

Probabilidad I

Miguel González
mgonzalez.contacto@gmail.com
miguelgg.com

Mayo de 2019

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} Z \sim N(0, 1)$$

Revisado en 2022

Apuntes de la asignatura impartida por Javier Cárcamo
en la Universidad Autónoma de Madrid en Mayo de 2019.

Acerca de este documento

Estos apuntes son una versión revisada de los de la asignatura Probabilidad I del grado en matemáticas, tomados en Mayo de 2019 por Miguel González. La asignatura fue impartida por Javier Cárcamo. A los apuntes originales se les ha añadido esta página, una imagen de portada, y breves párrafos explicativos en las zonas menos completas. Asimismo se han revisado las erratas y completado los contenidos faltantes.

Este documento es:

- Una recopilación ordenada y directa de las definiciones y resultados más importantes del tema en cuestión, al nivel de los estudios de grado.
- Una colección de demostraciones completas de dichos resultados (salvo en los casos más básicos).
- Una *guía* para revisar de manera rápida las ideas que se han adquirido previamente, o para consultar enunciados puntuales que puedan no haberse comprendido en su totalidad.

Este documento NO es:

- Un libro de texto de la asignatura.
- Una colección de ejercicios para practicar los conceptos adquiridos.
- Un listado de ejemplos para ilustrar las ideas tratadas. A pesar de ello, en ocasiones se incluyen ejemplos puntuales que puedan ser de especial interés o curiosidad, pero se intentan reducir al mínimo en virtud del primer punto de la lista anterior.

Sobre Probabilidad I

Esta asignatura se centra en la teoría de la probabilidad desde el punto de vista más clásico, sin recurrir a la formulación moderna mediante teoría de la medida. El principal objetivo es introducir el concepto de variable aleatoria, continua y discreta, y establecer algunos de los resultados más importantes en relación a las distribuciones de probabilidad, como el teorema central del límite.

Requisitos previos

1. Familiaridad con la notación matemática básica.
2. Conocimientos de cálculo equivalentes a la asignatura de Cálculo I. (Sucesiones, funciones reales de una variable, límites, derivadas, integración).

Índice

1. Espacios de Probabilidad	3
2. Variables aleatorias	8
2.1. Distribuciones discretas útiles	10
2.2. Distribuciones continuas útiles	11
3. Vectores aleatorios	13
3.1. Algunas distribuciones utiles	16
4. Esperanza y momentos	18
5. Funciones Características	24
6. Teoremas límite	26

1. Espacios de Probabilidad

La teoría de la probabilidad estudia el comportamiento de **experimentos aleatorios**, de los que conocemos los posibles resultados, pero no cuál de ellos se dará. Permite medir la incertidumbre asociada al resultado de un experimento.

El espacio de resultados posibles, o espacio muestral, se denotará por Ω . Un suceso es un conjunto $A \subset \Omega$. El significado de un suceso es que este se da si el resultado del experimento pertenece al suceso. Por ejemplo, al tirar un dado, $\Omega = \{1, 2, 3, 4, 5, 6\}$, y sacar un número par, es el suceso $A = \{2, 4, 6\}$.

Definición 1 (Terminología acerca de sucesos). Dado un suceso A , si $|A| = 1$, se denomina **suceso elemental**. Si no, es un **suceso compuesto**. Ω también se denomina **suceso seguro**, y \emptyset **suceso imposible**.

El **suceso contrario** a A es $A^c = \Omega \setminus A$. Representa que no se da A .

El **suceso unión o intersección** de A y B es el conjunto $A \cup B$ o $A \cap B$. Representa que se da alguno, o que se dan ambos, respectivamente.

La **inclusión de sucesos**, $A \subset B$, representa que si se tiene A , también se tiene B .

La **diferencia de sucesos**, $A \setminus B = A \cap B^c$, representa que se tiene A pero no B .

Dos sucesos son **disjuntos o incompatibles** si $A \cap B = \emptyset$. Su unión se suele representar $A \uplus B$ o simplemente $A + B$.

Por supuesto, se tienen las propiedades usuales de operaciones con conjuntos, dado que, al fin y al cabo, es lo que son los sucesos.

De manera clásica, la probabilidad de un suceso se calcula dividiendo los casos favorables al suceso entre todos los casos posibles. Esto presenta inconvenientes: los casos posibles pueden ser infinitos, los casos pueden no aparecer con la misma frecuencia... Otra forma de obtener probabilidades es la frecuentista: realizar multitud de veces el experimento y determinar cuántas veces se obtiene cada suceso. Esto no siempre es posible, por lo que es mejor y más general plantear la probabilidad de manera axiomática:

Definición 2. Un **espacio de probabilidad** es una triplete (Ω, \mathcal{F}, P) , donde:

- Ω es un conjunto no vacío llamado **espacio muestral**, el conjunto de todos los resultados posibles del experimento aleatorio.
- $\mathcal{F} \subset \mathcal{P}(\Omega)$ es el **espacio de sucesos aleatorios**, donde cada conjunto es un **suceso** cuya probabilidad puede medirse. \mathcal{F} debe ser una σ -álgebra sobre Ω , es decir:

1. $\Omega \in \mathcal{F}$

2. $A \in \mathcal{F} \implies A^c = \Omega \setminus A \in \mathcal{F}$

3. $\{A_i\}_1^\infty \subset \mathcal{F} \implies \bigcup_{i=1}^\infty A_i \in \mathcal{F}$

- $P : \mathcal{F} \rightarrow [0, 1]$ es una **medida de probabilidad** sobre \mathcal{F} , tal que verifica los **axiomas de Kolmogorov**:

1. $P(\Omega) = 1$

2. Si $\{A_i\}_1^\infty \subset \mathcal{F}$ son sucesos disjuntos dos a dos, entonces $P\left(\biguplus_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty P(A_i)$.

La razón de distinguir \mathcal{F} de $\mathcal{P}(\Omega)$ es que en ocasiones, no existe una medida de probabilidad sobre todas las partes de un conjunto, teniendo que restringir el espacio de sucesos permitidos a una colección menor.

Proposición 1 (Algunas propiedades sobre σ -álgebras). *Dada la σ -álgebra \mathcal{F} , se tiene:*

1. $\emptyset \in \mathcal{F}$
2. $\{A_i\}_1^\infty \subset \mathcal{F} \implies \bigcap_{i=1}^\infty A_i \in \mathcal{F}$
3. $\{A_i\}_1^n \subset \mathcal{F} \implies \bigcup_{i=1}^n A_i \in \mathcal{F} \wedge \bigcap_{i=1}^n A_i \in \mathcal{F}$
4. Si $\{A_i\}_1^n$ cumple $A_1 \subset A_2 \subset \dots \subset A_i \subset \dots$ y además $\bigcup_{i=1}^\infty A_i = A$ (lo que se suele denotar $A_i \uparrow A$), o bien, $A_1 \supset A_2 \supset \dots \supset A_i \supset \dots$ y además $\bigcap_{i=1}^\infty A_i = A$ (lo que se suele denotar $A_i \downarrow A$), entonces $A \in \mathcal{F}$.

Demostración.

1. Como $\Omega \in \mathcal{F}$, $\Omega^c = \emptyset \in \mathcal{F}$.
2. Como $\bigcap_{i=1}^\infty A_i = (\bigcup_{i=1}^\infty A_i^c)^c$, y es cerrado a complemento y unión, se tiene.
3. Basta con observar que $\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n A_i \cup \bigcup_{i=n+1}^\infty \emptyset$, y como $\emptyset \in \mathcal{F}$, esta unión también. Para la intersección análogo, completando con Ω .
4. Es un caso particular de ser cerrado a unión e intersección. □

Por el segundo axioma de Kolmogorov, parece razonable obtener formas de calcular probabilidades de distintos sucesos **partiéndolos en partes disjuntas**, lo que permite derivar una serie de propiedades.

A continuación, veremos estas **propiedades de un espacio de probabilidad** que resultarán cruciales a la hora de calcular probabilidades de sucesos:

Proposición 2. En un espacio de probabilidad, (Ω, \mathcal{F}, P) , se tiene:

1. $P(\emptyset) = 0$ (**Probabilidad del vacío**)
2. $P(\bigsqcup_1^n A_i) = \sum_1^n P(A_i)$ (**Aditividad finita**)
3. $P(B \setminus A) = P(B) - P(A \cap B)$ (**Probabilidad de la resta**)
4. Si $A \subset B$, entonces $P(B \setminus A) = P(B) - P(A)$, y $P(A) \leq P(B)$ (**Monotonía**)
5. $P(A^c) = 1 - P(A)$ (**Probabilidad del complementario**)
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, y, en general:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} (A_i \cap A_j \cap A_k) - \dots + (-1)^{r+1} \sum_{i_1 < \dots < i_r} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_r}) + \dots$$

(**Principio de inclusión-exclusión**)

7. Si $\{A_i\}_1^n \subset \mathcal{F}$, entonces $P(\bigcup_1^n A_i) \leq \sum_1^n P(A_i)$ (**Subaditividad finita**)
8. Si $\{A_i\}_1^\infty \subset \mathcal{F}$ y $A_n \uparrow A$, entonces $P(A_n) \uparrow P(A)$. Si, por el contrario, $A_n \downarrow A$, entonces $P(A_n) \downarrow P(A)$. (**Continuidad hacia arriba y hacia abajo**)
9. Si $\{A_i\}_1^\infty \subset \mathcal{F}$, entonces $P(\bigcup_1^\infty A_i) \leq \sum_1^\infty P(A_i)$ (**Subaditividad numerable**)

Demostración. Para 1, veamos que, como $\emptyset = \bigcup_1^\infty \emptyset$, y son disjuntos, se tiene $P(\emptyset) = P(\bigcup_1^\infty \emptyset) = \sum_1^\infty P(\emptyset)$. Ahora bien, como $P(\emptyset) \in [0, 1]$, debe ser que $P(\emptyset) = 0$.

Para 2, basta con considerar $P(\biguplus_1^n A_i) = P(\biguplus_1^n A_i \cup \biguplus_{i+1}^\infty \emptyset) = \sum_1^n P(A_i) + \sum_{i+1}^\infty 0 = \sum_1^n P(A_i)$.

Para 3, basta con escribir $B = B \cap (A \cup A^c) = (A \cap B) \cup (A^c \cap B) = (A \cap B) \cup (B \setminus A)$, claramente disjuntos, luego $P(B) = P(A \cap B) + P(B \setminus A)$ y se tiene lo que se quería.

Para 4, veamos que es un caso particular de 3 donde $A \cap B = A$, y como las probabilidades están en $[0, 1]$, se tiene $P(B) - P(A) \geq 0$ de donde se tiene la desigualdad.

