text{con inversa} \quad \begin{cases} x = \frac{uv}{1+u} \\ y = \frac{v}{1+u} \end{cases}$$

que transforma el primer cuadrante $(0, \infty) \times (0, \infty)$ en sí mismo. El jacobiano de la transformación es

$$\frac{\partial(x, y)}{\partial(u, v)} = \frac{v}{(1+u)^2}$$

La función de densidad de (X, Y) vale para $(x, y) \in (0, \infty) \times (0, \infty)$ y es (X, Y independientes)

$$f_{(X, Y)}(x, y) = \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} \frac{a^q}{\Gamma(q)} y^{q-1} e^{-ay} = \frac{a^{p+q}}{\Gamma(p)\Gamma(q)} x^{p-1} y^{q-1} e^{-a(x+y)}$$

Y para obtener la función de densidad de (U, V) hay que sustituir en la anterior función de densidad las variables x, y por sus valores en función de u, v y multiplicar por el módulo del jacobiano de la transformación, obteniendo

$$f_{(U, V)}(u, v) = \frac{a^{p+q}}{\Gamma(p)\Gamma(q)} \left(\frac{uv}{1+u}\right)^{p-1} \left(\frac{v}{1+u}\right)^{q-1} e^{-av} \frac{v}{(1+u)^2} = \frac{a^{p+q}}{\Gamma(p)\Gamma(q)} \frac{u^{p-1}}{(1+u)^{p+q}} v^{p+q-1} e^{-av}, \quad u, v > 0$$

y vale 0 en el resto de cuadrantes.

La densidad de U se obtiene integrando el anterior respecto de v en $(0, \infty)$ y resulta

$$f_U(u) = \frac{a^{p+q}}{\Gamma(p)\Gamma(q)} \frac{u^{p-1}}{(1+u)^{p+q}} \int_0^\infty v^{p+q-1} e^{-av} dv \stackrel{(1)}{=} \frac{a^{p+q}}{\Gamma(p)\Gamma(q)} \frac{u^{p-1}}{(1+u)^{p+q}} \frac{\Gamma(p+q)}{a^{p+q}} = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{u^{p-1}}{(1+u)^{p+q}}$$

Ahora bien, como sabemos que

$$1 = \int_0^\infty f_U(u) du = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^\infty \frac{u^{p-1}}{(1+u)^{p+q}} du = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} B(p, q) \implies \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} = \frac{1}{B(p, q)}$$

que es una propiedad conocida de la relación entre las funciones B y Γ .

De aquí podemos ver que $\Gamma(p)\Gamma(q) = \Gamma(p+q)B(p, q)$, y entonces

$$\begin{aligned} f_{(U,V)}(u, v) &= \frac{a^{p+q}}{\Gamma(p)\Gamma(q)} \frac{u^{p-1}}{(1+u)^{p+q}} v^{p+q-1} e^{-av} = \frac{a^{p+q}}{B(p, q)\Gamma(p+q)} \frac{u^{p-1}}{(1+u)^{p+q}} v^{p+q-1} e^{-av} = \\ &= \left[\frac{1}{B(p, q)} \frac{u^{p-1}}{(1+u)^{p+q}} \right] \cdot \left[\frac{a^{p+q}}{\Gamma(p+q)} v^{p+q-1} e^{-av} \right] = f_U(u) \cdot f_V(v) \end{aligned}$$

donde $f_U(u)$, $f_V(v)$ son las densidades dadas en el enunciado, lo que prueba la independencia de las variables U, V . \square

Nota: Algunas obras utilizan en vez del parámetro « a » de la distribución $\Gamma(a, p)$ el parámetro $\alpha = \frac{1}{a}$.

Teorema 17.6. Si X_j son n variables aleatorias independientes con distribuciones $\Gamma(a, p_j)$ para $j = 1, \dots, n$, entonces la suma

$$S_n = X_1 + \dots + X_n$$

tiene una distribución $\Gamma(a, p)$ donde $p = p_1 + \dots + p_n$

Demostración. Procedemos por inducción.