Para 5, basta con poner $\omega = A \uplus A^c$, luego $1 = P(A) + P(A^c)$.

Para 6, el primer caso se puede probar viendo que $A \cup B = A \uplus (B \setminus A)$, luego $P(A \cup B) = P(A) + P(B \setminus A) = P(A) + P(B) - P(A \cap B)$. Para el caso general se procede por inducción:

$$\begin{aligned} P\left(\bigcup_1^{n+1} A_i\right) &= P\left(\bigcup_1^n A_i\right) + P(A_{n+1}) - P\left(\bigcup_1^n A_i \cap A_{n+1}\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) - \dots \\ &\dots + (-1)^{r+1} \sum_{1 \leq i_1 < \dots < i_r \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_r}) + \dots + P(A_{n+1}) - P\left(\bigcup_1^n (A_i \cap A_{n+1})\right) \end{aligned}$$

Ahora podemos aplicar la hipótesis a este último conjunto, que es unión de n conjuntos, para observar que:

$$\begin{aligned} -P\left(\bigcup_1^n (A_i \cap A_{n+1})\right) &= -\left(\sum_{i=1}^n P(A_i \cap A_{n+1}) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j \cap A_{n+1}) - \dots \right. \\ &\dots + (-1)^{r+1} \sum_{1 \leq i_1 < \dots < i_r \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_r} \cap A_{n+1}) + \dots) = \\ &= -\sum_{i=1}^n P(A_i \cap A_{n+1}) + \sum_{i < j} P(A_i \cap A_j \cap A_{n+1}) - \dots \\ &\dots + (-1)^r \sum_{i_1 < \dots < i_r} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_r} \cap A_{n+1}) + \dots \end{aligned}$$

Donde en el último paso utilizamos la distributividad de la intersección respecto de la unión. Como se observa, aparecen todas las combinaciones preexistentes con el nuevo conjunto A_{n+1} , y el signo esperado. Combinando ambas expresiones, queda que:

$$\begin{aligned} P\left(\bigcup_1^{n+1} A_i\right) &= \sum_{i=1}^{n+1} P(A_i) - \sum_{1 \leq i < j \leq n+1} P(A_i \cap A_j) - \dots \\ &\dots + (-1)^{r+1} \sum_{1 \leq i_1 < \dots < i_r \leq n+1} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_r}) + \dots \end{aligned}$$

Como se quería.

Para 7, el caso base para $n = 2$ se tiene dado que $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2)$ al ser las probabilidades no negativas. En general, procedemos por inducción: $P(\biguplus_1^{n+1} A_i) = P(\biguplus_1^n A_i) + P(A_{n+1}) - P(\biguplus_1^n A_i \cap A_{n+1}) \leq P(\biguplus_1^n A_i) + P(A_{n+1}) \leq \sum_1^n P(A_i) + P(A_{n+1}) = \sum_1^{n+1} P(A_i)$.

Para el 8, utilizando la monotonía, sabemos que $P(A_n)$ crece, y tiene límite dado que está acotada por 1 en todo caso. Ahora, si expresamos $A = \biguplus_1^\infty (A_i \setminus A_{i-1})$, que es unión de disjuntos, aceptando que $A_0 = \emptyset$, tenemos que $P(A) = \lim_{n \rightarrow \infty} \sum_1^n P(A_i - A_{i-1}) = \lim_{n \rightarrow \infty} \sum_1^n P(A_i) - P(A_{i-1}) = \lim_{n \rightarrow \infty} (P(A_n) - P(\emptyset)) =$

$\lim_{n \rightarrow \infty} P(A_n)$. En caso de que decrezca la sucesión, basta con ver que $A = \bigcap_1^\infty A_i = (\bigcup_1^\infty A_i^c)^c$, luego $A_n^c \uparrow A$, dado que si $A_i \subset A_{i+1}$, se tiene $A_i^c \supset A_{i+1}^c$. Por tanto, aplicamos lo que acabamos de ver para obtener que $\lim_{n \rightarrow \infty} P(A_n^c) = P(A^c)$, luego $\lim_{n \rightarrow \infty} (1 - P(A_n)) = 1 - P(A)$ con lo que $\lim_{n \rightarrow \infty} P(A_n) = P(A)$.

Para el 9, como por definición $(\bigcup_1^n A_i) \uparrow (\bigcup_1^\infty A_i)$, entonces $P(\bigcup_1^n A_i) \uparrow P(\bigcup_1^\infty A_i)$, es decir, $P(\bigcup_1^\infty A_i) = \lim_{n \rightarrow \infty} P(\bigcup_1^n A_i) \leq \lim_{n \rightarrow \infty} \sum_1^n P(A_i) = \sum_1^\infty P(A_i)$, que existe (aunque sea ∞) porque crece la sucesión de sumas parciales. \square

Observación 1. Algunos ejemplos de medidas de probabilidad son:

- **Modelo clásico o de Laplace.** En un espacio muestral finito $\omega = \{w_1, \dots, w_n\}$, la medida $P(A) = \frac{|A|}{n}$ para $A \subseteq \omega$ es una medida sobre $(\omega, \mathcal{P}(\Omega))$.
- **Modelo finito.** En un espacio muestral finito $\omega = \{w_1, \dots, w_n\}$, y $\{p_1, \dots, p_n\}$ positivos tales que $\sum_1^n p_i = 1$, la medida $P(A) = \sum_{w_i \in A} P(\{w_i\})$ para $A \subseteq \omega$, con $P(\{w_i\}) = p_i$, es una medida sobre $(\omega, \mathcal{P}(\Omega))$.
- **Modelo discreto.** En un espacio muestral numerable $\omega = \{w_i\}_1^\infty$, y $P(\{w_i\}) \in [0, 1]$ tales que $\sum_1^\infty P(\{w_i\}) = 1$, con $P(A) = \sum_{w_i \in A} P(\{w_i\})$ es una medida sobre $(\omega, \mathcal{P}(\Omega))$.

En ocasiones, saber que se ha dado un suceso altera la probabilidad de otros sucesos. Para ello, se puede definir una nueva medida de probabilidad, reajustando la probabilidad de cada suceso habiéndose dado el anterior y normalizando para que el suceso conocido tenga probabilidad 1:

Definición 3. En un espacio de probabilidad (Ω, \mathcal{F}, P) , con $A, B \in \mathcal{F}$ y $P(B) > 0$, la **probabilidad de A condicionada a B** se define:

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

Esto da lugar a una nueva medida, $P(\cdot|B) : \mathcal{F} \rightarrow [0, 1]$, sobre (Ω, \mathcal{F}) , y por tanto verifica las propiedades de la probabilidad.

A continuación veremos resultados básicos que permiten calcular probabilidades en numerosos casos:

Proposición 3 (Fórmula del producto). *Si (Ω, \mathcal{F}, P) es un espacio de probabilidad, y $\{A_i\}_1^n \subset \mathcal{F}$ son sucesos tales que $P(\bigcap_1^n A_i) > 0$, entonces, se tiene:*

$$P\left(\bigcap_1^n A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\bigcap_1^{n-1} A_i) = \prod_{i=1}^n P(A_i|\bigcap_{j<i} A_j)$$

Demostración. Por inducción sobre n , veamos que $P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)}$ por definición, y dado que $P(A_1) \neq 0$ por monotonía. Podemos despejar y obtener la fórmula para $n = 2$. Ahora, suponiendo la fórmula válida hasta conjuntos de n elementos, tendríamos $P(A_{n+1}|\bigcap_1^n A_i) = \frac{P(\bigcap_1^n A_i \cap A_{n+1})}{P(\bigcap_1^n A_i)}$. Despejando, $P(\bigcap_1^{n+1} A_i) = P(A_{n+1}|\bigcap_1^n A_i)P(\bigcap_1^n A_i) = P(A_{n+1}|\bigcap_1^n A_i)P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\bigcap_1^{n-1} A_i)$ por la hipótesis. \square

Definición 4. Sea (Ω, \mathcal{F}, P) un espacio de probabilidad. Una colección contable (finita o numerable) $\{A_i\}_{i \in I} \subset \mathcal{F}$ es un **sistema completo de sucesos o partición de Ω** si $P(A_i) > 0 \forall i \in I$ y además $\Omega = \bigsqcup_{i \in I} A_i$.

Proposición 4 (Probabilidad total). *En un espacio de probabilidad, con $\{A_i\}_{i \in I}$ una partición de Ω , y $B \in \mathcal{F}$, se tiene:*

$$P(B) = \sum_{i \in I} P(A_i)P(B|A_i)$$

Demostración. $P(B) = P(B \cap (\bigsqcup_{i \in I} A_i)) = P(\bigsqcup_{i \in I} (B \cap A_i)) = \sum_{i \in I} P(B \cap A_i) = \sum_{i \in I} P(A_i)P(B|A_i)$. \square

Esto nos permite calcular probabilidades particionando el espacio de sucesos en distintos casos, sencillos de calcular, y sumándolos. A continuación, combinando lo anterior, veremos una manera efectiva de revertir una probabilidad condicionada:

Proposición 5 (Bayes). *En un espacio de probabilidad, con $\{A_i\}_{i \in I}$ partición de Ω , y $B \in \mathcal{F}$ tal que $P(B) > 0$, se tiene:*

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i \in I} P(A_i)P(B|A_i)}$$

con $j \in I$. La probabilidad $P(A_j|B)$ se denomina **probabilidad a posteriori** y las $P(B|A_j)$ **probabilidades a priori**.

Demostración. Por definición, $P(A_j|B) = \frac{P(A_j \cap B)}{P(B)}$. Ahora solo hay que usar la fórmula de probabilidad total y la del producto para la intersección, y se obtiene lo que se quería. \square

En general, $\{A, A^c\}$ es una partición, luego con ella se pueden revertir fácilmente probabilidades condicionadas.

Definición 5. En un espacio de probabilidad, los sucesos $A, B \in \mathcal{F}$ son **independientes** si y solo si $P(A \cap B) = P(A)P(B)$. Se suele denotar $A \cap B \equiv AB$ en este caso.