El caso 2 queda demostrado en el teorema anterior, ya que $S_2 = V$, que es $\Gamma(a, p+q)$.

Supongámoslo cierto para $n-1$, y entonces tenemos que

$$S_n = S_{n-1} + X_n$$

y la hipótesis de inducción nos asegura que S_{n-1} es $\Gamma(a, \sum_{j=1}^{n-1} p_j)$, y aplicando el caso $n=2$ a S_{n-1} y X_n , tenemos el resultado. \square

Teorema 17.7. Los momentos de la distribución $\Gamma(a, p)$ son

$$\alpha_r = \frac{p(p+1) \cdot \dots \cdot (p+r-1)}{a^r} = \frac{(p+r-1)^{(r)}}{a^r}$$

En particular, la media y la varianza son

$$\mu = \frac{p}{a}$$

$$\sigma^2 = \frac{p}{a^2}$$

Demostración. Vamos a calcularlo:

$$\alpha_r = \frac{a^p}{\Gamma(p)} \int_0^\infty x^{r+p-1} e^{-ax} = \frac{a^p}{\Gamma(p)} \frac{\Gamma(r+p)}{a^{r+p}} = \frac{p(p+1) \cdot \dots \cdot (p+r-1) \cdot \Gamma(p)}{\Gamma(p) a^r} = \frac{(p+r-1)^{(r)}}{a^r}$$

\square

17.5. Distribución χ^2 (ji cuadrado)

Se llama **distribución χ^2 con n grados de libertad** a la distribución de una suma de n cuadrados de variables aleatorias independientes normales de media 0 y varianza 1, es decir, la distribución χ^2 es la de la variable aleatoria

$$X = X_1^2 + \dots + X_n^2$$

donde las X_j son independientes y tienen distribución $N(0, 1)$.

Teorema 17.8. *La distribución χ^2 con n grados de libertad es una distribución $\Gamma(\frac{1}{2}, \frac{n}{2})$, es decir, su función de densidad es*

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Demostración. Comencemos por el caso $n = 1$.

Si X_1 es $N(0, 1)$, su función de densidad es

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \quad y \in \mathbb{R}$$

y de la $X = X_1^2$ se obtiene por

$$f(x) = g(\sqrt{x}) \frac{1}{2\sqrt{x}} + g(-\sqrt{x}) \frac{1}{2\sqrt{x}} \stackrel{g \text{ par}}{=} g(\sqrt{x}) \frac{1}{\sqrt{x}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} x^{-\frac{1}{2}}, \quad x > 0$$

que es la función de densidad de una distribución $\Gamma(\frac{1}{2}, \frac{1}{2})$, y el caso $n = 1$ es cierto.

Demostrado esto, la variable X es una suma de n variables aleatorias independientes con distribución $\Gamma(\frac{1}{2}, \frac{1}{2})$, por lo que (penúltimo teorema de la sección anterior) X tiene distribución $\Gamma(\frac{1}{2}, \sum_{i=1}^n \frac{1}{2}) = \Gamma(\frac{1}{2}, \frac{n}{2})$, como queríamos ver. \square

Como la distribución χ^2 con n grados de libertad es una $\Gamma(\frac{1}{2}, \frac{n}{2})$, su función característica es

$$\varphi(t) = (1 - 2it)^{-\frac{n}{2}}$$

Por la misma razón, los momentos son

$$\alpha_r = 2^r \left(\frac{n}{2} + r - 1 \right)^{(r)} = n(n+2)(n+4) \cdot \dots \cdot (n+2r-2)$$

y la media y varianza

$$\alpha_1 = n$$

$$\sigma^2 = 2n$$

17.6. Distribución \mathcal{F} de Snédecor

Se llama **distribución \mathcal{F} de Snédecor con m y n grados de libertad** a la distribución de una variable aleatoria que se define por la expresión

$$\mathcal{F} = \frac{X}{Y}, \quad m, n \in \mathbb{Z}^+$$

donde X, Y son dos variables aleatorias independientes con distribuciones χ^2 de m y n grados de libertad, respectivamente.