Proposición 6. \blacksquare *Si $P(A) = 0$ o $P(A) = 1$, entonces A y B son independientes para $B \in \mathcal{F}$.*

- \blacksquare *Si A, B son independientes y $P(B) > 0$, entonces $P(A|B) = P(A)$.*
- \blacksquare *Si A, B son independientes, lo son A^c y B . (Y, por tanto, A^c y B^c , y A y B^c).*

Demostración. Para la primera, si $P(A) = 0$, entonces $P(A \cap B) = 0$ por monotonía, y $P(A)P(B) = 0 \cdot P(B) = 0 = P(A \cap B)$. Si $P(A) = 1$, por monotonía, $P(A \cup B) = 1$ y por tanto $P(A \cap B) = P(A) + P(B) - P(A \cup B) = P(B) + 1 - 1 = P(B) = P(A)P(B)$. Para la segunda, por definición $P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$. Finalmente, para la tercera, tenemos $P(A^c)P(B) = (1 - P(A))P(B) = P(B) - P(A)P(B) = P(B) - P(A \cap B) = P(B - A) = P(B \cap A^c)$. \square

Definición 6. En general, los sucesos $\{A_i\}_{i \in I} \subset \mathcal{F}$ son **mutuamente independientes** si, fijado $J \subset I$ con $|J| < \infty$, siempre se tiene:

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j)$$

Es decir, si para **todo** subconjunto finito en la familia de sucesos, de tamaño superior a 1, se puede obtener la probabilidad de su intersección multiplicando individualmente las probabilidades de cada suceso.

Por ejemplo, en el caso $n = 3$, no basta con que sean independientes 2 a 2, sino que además han de verificar que $P(A \cap B \cap C) = P(A)P(B)P(C)$. En particular, si son n conjuntos, deberán verificar una condición por cada subconjunto de tamaño 2 o más de la familia de sucesos, es decir: $2^n - n - 1$ condiciones.

Proposición 7. *Dados $\{A_i\}_{i \in I}$ independientes, se tiene que $\{A_i^c\}_{i \in I}$ también son independientes.*

Demostración. Tomemos $\{A_1, \dots, A_n\} \subset \{A_i\}_{i \in I}$ cualesquiera sucesos de la familia. Veamos que $P(\bigcap_1^n A_i^c) = P((\bigcup_1^n A_i)^c) = 1 - P(\bigcup_1^n A_i) = 1 - \sum_1^n P(A_i) + \sum_{i < j} P(A_i \cap A_j) - \dots + (-1)^n P(\bigcap_1^n A_i) = 1 - \sum_1^n P(A_i) + \sum_{i < j} P(A_i)P(A_j) + \dots + (-1)^n \prod_1^n P(A_i) = \prod_1^n (1 - P(A_i)) = \prod_1^n P(A_i^c)$, donde usamos primero la Ley de De Morgan, después el principio de inclusión-exclusión y finalmente factorizamos la expresión resultante. \square

2. Variables aleatorias

En ocasiones nos interesa cierto aspecto concreto de un experimento y no el resultado obtenido. Para ello, se definen las variables aleatorias:

Definición 7. Una **variable aleatoria** es una aplicación $X : \Omega \rightarrow E$, que asocia a cada resultado de un experimento un elemento del conjunto E .

En adelante nos centraremos en variables cuantitativas, es decir, en las que $E = \mathbb{R}$. Estas deben cumplir, además, que $\forall a \in \mathbb{R}$, el conjunto $\{\omega \in \Omega : X(\omega) \leq a\} \subset \mathcal{F}$, aunque si $\mathcal{F} = \mathcal{P}(\Omega)$, esto se tiene siempre.

Las variables inducen una medida de probabilidad en \mathbb{R} :

Proposición 8. *La medida $P_X : \mathcal{B} \subset \mathcal{P}(\mathbb{R}) \rightarrow [0, 1]$ con $P_X(B) = P(\{\omega \in \Omega : X(\omega) \in B\}) \equiv P(X \in B)$ es de probabilidad sobre \mathbb{R} . Es la **distribución** de la variable X , e indica la probabilidad de que el valor de dicha variable esté en el subconjunto B de los reales.*

Demostración. $P_X(\mathbb{R}) = P(X \in \mathbb{R}) = P(\Omega) = 1$. Asimismo, $P(\biguplus B_i) = P(X \in \biguplus B_i) = P(\{\omega \in \Omega : X(\omega) \in \biguplus B_i\}) = P(\biguplus \{\omega \in \Omega : X(\omega) \in B_i\}) = \sum P_X(B_i)$. \square

Definición 8. Dada una variable aleatoria, la **función de distribución** de dicha variable es $F_X : \mathbb{R} \rightarrow \mathbb{R}$ con $F_X(x) = P(X \leq x) \equiv P_X((-\infty, x])$. Es decir, indica cuál es la probabilidad de que la variable aleatoria tome un valor a la izquierda de x .

Proposición 9. *Se tienen las siguientes propiedades en una función de distribución F :*

1. $x \leq y \implies F(x) \leq F(y)$ (F es no decreciente)
2. $x_n \downarrow x \implies F(x_n) \downarrow F(x)$ (Continuidad por la derecha)
3. $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$.

De hecho, toda función que cumpla estas propiedades verifica que existe una variable aleatoria que la genera, y es única.

Demostración. Para 1, si $x \leq y$ entonces $(-\infty, x] \subset (-\infty, y]$, luego $P_X((-\infty, x]) \leq P_X((-\infty, y])$, como se quería. Para 2, como en ese caso $(-\infty, x_n] \downarrow (-\infty, x]$, entonces $P_X((-\infty, x_n]) \downarrow P_X((-\infty, x])$, como se quería. Cabe observar que este argumento no funciona por la izquierda dado que $(-\infty, x_n] \uparrow (-\infty, x)$ si $x_n \uparrow x$, con lo cual lo único que se podría decir es que $F(x_n) \uparrow F(x^-)$, es decir, el límite por la izquierda. Finalmente, para 3, tenemos que $\lim_{x \rightarrow x \rightarrow -\infty} P_X((-\inf, X]) = P_X(\emptyset) = 0$ dado que si $x \downarrow -\infty$, $(-\infty, x] \downarrow \emptyset$. Asimismo, $\lim_{x \rightarrow \infty} P_X((-\infty, x]) = \lim_{x \rightarrow \infty} P_X(\mathbb{R}) = 1$. \square

Observemos que estas funciones nos permiten calcular probabilidades de distintos intervalos de \mathbb{R} . Por ejemplo, $P(x < X \leq y) = P((-\inf, y] \setminus (-\inf, x]) = F(y) - F(x)$, $P(x = X) = F(x) - F(x^-)$ o $P(X \geq x) = 1 - F(x^-)$.

Definición 9. Se dice que X tiene **distribución discreta** si $\exists S \subset \mathbb{R}$ contable (finito o numerable) tal que $P(X \in S) = 1$. Se denomina **soporte de la distribución**. Por tanto, $P(X \in B) = \sum_{s \in S \cap B} P(X = s)$, $B \in \mathbb{R}$, dado que $P(X \in B) = P(X \in B \cap S) + P(X \in B \cap S^c) = P(X \in B \cap S) = P(X \in \bigcup_{b \in B \cap S} \{b\}) = \sum_{b \in S \cap B} P(X = b)$ al ser la unión numerable.

Definición 10. Si X es variable aleatoria discreta de soporte S , la función $p_X : \mathbb{R} \rightarrow \mathbb{R}$ dada por $p_X(x) = P(X = x) = P_X(\{x\})$ se denomina **función (de masa de) probabilidad de X** .

Esta verifica que $p(x) = 0$ si $x \in S^c$, dado que por definición $P_X(S^c) = 0$ luego por monotonía se tiene. Además, $P(X \in A) = \sum_{x \in A \cap S} p(x)$, a partir de lo que vimos anteriormente, lo que da pie a $\sum_{x \in S} p(x) = 1$.

Definición 11. $f : \mathbb{R} \rightarrow \mathbb{R}$ es una **función de densidad de probabilidad** si:

1. $f(x) \geq 0 \forall x \in \mathbb{R}$
2. f es Riemann-integrable.
3. $\int_{-\infty}^{\infty} f(t)dt = 1$

Definición 12. Una variable aleatoria X es **(absolutamente) continua** si $\exists f : \mathbb{R} \rightarrow \mathbb{R}$ de densidad tal que $F_X(x) = \int_{-\infty}^x f(t)dt \forall x \in \mathbb{R}$.

Esta función densidad, entonces representa la probabilidad de que la variable aleatoria tome un valor cercano a x , valiendo más $f(x)$ si es muy probable que el valor de X acabe cerca de x . De su definición se extraen las siguientes propiedades:

Observación 2. Sea f de densidad de probabilidad de X variable continua.

1. $P(X = x) = 0$
2. $F_X(x)$ es continua.
3. $\frac{d}{dx} F(x) = f(x)$
4. $P(X \leq x) = \int_{-\infty}^x f(t)dt$
5. $P(x \leq X \leq y) = P(x < X \leq y) = P(x \leq X < y) = P(x < X < y) = F_X(y) - F_X(x) = \int_x^y f(t)dt$

Razón. Es un resultado conocido que las integrales indefinidas de este estilo ($F(x) = \int_a^x f(t)dt$) siempre son continuas (Lipschitz), luego se tiene 2 y por tanto 1, dado que esa probabilidad es $F_X(x) - F_X(x^-) = 0$ por continuidad. Por el Teorema Fundamental del Cálculo se tiene 3, 4 por definición y 5 por definición, teniendo en cuenta que los intervalos pueden estar cerrados o no, al ser continua F . \square

Definición 13. Dada una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ y $g : \mathbb{R} \rightarrow \mathbb{R}$ continua, se induce una nueva variable aleatoria $Y : \Omega \rightarrow \mathbb{R}$ dada por $Y(\omega) = g(X(\omega))$ conocida como **transformada de X por g** .

Proposición 10 (Transformadas discretas). Si X es discreta e Y es la transformada por g , se tiene que $p_Y(y) = \sum_{x:g(x)=y} p_X(x)$ y $F_Y(y) = \sum_{x:g(x) \leq y} p_X(x)$.

Demostración. $p_Y(y) = P(Y = y) = P(\{\omega \in \Omega : g(X(\omega)) = y\}) = \sum_{x:g(x)=y} P(\{\omega \in \Omega : X(\omega) = x\}) = \sum_{x:g(x)=y} p_X(x)$. Por otro lado, $F_Y(y) = P(Y \leq y) = P(\{\omega \in \Omega : g(X(\omega)) \leq y\}) = \sum_{x:g(x) \leq y} P(\{\omega \in \Omega : X(\omega) = x\}) = \sum_{x:g(x) \leq y} p_X(x)$ \square

Proposición 11 (Transformadas continuas). Si X es continua e Y es la transformada por g monótona, se tiene:

$$F_Y(y) = \begin{cases} F_X(g^{-1}(y)) & \text{si } g \text{ crece} \\ 1 - F_X(g^{-1}(y)) & \text{si } g \text{ decrece} \end{cases}$$

Con esto, la densidad se puede obtener derivando con la regla de la cadena: $f_Y(y) = f_X(x) \left| \frac{dx(y)}{dy} \right|$ si $g(x) = y$.

Demostración. $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$ si g crece (y por tanto g^{-1}) y análogo si decrece. \square

2.1. Distribuciones discretas útiles

A continuación veremos algunas distribuciones discretas de probabilidad (daremos la función de masa para cada una).

Sea X una variable aleatoria.

1. **Degenerada.** Fijado $a \in \mathbb{R}$, X sigue una distribución de este tipo si $P(X = a) = 1$, de modo que realmente no es aleatoria como tal.
2. **Bernoulli.** Supongamos que realizamos un experimento con dos posibles resultados (éxito/fracaso). Si la probabilidad de éxito viene dada por el parámetro $p \in (0, 1)$, y X representa el número de éxitos, está claro que $P(X = 0) = (1 - p)$ y $P(X = 1) = p$. En este caso X sigue la distribución de Bernoulli de parámetro p , y se denota $X \sim B(1; p)$.
3. **Uniforme.** Si la variable X cumple $P(X = k) = \frac{1}{n}$ para $k \in \{a_1, \dots, a_n\}$, se dice que sigue una distribución uniforme y se denota $X \sim U(\{a_1, \dots, a_n\})$.
4. **Binomial.** Generalizando el caso anterior, supongamos que realizamos n veces el experimento, siendo cada resultado independiente de los anteriores. Entonces, está claro que si X mide el número de aciertos, se tiene:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (k \in \{0, 1, \dots, n\})$$

Esta distribución es la binomial de parámetros n y p : $X \sim B(n; p)$. Está claro que el resto de valores tienen masa 0, ya que la suma de estos $n + 1$ valores es precisamente $(p + 1 - p)^n = 1$.

Las variables de Bernoulli independientes se pueden sumar: si X_1, \dots, X_k tienen $X_i \sim B(n_i; p)$ y $n = n_1 + n_2 + \dots + n_k$, entonces $X_1 + X_2 + \dots + X_k = X \sim B(n; p)$. Es decir, si cada variable mide el número de éxitos en n_i intentos, y los intentos son independientes, al sumarlas estamos midiendo efectivamente el número de éxitos en $n_1 + n_2 + \dots + n_k$ intentos, y como cada variable puede valer de 0 a n_i , la suma irá de 0 a n como se espera.

5. **Geométrica.** Si en los experimentos anteriores la variable X representa el número de fracasos antes que un éxito, está claro que:

$$P(X = k) = p(1 - p)^k \quad (k \in \{0, 1, \dots\})$$

Y se dice que sigue una distribución geométrica: $X \sim G(p)$. En ocasiones el exponente es $k - 1$, y sería el número de intentos hasta un éxito. Está claro que suma 1 al ser una serie geométrica de suma $\frac{p}{1 - 1 + p}$.

6. **Binomial negativa** Si X mide el número de fracasos antes de obtener t éxitos, donde p es la probabilidad de éxito en cada intento independiente del experimento, se tiene que:

$$P(X = k) = \binom{t+k-1}{k} p^t (1-p)^k \quad (k \in \{0, 1, \dots\})$$

Dado que, fijando el acierto t -ésimo final, hay que elegir en que orden suceden los anteriores $t-1$ aciertos y k fallos. Las variables que cumplen esta función de masas siguen una distribución binomial negativa de parámetros t, p y se denotan $X \sim BN(t; p)$.

7. **Poisson.** Supongamos que sabemos que un evento ocurre, de media, λ veces por unidad de tiempo (también valen otras unidades, por supuesto). Sea la variable aleatoria X el número de eventos que ocurren por unidad de tiempo.

Para calcular la probabilidad de que $X = k$, dividimos dicha unidad de tiempo en n partes iguales, de tal modo que en cada parte, de media, esperamos $\frac{\lambda}{n}$ sucesos. Veamos que si $n \rightarrow \infty$, la probabilidad de que ocurran 2 sucesos o más en cada una de estas partes es sumamente pequeña (debemos esperar, si ocurren k sucesos, que ocurra 1 en k partes y 0 en las restantes, siendo altamente improbable que 2 ocurran en una misma parte al ser tan pequeña), con lo cual si X_i es el número de sucesos en la i -ésima parte, tenemos que $X_i \sim B(1, \frac{\lambda}{n})$ y por tanto $X = \sum X_i \sim B(n, \frac{\lambda}{n})$.

Es decir, con esta estrategia, y $n \rightarrow \infty$, se tiene: $P(X = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{(n-k)! k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$.

Las variables que siguen esta distribución:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Se dice que siguen una distribución de Poisson y se denotan $X \sim P(\lambda)$.

Con lo anterior, las binomiales con n grande (digamos $n \geq 30$) y p o $(1-p)$ pequeñas (digamos alguna menor que 0,1), pueden ser estimadas por una Poisson con $\lambda = np$.

8. **Hipergeométrica.** Si X cuenta el número de bolas rojas (o 'fracasos') extraídas sin reemplazamiento extraídas de una urna de N bolas tras n extracciones con r rojas inicialmente, se tiene:

$$P(X = k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}} \quad \max\{0, n - N + r\} \leq k \leq \min\{n, r\}$$

Y se dice que X sigue una distribución hipergeométrica. Se suele usar en controles de calidad.

2.2. Distribuciones continuas útiles

A continuación veremos las densidades de algunas distribuciones continuas útiles:

1. **Uniforme en (a,b).** Aparece en la generación de números aleatorios entre a y b , y está dada por:

$$f(x) = \frac{1}{b-a} 1_{(a,b)}(x)$$

Se denota $X \sim U(a, b)$. Se tiene para cualquier aleatoria X que, si F_X es continua, $Y = F_X(X)$ es uniforme en $[0, 1]$.

2. **Normal.** De parámetros $\mu \in \mathbb{R}$ y $\sigma > 0$. Aparece con frecuencia en los experimentos aleatorios. Tiene forma de "campana", centrada en μ y más o menos estirada según menor o mayor sea σ , y está dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

La **variable normal tipificada** es aquella que sigue la distribución $N(0;1)$, y sus valores de distribución se suelen recoger en tablas, dado que la integral de esta densidad es trascendente y no puede ser calculada con facilidad.

Toda variable normal X se puede **tipificar o normalizar** tras la transformación $Z = \frac{X-\mu}{\sigma}$, y de ahí calcularse con facilidad su distribución.

3. **De Cauchy.** Es la dada por $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$. Es una distribución de *Cola pesada*, donde la masa que se condensa en los extremos de la distribución es considerable, a diferencia de en la normal.

4. **Exponencial.** Dada por $f(x) = \lambda e^{-\lambda x}$ para $\lambda > 0$ y $x \geq 0$ (si no, es cero). Modela el tiempo que transcurre hasta que sucede un suceso, de intensidad λ .

Se caracteriza por ser una distribución que **no tiene memoria**. Se dice que una variable aleatoria **no tiene memoria** si $P(X > (s+t) | X > s) = P(X > t) \forall s, t > 0$. Es decir, si lo interpretamos como ocurrencia pasado un tiempo, la probabilidad de que algo ocurra empezando desde el instante s del tiempo, t unidades después, es la misma que la de que ocurra t unidades desde el comienzo.

De hecho, se tiene la proposición:

Proposición 12. *Se tiene que X variable continua positiva no tiene memoria \iff sigue una distribución exponencial.*

Demostración. Si sigue una exponencial, $P(X > t) = \int_t^\infty \lambda e^{-\lambda x} dx = e^{-\lambda t}$, y por definición $P(X > (s+t) | X > s) = \frac{P(X > (s+t) \cap X > s)}{P(X > s)} = \frac{P(X > (s+t))}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}$.

Para lo contrario, sea F la distribución de una variable sin memoria. En ese caso, $(1 - F(t)) = \frac{1 - F(t+s)}{1 - F(s)}$, luego $(1 - F(t))(1 - F(s)) = (1 - F(t+s))$. Ahora, sea $G(x) = 1 - F(x)$. Vemos que G ha de satisfacer $G(t)G(s) = G(t+s)$ para $t, s > 0$. La única ecuación que satisface esta ecuación funcional es $G(x) = e^{-\lambda x}$ con x positivo, luego $F(x) = 1 - e^{-\lambda x}$ y por tanto, integrando, se llega a la distribución exponencial. Para llegar a este resultado con la ecuación funcional, basta ver que $G(0) = 1$, dado que $G(a) = G(0)G(a)$, y G es no nula, y por tanto: $G'(x) = \lim_{h \rightarrow 0} \frac{G(x+h) - G(x)}{h} = \lim_{h \rightarrow 0} \frac{G(x)G(h) - G(x)}{h} = \lim_{h \rightarrow 0} \frac{G(x)(G(h) - G(0))}{h} = G(x)G'(0)$, con lo que $\log(G(x))' = G'(0)$, y basta con integrar y darse cuenta de que la constante de integración debe ser 1 para que sea distribución, y que $G'(0) < 0$ al ser G decreciente. \square

5. **Gamma.** Generaliza las exponenciales, en tanto que si una variable representa la probabilidad de tener que esperar un tiempo determinado para que un evento, que ocurre de media β veces por unidad de tiempo, ocurra α veces, entonces esa variable sigue esa distribución. Tiene densidad:

$$f(x) = \frac{x^{\alpha-1} e^{-\beta x} \beta^\alpha}{\Gamma(\alpha)} \quad (x, \alpha, \beta > 0)$$

Donde:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

Que como vemos varía como $x^{\alpha-1} e^{-\beta x}$. El resto de constantes se agregan como siempre para que integre 1. Según hemos comentado antes, y como se observa, tomar $\alpha = 1$ y $\beta = \lambda$ da una exponencial. Asimismo, con esta expresión, los parámetros α, β no necesariamente han de ser naturales. Si $\alpha = \frac{n}{2}$ con $n \in \mathbb{N}$ y $\beta = \frac{1}{2}$ se conoce como **distribución χ -cuadrado de n grados de libertad**.

6. **Distribución Beta** Permiten formar distribuciones en intervalos compactos. Viene dada por dos parámetros, $\alpha, \beta > 0$ y tiene densidad:

$$f(x) = \frac{1}{\beta(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad (x \in [0, 1])$$

Donde $\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ para valores positivos.