Teorema 17.9. *La función de densidad de una variable aleatoria con una distribución \mathcal{F} de Snédecor con m y n grados de libertad viene dada por*

$$f(x) = \begin{cases} \frac{m^{\frac{m}{2}} n^{\frac{n}{2}}}{B(\frac{m}{2}, \frac{n}{2})} \frac{x^{\frac{m}{2}-1}}{(n+mx)^{\frac{m+n}{2}}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Demostración. La función de densidad de $Z = \frac{X}{Y}$, según el primer teorema de la sección dedicada a la distribución Γ es (recordemos que χ^2 con k grados de libertad tiene distribución $\Gamma(\frac{1}{2}, \frac{k}{2})$):

$$g(z) = \frac{1}{B(\frac{m}{2}, \frac{n}{2})} \frac{z^{\frac{m}{2}-1}}{(1+z)^{\frac{m+n}{2}}}, \quad z > 0$$

y la variable

$$\mathcal{F} = \frac{n}{m} Z$$

tendrá función de densidad

$$\begin{aligned} f(x) &= g\left(\frac{m}{n}x\right) \frac{m}{n} = \frac{m}{n} \frac{1}{B(\frac{m}{2}, \frac{n}{2})} \frac{\left(\frac{m}{n}x\right)^{\frac{m}{2}-1}}{\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} = \frac{m}{n} \frac{1}{B(\frac{m}{2}, \frac{n}{2})} \frac{m^{\frac{m}{2}-1} \cdot n^{1-\frac{m}{2}} \cdot x^{\frac{m}{2}-1}}{(n+mx)^{\frac{m+n}{2}} \cdot n^{-\frac{m+n}{2}}} = \\ &= \frac{1}{B(\frac{m}{2}, \frac{n}{2})} \frac{m^{\frac{m}{2}} \cdot n^{-\frac{m}{2}+\frac{m+n}{2}} \cdot x^{\frac{m}{2}-1}}{(n+mx)^{\frac{m+n}{2}}} = \frac{m^{\frac{m}{2}} n^{\frac{n}{2}}}{B(\frac{m}{2}, \frac{n}{2})} \frac{x^{\frac{m}{2}-1}}{(n+mx)^{\frac{m+n}{2}}} \end{aligned}$$

y tenemos el resultado. □

Teorema 17.10. *Los momentos α_r de la distribución \mathcal{F} de Snédecor con m y n grados de libertad existen para $r < \frac{n}{2}$ y vienen dados por*

$$\alpha_r = \left(\frac{n}{m}\right)^r \frac{\left(\frac{m}{2} + r - 1\right)^{(r)}}{\left(\frac{n}{2} - 1\right)^{(r)}}, \quad r < \frac{n}{2}$$

Demostración. Primero notamos que

$$E(\mathcal{F}^r) = \frac{n^r}{m^r} E(Z^r)$$

donde $Z = \frac{X}{Y}$ tiene una distribución transformada de B con $p = \frac{m}{2}$ y $q = \frac{n}{2}$. Podemos, entonces, aplicar la fórmula que vimos para los momentos de esta distribución:

$$\alpha_r(B) = \frac{(p+r-1)^{(r)}}{(q-1)^{(r)}}$$

que existe para $r < q = \frac{n}{2}$. Por tanto, es

$$\alpha_r(\mathcal{F}) = \left(\frac{n}{m}\right)^r \frac{\left(\frac{m}{2} + r - 1\right)^{(r)}}{\left(\frac{n}{2} - 1\right)^{(r)}}$$

como queríamos. □

Podemos ahora particularizar y obtener los dos primeros momentos:

$$\alpha_1 = \frac{n}{m} \frac{\frac{m}{2}}{\frac{n}{2} - 1} = \frac{n}{m} \frac{m}{n - 2} = \frac{n}{n - 2}$$

$$\alpha_2 = \left(\frac{n}{m}\right)^2 \frac{\frac{m}{2} \left(\frac{m}{2} + 1\right)}{\left(\frac{n}{2} - 1\right) \left(\frac{n}{2} - 2\right)} = \left(\frac{n}{m}\right)^2 \frac{m(m+2)}{(n-2)(n-4)} = \frac{n^2(n+2)}{m(n-2)(n-4)}$$