7. **Weibull**. Modeliza variables que representan la distribución de fallos en sistemas, cuando la tasa de fallos es proporcional al tiempo que pasa. Esto último lo indica el parámetro $k > 0$, y pueden ser escaladas con el parámetro $\theta > 0$. Está dada por:

$$f(x) = \frac{k}{\theta} \left(\frac{x}{\theta}\right)^{k-1} e^{-\left(\frac{x}{\theta}\right)^k}$$

Tambien degeneran en una exponencial si $k = 1$ (los fallos no aumentan con el tiempo ni disminuyen) y $\theta = \frac{1}{\lambda}$.

8. **Pareto**. Modeliza distribuciones de ingresos, sobre todo, debido a ser otra distribución *Heavy-tail*, es decir, no decae tan rápido como las demás que decaían como exponenciales. Tiene parámetros $a, \theta > 0$ y está dada por:

$$f(x) = \frac{\theta a^\theta}{x^{\theta+1}} \quad (x > a)$$

9. **Lognormal**. Es la que se obtiene de transformar por $Y = e^X$ una variable normal X , es decir, es la distribución que sigue la exponencial de una normal. Si la normal representaba la suma de muchos factores independientes, esta representa el producto de muchos factores independientes, y tiene densidad:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2}$$

3. Vectores aleatorios

Al igual que las variables aleatorias permitían asignar valores a los sucesos aleatorios de un espacio de probabilidad, los vectores aleatorios permiten asignar elementos de \mathbb{R}^n a sucesos. Estudiar varias cantidades de un suceso a la vez dará más información que estudiarlas individualmente.

Definición 14. Dado un espacio de probabilidad (Ω, \mathcal{F}, P) , un **vector aleatorio** es una aplicación $X : \Omega \rightarrow \mathbb{R}^d$ que asigna a cada elemento del espacio muestral un vector de d entradas en \mathbb{R} . Además debe verificar la condición de medibilidad que vimos al comienzo de la sección 2 en cada componente individual.

Al igual que las variables aleatorias, inducen una medida:

Definición 15. Los vectores aleatorios inducen una **medida de probabilidad o distribución de X** en \mathbb{R}^d dada por $P_X : B \subset \mathcal{P}(\mathbb{R}^d) \rightarrow [0, 1]$, con $P_X(A) = P(\mathbf{X} \in A) = P(\{\omega : \mathbf{X}(\omega) \in A\})$

Si solo atendemos a una componente de $\mathbf{X} = (X_1, \dots, X_d)$, tenemos una variable aleatoria conocida como **proyección o variable marginal**. Conociendo la distribución completa es fácil conocer la marginal en cualquier variable.

Observación 3. Para obtener la probabilidad marginal conociendo la completa, basta con observar que $(X_j \in \mathbb{R}) = \Omega$, y entonces: $P_{X_i}(A) = P(X_i \in A) = P((X_i \in A) \cap (X_1 \in \mathbb{R}) \cap (X_2 \in \mathbb{R}) \cap \dots \cap (X_d \in \mathbb{R})) = P(\mathbf{X} \in \mathbb{R} \times \mathbb{R} \times \dots \times A \times \dots \times \mathbb{R})$. Por tanto, basta con tener en cuenta todas las probabilidades de cuando $X_i \in A$, para cualquier otro valor de las variables.

Por lo general no es posible obtener la distribución de un vector aleatorio sabidas las marginales.

Ahora pasaremos a tratar de definir la función de distribución de un vector, que nos da, como sabemos, la probabilidad de que el vector tenga un valor *menor* a uno dado.

Definición 16. Dado $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, la **región suroeste** de x es el conjunto: $S_x = (-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_d]$.

Definición 17. La **función de distribución del vector aleatorio X** es $F_X : \mathbb{R}^d \rightarrow [0, 1]$ dada por $F_X(x) = P_X(S_x) = P(X \in S_x) = P(X_1 \leq x_1; X_2 \leq x_2; \dots; X_d \leq x_d)$.

Veremos que verifica las mismas propiedades que la función de distribución para variables reales, y que es equivalente dar una función de distribución a una vector aleatorio (dado que describe por completo su comportamiento).

Definición 18. Una función $F : \mathbb{R}^d \rightarrow \mathbb{R}$ es **no decreciente** si, dados $(a_i)_1^d, (b_i)_1^d$ con $a_i \leq b_i$ en todas las coordenadas, se tiene:

$$\Delta(a_1, b_1)\Delta(a_2, b_2) \dots \Delta(a_d, b_d)F(x) \geq 0$$

Donde $\Delta(a_i, b_i)F(x_1, x_2, \dots, x_d) = F(x_1, x_2, \dots, b_i, \dots, x_d) - F(x_1, x_2, \dots, a_i, \dots, x_d)$.

Es decir, se pide que crezca en cada componente y no solo globalmente.

Observación 4 (Propiedades de las funciones distribución). Una función F de distribución de un vector aleatorio verifica:

1. F es no decreciente.
2. F es continua por la derecha en cada componente.
3. $\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_d) = 0$. Para cualquier componente.
4. $\lim_{(x_1, \dots, x_d) \rightarrow \infty} F(x_1, \dots, x_d) = 1$

Razón. Para las dos primeras se puede operar como en variables aleatorias, solo que fijando cada conjunto de la región suroeste salvo el que se vaya a considerar. La tercera es cierta dado que llevar cualquier componente de la región suroeste a cero da lugar al conjunto vacío, y por continuidad hacia abajo de la probabilidad se tiene. La cuarta se tiene porque llevar todos los componentes a infinito cubre todo \mathbb{R}^d y por continuidad hacia arriba se tiene.

Proposición 13. *Se puede obtener la función de distribución de cada componente del vector, sabiendo que:*

$$F_{X_i}(x_i) = \lim_{x_j \rightarrow \infty, j \neq i} F_X(x_1, \dots, x_d)$$

Demostración. $F_{X_i}(x_i) = P(X_i \leq x_i) = P(X_i \leq x_i, X_k \in \mathbb{R} \forall k \neq i) = \lim_{x_j \rightarrow \infty, j \neq i} P(X_1 \leq x_i; X_2 \leq x_2; \dots; X_d \leq x_d) = \lim_{x_j \rightarrow \infty, j \neq i} F_X(x_1, \dots, x_d)$ \square

Definición 19. Un **vector aleatorio discreto** es aquel tal que $\exists S \subset \mathbb{R}^d$ contable tal que $P(\vec{X} \in S) = 1$. Por tanto, $P(\vec{X} \in B) = \sum_{s \in S \cap B} p(s)$, donde $p : \mathbb{R}^d \rightarrow \mathbb{R}$ es la **función de masa conjunta**, con $p(x) = P(\vec{X} = x)$.

Teorema 1. $\vec{X} = (X_1, X_2, \dots, X_d)$ es discreto $\iff X_i$ es variable aleatoria discreta con $i \in \{1, \dots, d\}$. Además, $p_{X_i}(x_i) = \sum_{j_1, \dots, j_{i-1}, j_{i+1}, \dots, j_d} p(x_{j_1}, \dots, x_i, \dots, x_{j_d})$, es decir, las masas marginales se obtienen sumando por todos los posibles valores en el resto de las coordenadas, fijada la que queremos.

Demostración. Si \vec{X} es discreto, y S es su soporte, definimos $S_i = \{x \in \mathbb{R} : (a_1, a_2, \dots, a_{i-1}, x, a_{i+1}, \dots, a_d) \in S, a_j \in \mathbb{R}\}$, la proyección de S sobre la i -ésima coordenada. Es decir, si $\vec{X}(\omega) \in S$, ha de darse que $X_i(\omega) \in S_i$ por construcción. Es decir, $\{\vec{X} \in S\} \subset \{X_i \in S_i\}$, y por tanto $P(\{X_i \in S_i\}) = 1$. Además ha de ser contable por construcción. (Su cardinal será menor o igual que el de S).

Para el recíproco, si S_i es el soporte de cada variable marginal, tomando $S = S_1 \times S_2 \times \dots \times S_d$, que es contable, se tiene $\vec{X}(\omega) \in S \iff (X_1(\omega), X_2(\omega), \dots, X_d(\omega)) \in S \iff X_i(\omega) \in S_i$ por construcción, de modo que $P(\{\vec{X} \in S\}) = P(\cap_1^d \{X_i(\omega) \in S_i\})$ (son el mismo conjunto). Como todos los conjuntos de la intersección tienen probabilidad 1, su intersección la tiene.

Para la expresión de las masas marginales, basta con ver que $P(X_i = x_i) = P(X_i = x_i, X_j \in \mathbb{R} \text{ si } j \neq i)$, y por tanto basta con sumar las masas no nulas de esos valores (fijando la i -ésima coordenada y tomando las coordenadas restantes que dan lugar a vectores de S). La suma puede hacerse al ser S contable. \square

Definición 20. Una función $f : \mathbb{R}^d \rightarrow \mathbb{R}$ es una **densidad de probabilidad** si es positiva, integrable Riemann, y verifica:

$$\int \dots \int_{\mathbb{R}^d} f = 1$$

Definición 21. Un **vector aleatorio continuo** es aquel tal que $\exists f$ de densidad que verifica:

$$F(x_1, \dots, x_d) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_d} f(t_1, \dots, t_d) dt_1 dt_2 \dots dt_d = \int \dots \int_{S_x} f$$

Para todo $(x_1, \dots, x_d) \in \mathbb{R}^d$.

Por tanto, la probabilidad de A se obtiene: $P(\vec{X} \in A) = \int \dots \int_A f$.

Asimismo, por el teorema fundamental del cálculo, y bajo condiciones de continuidad de las derivadas parciales, se tiene: $\frac{\partial^d F}{\partial x_1 \partial x_2 \dots \partial x_d} = f$.

Teorema 2. Si \vec{X} es continuo, lo son sus componentes X_i y además:

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_d) dx_1 dx_2 \dots dx_{i-1} dx_{i+1} \dots dx_d$$

Demostración. Sea f la densidad de \vec{X} . Ahora, como es habitual, $F_i(x_i) = P(X_i \leq x_i) = P(X_i \leq x_i, X_j \in \mathbb{R} \text{ si } j \neq i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{x_i} \dots \int_{-\infty}^{\infty} f(t_1, \dots, t_d) dt_1 \dots dt_d = \int_{-\infty}^{x_i} [\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(t_1, \dots, t_d) dt_1 dt_2 \dots dt_{i-1} dt_{i+1} \dots dt_d] dt_i$.

Por tanto, la función entre corchetes es su densidad de probabilidad y por tanto es continua la componente. \square

3.1. Algunas distribuciones utiles

A continuación veremos algunos ejemplos de distribuciones discretas para vectores aleatorios:

1. **Producto de discretas.** El vector (X, Y) tiene distribución de producto de binomiales, $(X, Y) \simeq B(n_1, p_1) \otimes B(n_2, p_2)$, si su masa está dada por:

$$P(X = k, Y = l) = \binom{n_1}{k} p_1^k (1 - p_1)^{n_1 - k} \binom{n_2}{l} p_2^l (1 - p_2)^{n_2 - l}$$

Sumando en una de las variables para obtener las marginales, está claro que la otra sale de factor común y la suma da 1, luego las distribuciones marginales son cada una de las binomiales, independientemente.

Sumando en k y en l se sacan factor común las constantes y se obtiene que la suma es el producto de las sumas en k y en l de cada binomial, es decir, $1 \cdot 1 = 1$. Esto se cumple para cualquier distribución discreta, por tanto. Por ejemplo, el producto de Poisson:

$$P(X = k, Y = l) = e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^l}{l!}$$

2. **Trinomial.** Supongamos que un experimento tiene tres resultados (por ejemplo, victoria, empate y derrota), con probabilidades p_1 y p_2 , ($p_1 + p_2 \leq 1$) para los dos primeros. Si el vector (X, Y) representa la probabilidad de obtener X veces el primer resultado e Y veces el segundo, tras n intentos independientes, está claro que la masa es:

$$P((X, Y) = (k, l)) = \binom{n}{k} \binom{n-k}{l} p_1^k p_2^l (1 - p_1 - p_2)^{n-k-l} \quad (k + l \leq n)$$

Se dice que sigue una distribución trinomial. Un interesante hecho es que las distribuciones marginales son también binomiales de parámetros n y p_1 o p_2 según qué variable. En efecto:

$$P(X = k) = \sum_{l=0}^{n-k} \binom{n}{k} \binom{n-k}{l} p_1^k p_2^l (1 - p_1 - p_2)^{n-k-l} = \binom{n}{k} p_1^k \sum_{l=0}^{n-k} \binom{n-k}{l} p_2^l (1 - p_1 - p_2)^{n-k-l} = \binom{n}{k} p_1^k (1 - p_1 - p_2 + p_2)^{n-k} \text{ y análogo para la otra variable.}$$

Al igual que sucede con el producto de discretas, es razonable que el producto de continuas también sea continuo.

Lema 1. Si f y g son densidades en \mathbb{R} , lo es $h(x, y) = f(x)g(y)$ en \mathbb{R}^2 .

Demostración. Está claro que sigue siendo positiva e integrable. Asimismo,

$$\iint_{\mathbb{R}} f(x)g(y)dA = \int_{\mathbb{R}} f(x)dx \int_{\mathbb{R}} g(y)dy = 1 \cdot 1 = 1. \quad \square$$

Además, siguiendo el mismo argumento, sus marginales serían f y g respectivamente.

Algunos ejemplos son:

1. **Uniforme en rectángulo.** Está dada en por:

$$f(x, y) = \frac{1}{(b-a)(d-c)} \chi_{[a,b] \times [c,d]}(x, y)$$

2. **Uniforme en disco unidad.** Está dada para $(x, y) \in B_1((0, 0))$, es decir, con $x^2 + y^2 \leq 1$, por:

$$f(x, y) = \frac{1}{\pi}$$

Y 0 fuera. Sus densidades marginales son $f_i(x) = \frac{2\sqrt{1-x^2}}{\pi}$, lo que se obtiene integrando como sabemos.

3. **Producto de exponenciales.** Está dada para $x, y > 0$ por:

$$f(x, y) = \lambda \mu e^{-(\lambda x + \mu y)}$$

4. **Normal multivariante.** De parámetros:

$$\vec{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix}, \quad \vec{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{dd} \end{pmatrix}$$

Con cada $\mu_i, \sigma_{ij} > 0$. Está dada por:

$$f(\vec{x}) = \frac{1}{\sqrt{|\Sigma|}(2\pi)^d} \cdot e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^t \cdot \Sigma^{-1} \cdot (\vec{x}-\vec{\mu})}$$

Definición 22 (Distribución condicionada). Sea (X, Y) discreto. Se define la **distribución condicionada discreta** de Y cuando $X = x_0$ como la dada por la masa $p_{Y|X=x_0} = \frac{p(x_0, y)}{p_1(x_0)}$, donde p es la masa conjunta de (X, Y) y p_1 es la masa marginal de X .

Sea (X, Y) continuo. Se define la **distribución condicionada continua** de Y cuando $X = x_0$ como la dada por la densidad $f_{Y|X=x_0} = \frac{f(x_0, y)}{f_1(x_0)}$, donde f es la densidad conjunta de (X, Y) y p_1 es la densidad marginal de X .

Definición 23 (Independencia de variables discretas). Si el vector (X_1, \dots, X_d) es discreto, las variables X_i **son independientes** si se tiene $P(X_1 = x_1, X_2 = x_2, \dots, X_d = x_d) = \prod_1^d P(X_i = x_i) \forall x_1, \dots, x_d \in \mathbb{R}$.

Observamos que en ese caso la distribución condicionada coincide con la marginal. Además:

Proposición 14. En caso de que X_1, \dots, X_d sean independientes, los conjuntos $\{\{X_i = x_i\}\}_{i=1}^d$ son mutuamente independientes.

Demostración. Vamos a ver un ejemplo tomando 2 conjuntos: $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1, X_2 = x_2; X_3, X_4, \dots, X_d \in \mathbb{R}) = \sum_{x_3, x_4, \dots, x_d} P(X_1 = x_1, X_2 = x_2, \dots, X_d = x_d) = P(X_1 = x_1)P(X_2 = x_2) \sum_{x_3, \dots, x_d} \prod_3^d P(X_i = x_i) = P(X_1 = x_1)P(X_2 = x_2) \prod_3^d (1) = P(X_1 = x_1)P(X_2 = x_2)$. Análogamente se hace para cualquier otro número de variables, con lo que se tiene la independencia. \square

Definición 24 (Independencia de variables continuas). Si el vector (X_1, \dots, X_d) es continuo, las variables X_i **son independientes** si se tiene $f(x_1, \dots, x_d) = f_1(x_1) \dots f_d(x_d)$.

Teorema 3. Sea (X_1, \dots, X_d) un vector aleatorio, continuo o discreto. Entonces:

$$X_1, \dots, X_d \text{ son independientes} \iff F(x_1, x_2, \dots, x_d) = F_1(x_1) \dots F_d(x_d)$$

Demostración. Para el caso discreto, el \implies se puede probar sabiendo que $F(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d) = \sum_{y_i \leq x_i} P(X_1 = y_1, X_2 = y_2, \dots, X_d = y_d) = \sum_{y_i \leq x_i} \prod_{j=1}^d P(X_j = y_j) = \prod_{j=1}^d \sum_{y_i \leq x_i} P(X_j = y_j) = \prod_{j=1}^d F(x_i)$. Para \impliedby , haremos el caso $d = 2$, siendo el resto análogos. Debemos darnos cuenta

de que $\lim_{h \rightarrow 0} P((X_1, X_2) \in (x_1 - h, x_1] \times (x_2 - h, x_2]) = P((X_1, X_2) = (x_1, x_2))$, al converger el cuadrado que consideramos a un punto. Ahora, veamos lo siguiente: $P((x, y) \in (x_1 - h, x_1] \times (x_2 - h, x_2]) = F(x_1, x_2) - F(x_1 - h, x_2) - F(x_1, x_2 - h) + F(x_1 - h, x_2 - h) = F_1(x_1)F_2(x_2) - F_1(x_1 - h)F_2(x_2) - F_1(x_1)F_2(x_2 - h) + F_1(x_1 - h)F_2(x_2 - h) = (F_1(x_1) - F_1(x_1 - h))(F_2(x_2) - F_2(x_2 - h))$. Tomando límites a ambos lados cuando h tiende a 0, se tiene que $P((X, Y) = (x_1, x_2)) = P(X_1 = x_1)P(X_2 = x_2)$.

Para el caso continuo, el \implies se obtiene sabiendo que $F(x_1, \dots, x_d) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f(t_1, \dots, t_d) dt_1 \dots dt_d = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f(t_1) \dots f(t_d) dt_1 \dots dt_d = \prod_{i=1}^d \int_{-\infty}^{x_i} f_i(t_i) dt_i = \prod_{i=1}^d f_i(x_i)$. El \impliedby se obtiene sabiendo que $f(x_1, \dots, x_d) = \frac{\partial^d F(x_1, \dots, x_d)}{\partial x_1 \dots \partial x_d} = \frac{\partial^d F_1(x_1) \dots F_d(x_d)}{\partial x_1 \dots \partial x_d} = \frac{\partial F_1(x_1)}{\partial x_1} \dots \frac{\partial F_d(x_d)}{\partial x_d} = f_1(x_1) \dots f_d(x_d)$. \square

Definición 25 (Distribución derivada). Dado $X = (X_1, \dots, X_d)$ vector aleatorio, y una transformación continua $T : \mathbb{R}^d \rightarrow \mathbb{R}^k$, conocida su distribución, se quiere encontrar la distribución de $Y = T(X_1, \dots, X_d)$, conocida como **distribución derivada**. Esta viene dada por $P_Y(B) = P(Y \in B) = P(T(X) \in B) = P(X \in T^{-1}(B))$, aunque en la práctica no siempre es fácil hallar estas preimágenes, y veremos formas de hacerlo.

Teorema 4. Sea $(X, Y) : \Omega \rightarrow A \subset \mathbb{R}^2$ un vector aleatorio continuo de densidad $h(x, y)$. Sea $T : A \rightarrow B \subset \mathbb{R}^2$ un difeomorfismo (morfismo biyectivo diferenciable en ambos sentidos) con $T^{-1}(u, v) = (x(u, v), y(u, v))$. El vector $(U, V) = T(X, Y)$ es aleatorio, de densidad:

$$g(u, v) = \begin{cases} h(x(u, v), y(u, v)) \cdot |J_{T^{-1}}| & \text{si } (u, v) \in B \\ 0 & \text{si no.} \end{cases}$$

$$\text{Donde } J_{T^{-1}} = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}.$$

Demostración. Sea $C \subset B$. Entonces $P((U, V) \in C) = P((X, Y) \in T^{-1}(C)) = \iint_{T^{-1}(C)} h(x, y) dx dy = \iint_C h(x(u, v), y(u, v)) \cdot |J_{T^{-1}}| du dv$, por el teorema de cambio de variable. Por tanto, esa es la densidad. Fuera de B , como es la imagen de T , no puede acabar ningún valor, por lo que no tienen probabilidad. \square

4. Esperanza y momentos

Definición 26. Si X es una variable aleatoria discreta, su **esperanza o media** es:

$$\mathbb{E}(X) = \sum_{x \in S} x P(X = x) = \sum_{x \in S} xp(x)$$

Siempre y cuando la suma converja absolutamente.