Y la varianza

$$\sigma^2 = \alpha_2 - \alpha_1^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$$

Sean ahora

$$X_1, \dots, X_m, X_{m+1}, \dots, X_{m+n}$$

$m+n$ variables aleatorias independientes con distribución $N(0, \sigma^2)$. Las variables X, Y definidas por

$$X = X_1^2 + \dots + X_m^2$$

$$Y = X_{m+1}^2 + \dots + X_{m+n}^2$$

son independientes y las variables

$$\frac{X}{\sigma^2}, \quad \frac{Y}{\sigma^2}$$

tienen distribución χ^2 con m y n grados de libertad, respectivamente. Por ello se dice que X, Y tienen distribuciones $\sigma^2 \chi^2$ con m y n grados de libertad, respectivamente.

Es inmediato que

$$\mathcal{F} = \frac{\frac{X}{\sigma^2}}{\frac{Y}{\sigma^2}}$$

tiene distribución \mathcal{F} de Snédecor con m y n grados de libertad.

17.7. Distribución t de Student

Se llama **distribución t de Student con n grado de libertad** a la distribución de la variable aleatoria

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

donde X, Y son variables aleatorias independientes, la variable X con distribución $N(0, 1)$ y la variable Y con distribución χ^2 con n grados de libertad.

Teorema 17.11. *La función de densidad de la distribución t de Student con n grados de libertad es*

$$f(x) = \frac{n^{\frac{n}{2}}}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \frac{1}{(n+x^2)^{\frac{n+1}{2}}}, \quad x \in \mathbb{R}$$

Demostración. La variable T tiene una distribución simétrica, ya que

$$-T = \frac{-X}{\sqrt{\frac{Y}{n}}} = \frac{X'}{\sqrt{\frac{Y}{n}}}$$

con $X' = -X$, que también es $N(0, 1)$, por lo que $-T$ y T tienen la misma distribución.

Si es f la función de densidad de T , la de T^2 es

$$g(x) = f(\sqrt{x}) \frac{1}{2\sqrt{x}} + f(-\sqrt{x}) \frac{1}{2\sqrt{x}} = \frac{1}{\sqrt{x}} f(\sqrt{x}), \quad x > 0$$

y T^2 es una \mathcal{F} de Snédecor con 1 y n grados de libertad, luego

$$g(x) = \frac{n^{\frac{n}{2}}}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \frac{x^{\frac{1}{2}-1}}{(n+x)^{\frac{n+1}{2}}}, \quad x \in \mathbb{R}$$

y entonces

$$f(x) = xg(x^2) = x \frac{n^{\frac{n}{2}}}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \frac{x^{-1}}{(n+x^2)^{\frac{n+1}{2}}} = \frac{n^{\frac{n}{2}}}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \frac{1}{(n+x^2)^{\frac{n+1}{2}}}$$

tal como queríamos ver. □

Para esta distribución existen solo los momentos de orden $r < n$. Si son de orden impar, son nulos, por la simetría de la distribución y el momento de orden par $r = 2h$ de T es el momento de orden h de la distribución \mathcal{F} de 1 y n grados de libertad:

$$\alpha_r = \alpha_{2h} = E(T^{2h}) = E(\mathcal{F}^h) = n^h \frac{(h - \frac{1}{2})^{(h)}}{(\frac{n}{2} - 1)^{(h)}} = \frac{1 \cdot 3 \cdot 5 \cdot \dots \cdot (2h - 1) \cdot n^h}{(n - 2)(n - 4) \cdot \dots \cdot (n - 2h)}$$