Definición 27. Si X es una variable aleatoria continua, su **esperanza o media** es:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x) dx$$

Siempre y cuando la integral converja absolutamente.

Interpretación. La esperanza es el valor real alrededor del cual se concentra la masa de probabilidad. La distribución se equilibra en torno a ese valor (es el *centro de gravedad* de la distribución).

Hay distribuciones sin esperanza, por no converger absolutamente las sumas o integrales. Un ejemplo es la de Cauchy.

A continuación nos planteamos cuál es la esperanza de la variable $Y = g(X)$ si X es una transformación continua.

Proposición 15 (Ley del estadístico inconsciente). *Si g es transformación continua, la variable $Y = g(X)$ tiene esperanza:*

$$\mathbb{E}(Y) = \sum_{x \in S} g(x)P(X = x) = \sum_{x \in S} g(x)p(x)$$

Si es discreta, y :

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Si es continua. p o f son las funciones de X .

Demostración. Se dará para el caso discreto, siendo la del caso continuo algo más larga pero con la misma idea. Siendo I la imagen de X (valores no nulos que toma), entonces Y tiene imagen $g(I)$. Por tanto, $\mathbb{E}(Y) = \sum_{y \in g(I)} yP(Y = y) = \sum_{y \in g(I)} y \sum_{x \in I, g(x)=y} P(X = x) = \sum_{x \in I} g(x)P(X = x)$. \square

Proposición 16 (Esperanza de transformaciones de vectores). *Sea (X, Y) un vector aleatorio, y $Z = g(x, y)$ una variable aleatoria tras una transformación. Entonces, si el vector es discreto:*

$$\mathbb{E}(Z) = \sum_x \sum_y g(x, y)P(X = x, Y = y)$$

Y si es continuo:

$$\mathbb{E}(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)dxdy$$

Donde f es la densidad conjunta.

Demostración. Es análoga a la proposición anterior, solo que la imagen I está en \mathbb{R}^2 . \square

Proposición 17 (Propiedades de la esperanza). *Se tienen las siguientes propiedades:*

1. $\mathbb{E}(c) = c \quad \forall c \in \mathbb{R}$
2. $\mathbb{E}(cX) = c\mathbb{E}(X) \quad \forall c \in \mathbb{R}$
3. Si X, Y son integrables (con esperanza finita) entonces lo es $aX + bY \quad \forall a, b \in \mathbb{R}$, y se tiene $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.
4. $X \geq 0 \implies \mathbb{E}(X) \geq 0$.
5. $X \leq Y \implies \mathbb{E}(X) \leq \mathbb{E}(Y)$. En particular, $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$
6. Si X, Y son independientes, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Demostración.

1. Se tiene $\mathbb{E}(c) = c \cdot P(c = c) = c \cdot 1 = 1$.
2. En el caso discreto, $\mathbb{E}(cX) = \sum_{x \in S} cxP(X = x) = c \sum_{x \in S} xP(X = x) = c\mathbb{E}(X)$, y en el continuo, $\mathbb{E}(cX) = \int_{\mathbb{R}} cx f(x)dx = c \int_{\mathbb{R}} x f(x)dx = c\mathbb{E}(X)$. Para 3, primero probaremos que es integrable.

3. En el caso discreto, $\mathbb{E}(|aX + bY|) = \sum_x \sum_y |ax + by|P(X = x, Y = y) \leq \sum_x \sum_y (|ax| + |by|)P(X = x, Y = y) = |a| \sum_x |x| \sum_y P(X = x, Y = y) + |b| \sum_y |y| \sum_x P(X = x, Y = y) = |a| \sum_x |x|P(X = x) + |b| \sum_y |y|P(Y = y) = |a|\mathbb{E}(|X|) + |b|\mathbb{E}(|Y|) < +\infty$, luego es integrable.

$$\text{Asimismo: } \mathbb{E}(aX + bY) = \sum_x \sum_y (ax + by)P(X = x, Y = y) = a \sum_x x \sum_y P(X = x, Y = y) + b \sum_y y \sum_x P(X = x, Y = y) = a \sum_x xP(X = x) + b \sum_y yP(Y = y) = a\mathbb{E}(X) + b\mathbb{E}(Y) < +\infty$$

$$\text{En el caso continuo, } \mathbb{E}(|aX + bY|) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |ax + by|f(x, y)dxdy \leq \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (|ax| + |by|)f(x, y)dxdy = |a| \int_{-\infty}^{+\infty} |x| \int_{-\infty}^{+\infty} f(x, y)dxdy + |b| \int_{-\infty}^{+\infty} |y| \int_{-\infty}^{+\infty} f(x, y)dxdy = |a| \int_{-\infty}^{+\infty} |x|f_X(x)dx + |b| \int_{-\infty}^{+\infty} |y|f_Y(y)dy = |a|\mathbb{E}(|X|) + |b|\mathbb{E}(|Y|) < +\infty$$
, luego es integrable.

$$\text{Asimismo: } \mathbb{E}(aX + bY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (ax + by)f(x, y)dxdy =$$

$$a \int_{-\infty}^{+\infty} x \int_{-\infty}^{+\infty} f(x, y)dxdy + b \int_{-\infty}^{+\infty} y \int_{-\infty}^{+\infty} f(x, y)dxdy = a \int_{-\infty}^{+\infty} xf_X(x)dx + b \int_{-\infty}^{+\infty} yf_Y(y)dy = a\mathbb{E}(X) + b\mathbb{E}(Y) < +\infty$$

4. En el caso discreto, $\mathbb{E}(X) = \sum_x xP(X = x) \geq 0$ al ser los $x \geq 0$ y $P(X = x) \in [0, 1]$. En el continuo, $\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} xf(x)dx \geq 0$, ser $f(x) \geq 0$.
5. Como $Y - X \geq 0$, entonces $\mathbb{E}(Y - X) \geq 0$ y por linealidad $\mathbb{E}(Y) - \mathbb{E}(X) \geq 0$. Lo segundo se tiene porque como $-|X| \leq X \leq |X|$, entonces $-\mathbb{E}(|X|) \leq \mathbb{E}(X) \leq \mathbb{E}(|X|)$, luego $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.
6. En el caso discreto, $\mathbb{E}(XY) = \sum_x \sum_y xyP(X = x, Y = y) = \sum_x \sum_y xyP(X = x)P(Y = y) = \sum_x xP(X = x) \sum_y yP(Y = y) = \mathbb{E}(X)\mathbb{E}(Y)$. En el caso continuo, $\mathbb{E}(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dxdy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dxdy = \int_{-\infty}^{\infty} xf_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy = \mathbb{E}(X)\mathbb{E}(Y)$.

□

Definición 28. Sea (X, Y) discreto con $P(X = x) > 0$. Se define la **esperanza condicional** como el valor:

$$\mathbb{E}(Y|X = x) = \sum_y yP(X = x, Y = y) = \sum_y y \frac{P(X = x, Y = y)}{P(X = x)}$$

Representa el valor que se espera de Y si sabemos que X vale x .

Del mismo modo, si (X, Y) es continuo, su esperanza condicional es:

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y \frac{f(x, y)}{f_X(x)}$$

Si bien esta esperanza es un valor fijo, en ocasiones nos interesa plantearnos cuál es la probabilidad de que el valor esperado de Y sea uno dado, sabiendo que depende de X que es aleatoria. Esto da lugar a una nueva variable aleatoria:

Definición 29. Se define la **variable esperanza condicional** como $T(\omega) = \mathbb{E}(Y|X = X(\omega))$, para $\omega \in \Omega$. Se suele denotar también como $\mathbb{E}(Y|X)$ a esta variable. Vemos que es función de lo que valga la variable X .

Es interesante observar que $\mathbb{E}(Y|X)$ toma el valor $\mathbb{E}(Y|X = x)$ con una probabilidad $P(X = x)$. Conviene **definir** $\mathbb{E}(Y|X = x) = 0$ si $P(X = x) = 0$.

Teorema 5. Sea Y una variable integrable. Entonces, $\mathbb{E}(Y|X)$ también es integrable, y se verifica que $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$.

Demostración. Caso discreto. Como $\mathbb{E}(Y|X)$ se puede considerar como transformación de X , entonces $\mathbb{E}(|\mathbb{E}(Y|X)|) = \sum_x |\mathbb{E}(Y|X = x)|p_1(x) \leq \sum_x \mathbb{E}(|Y||X = x)p_1(x) = \sum_x (\sum_y |y| \frac{p(x,y)}{p_1(x)})p_1(x) = \sum_x \sum_y |y|p(x,y) = \sum_y \sum_x |y|p(x,y) = \sum_y |y| \sum_x p(x,y) = \sum_y |y|p_2(y) = \mathbb{E}(|Y|) \leq +\infty$, luego es integrable.

Además, $\mathbb{E}(\mathbb{E}(Y|X)) = \sum_x \mathbb{E}(Y|X = x)p_1(x) = \sum_x (\sum_y y \frac{p(x,y)}{p_1(x)})p_1(x) = \sum_x \sum_y yp(x,y) = \sum_y \sum_x yp(x,y) = \sum_y y \sum_x p(x,y) = \sum_y yp_2(y) = \mathbb{E}(Y)$.

Caso continuo. Es análogo: $\mathbb{E}(|\mathbb{E}(Y|X)|) = \int_{\mathbb{R}} |\mathbb{E}(Y|X = x)|f_1(x)dx \leq \int_{\mathbb{R}} \mathbb{E}(|Y||X = x)f_1(x)dx = \int_{\mathbb{R}} (\int_{\mathbb{R}} |y| \frac{f(x,y)}{f_1(x)})dyf_1(x)dx = \int_{\mathbb{R}} \int_{\mathbb{R}} |y|f(x,y)dxdy = \int_{\mathbb{R}} |y| \int_{\mathbb{R}} f(x,y)dxdy = \int_{\mathbb{R}} |y|f_2(y)dy = \mathbb{E}(|Y|) \leq +\infty$, luego es integrable.

Además, $\mathbb{E}(\mathbb{E}(Y|X)) = \int_{\mathbb{R}} \mathbb{E}(Y|X = x)f_1(x)dx = \int_{\mathbb{R}} (\int_{\mathbb{R}} y \frac{f(x,y)}{f_1(x)})dyf_1(x)dx = \int_{\mathbb{R}} \int_{\mathbb{R}} yf(x,y)dxdy = \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x,y)dxdy = \int_{\mathbb{R}} yf_2(y)dy = \mathbb{E}(Y)$. \square

La esperanza no es el único parámetro a considerar de una distribución, dado que hay muchas distribuciones distintas de misma esperanza. Para distinguirlas existen otros parámetros:

Definición 30. Si X es aleatoria y $n \in \mathbb{N}$, se define su **momento de orden n** por $\alpha_n = \mathbb{E}(X^n)$.

Definición 31. El **momento de orden n centrado en c** de X aleatoria con $n \in \mathbb{N}$, $c \in \mathbb{R}$, es $\mu_{n,c} \equiv \alpha_{n,c} = \mathbb{E}((X - c)^n)$. Si μ es la esperanza de X , el **momento centrado de orden n** es $\mu_n = \mathbb{E}((X - \mu)^n) = \mathbb{E}((X - \mathbb{E}(X))^n)$.

Definición 32. La **varianza** de X es μ_2 . Se denota también $Var(X)$ o σ_X^2 . Es decir, $Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$.

La idea de la varianza es el valor esperado de la distancia de la variable a su esperanza. Indica de media cuanto de alejada está la variable de su esperanza. Como las unidades de la varianza definidas de esta forma corresponderían con las unidades de X al cuadrado, conviene definir:

Definición 33. La **desviación típica** de X es $\sigma_X = \sqrt{Var(X)} = \sqrt{\mathbb{E}((X - \mu)^2)}$.

Proposición 18 (Propiedades de la varianza). *Se verifican:*

1. $Var(X) \geq 0$.
2. $Var(X) = 0 \iff X = c$ para algún $c \in \mathbb{R}$.
3. $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. Además, por 1, $\mathbb{E}(X^2) \geq (\mathbb{E}(X))^2$.
4. $Var(X + Y) = Var(X) + Var(Y) + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))$. En particular, si $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, por ejemplo en variables independientes, entonces $Var(X) + Var(Y) = Var(X + Y)$.
5. $Var(aX + b) = a^2Var(X)$ para $a, b \in \mathbb{R}$.
6. $Var(X) = \min_{c \in \mathbb{R}} \alpha_{2,c}$.

Demostración. Para 1, como $(X - \mu)^2$ es no negativa, sabemos que su esperanza lo es. Para 2, si $X = c$, entonces $(X - \mu)^2 = (X - c)^2 = (c - c)^2 = 0$ luego su esperanza es 0. Por otro lado, si $X \neq c$, entonces $(X - \mu)^2$ no es nulo siempre, dado que hay por lo menos un valor de X que no es μ con probabilidad, luego su esperanza es positiva. Para 3, veamos que $Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2 + \mathbb{E}(X)^2 - 2X\mathbb{E}(X)) = \mathbb{E}(X^2) + \mathbb{E}(X)^2 - 2\mathbb{E}(X)\mathbb{E}(X)$ por linealidad y al ser $\mathbb{E}(X)$ constante, y operando sigue que $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

Para 4, se tiene $Var(X + Y) = \mathbb{E}((X + Y)^2) - \mathbb{E}(X + Y)^2 = \mathbb{E}(X^2) + \mathbb{E}(Y^2) + 2\mathbb{E}(XY) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) = Var(X) + Var(Y) + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))$ donde usamos una vez más linealidad, y el apartado 3.

Para 5, tenemos que $Var(aX + b) = \mathbb{E}((aX + b - \mathbb{E}(aX + b))^2) = \mathbb{E}((aX - a\mathbb{E}(X))^2) = a^2\mathbb{E}((X - \mathbb{E}(X))^2) = a^2Var(X)$.

Para 6, se tiene que $\alpha_{2,c} = \mathbb{E}((X - c)^2) = \mathbb{E}((X - \mathbb{E}(X) + \mathbb{E}(X) - c)^2) = \mathbb{E}((X - \mathbb{E}(X))^2) + \mathbb{E}((\mathbb{E}(X) - c)^2) + 2\mathbb{E}((X - \mathbb{E}(X))(\mathbb{E}(X) - c)) = \mathbb{E}((X - \mathbb{E}(X))^2) + \mathbb{E}((\mathbb{E}(X) - c)^2) + 2(\mathbb{E}(X) - c)\mathbb{E}((X - \mathbb{E}(X))) = \mathbb{E}((X - \mathbb{E}(X))^2) + \mathbb{E}((\mathbb{E}(X) - c)^2) + 2(\mathbb{E}(X) - c)(\mathbb{E}(X) - \mathbb{E}(X)) = \mathbb{E}((X - \mathbb{E}(X))^2) + \mathbb{E}((\mathbb{E}(X) - c)^2) \geq \mathbb{E}((X - \mathbb{E}(X))^2) = Var(X)$. \square

Observación 5 (Esperanzas y varianzas de distribuciones frecuentes.). A partir de la definición, se pueden determinar las siguientes esperanzas y varianzas:

1. **Bernoulli parámetro p.** $\mathbb{E}X = p$, $Var(X) = p(1 - p)$.
2. **Binomial parámetros n,p.** $\mathbb{E}X = np$, $Var(X) = np(1 - p)$.
3. **Geométrica parámetro p.** $\mathbb{E}X = \frac{1-p}{p}$, $Var(X) = \frac{1-p}{p^2}$.
4. **Binomial negativa parámetros t,p.** $\mathbb{E}X = t\frac{1-p}{p}$, $Var(X) = t\frac{1-p}{p^2}$.
5. **Poisson parámetro λ .** $\mathbb{E}X = \lambda$, $Var(X) = \lambda$.
6. **Uniforme en $[a, b]$.** $\mathbb{E}X = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$.
7. **Normal parámetros μ, σ .** $\mathbb{E}X = \mu$, $Var(X) = \sigma^2$.
8. **Exponencial parámetro λ .** $\mathbb{E}X = \lambda^{-1}$. $Var(X) = \lambda^{-2}$.

Definición 34. Una variable X es tipificada si $\mathbb{E}X = 0$ y $Var(X) = 1$.

Proposición 19. Dada X , la variable $X^* = \frac{X - \mathbb{E}X}{\sqrt{Var(X)}}$ es tipificada y adimensional, de tal modo que dos variables de dimensiones distintas se pueden tipificar de esta manera para compararlas.

Demostración. $\mathbb{E}X^* = \frac{\mathbb{E}X - \mathbb{E}X}{\sqrt{Var(X)}} = 0$. $Var(X^*) = \frac{Var(X)}{\sqrt{Var(X)}^2} = 1$. \square

Definición 35. Se define el **coeficiente de simetría** $CA(X) = \frac{\mu_3}{\sigma^3} = \frac{\mathbb{E}((X - \mathbb{E}X)^3)}{(\mathbb{E}((X - \mathbb{E}X)^2))^{\frac{3}{2}}}$, e indica si una distribución es simétrica alrededor de su media (si $CA = 0$), asimétrica por la derecha (si $CA > 0$) o por la izquierda (si $CA < 0$).

Definición 36. Se define el **coeficiente de apuntamiento o curtosis** $CAP(X) = \frac{\mu_4}{\sigma^4} = \frac{\mathbb{E}((X - \mathbb{E}X)^4)}{(\mathbb{E}((X - \mathbb{E}X)^2))^2}$, que indica la forma (convexidad) de la distribución. Si $CAP < 1,8$, tiene forma de U , si $CAP = 1,8$ es plana, si $CAP > 1,8$ es con forma de campana, siendo $CAP = 3$ la convexidad de la normal 0, 1.

Definición 37. Los **momentos de orden** $n_1 + n_2$ de un vector (X, Y) son los $\alpha_{n_1, n_2} = \mathbb{E}(X^{n_1}Y^{n_2})$. Los **momentos centrados** son $\mu_{n_1, n_2} = \mathbb{E}((X - \mathbb{E}X)^{n_1}(Y - \mathbb{E}Y)^{n_2})$.

Definición 38. El momento centrado de (X, Y) con $n_1 = n_2 = 2$ se denomina **covarianza** y se denota $Cov(X, Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y))$. Será mayor si las variables X, Y varían del mismo modo, y menor si varían de modo contrario.

Proposición 20. Estas son algunas propiedades de la covarianza:

1. $Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y$.
2. $Cov(aX + b, cY + d) = acCov(X, Y)$.

3. $Cov(X, X) = Var(X)$.
4. $Cov(X, Y) = Cov(Y, X)$.
5. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$.

Demostración. Para 1, tenemos que $\mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) = \mathbb{E}(XY - X\mathbb{E}Y - Y\mathbb{E}X + \mathbb{E}X\mathbb{E}Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y - \mathbb{E}Y\mathbb{E}X + \mathbb{E}X\mathbb{E}Y = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y$. Para 2, se tiene que $\mathbb{E}((aX + b - \mathbb{E}(aX + b))(cY + d - \mathbb{E}(cY + d))) = \mathbb{E}(a(X - \mathbb{E}X)c(Y - \mathbb{E}Y)) = ac\mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) = acCov(X, Y)$.

Para 3 y 4, sigue directamente de la definición o de la propiedad 1. Para 5, se trata de reescribir la propiedad 4 de la proposición 18 usando la propiedad 1. \square

Definición 39. Dos variables son **incorreladas** si $Cov(X, Y) = 0$, o, equivalentemente, si $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$.

Observación 6. 1. X, Y incorreladas $\implies Var(X + Y) = Var(X) + Var(Y)$.

2. X, Y independientes $\implies X, Y$ incorreladas. El recíproco no es cierto en general.

Proposición 21 (Cauchy-Schwarz). *Se tiene $\mathbb{E}|XY| \leq \sqrt{(\mathbb{E}(X^2))(\mathbb{E}(Y^2))}$*

Demostración. Sea $s \in \mathbb{R}$. Sabemos que $0 \leq \mathbb{E}((s|X| + |Y|)^2) = s^2\mathbb{E}X^2 + 2s\mathbb{E}|XY| + \mathbb{E}Y^2$. Obtenemos un polinomio en s que es no negativo. Si $\mathbb{E}X^2 = 0$, entonces es porque $X = 0$, y se tiene $\mathbb{E}|XY| = \mathbb{E}0 = \sqrt{0\mathbb{E}Y^2} = \sqrt{\mathbb{E}X^2\mathbb{E}Y^2}$ y se tiene. Si no, entonces $\mathbb{E}X^2 > 0$ y el polinomio, al ser no negativo, debe tener a lo sumo una raíz, con lo que $4(\mathbb{E}|XY|)^2 - 4\mathbb{E}X^2\mathbb{E}Y^2 \leq 0$, y despejando se tiene. \square

Proposición 22. $|Cov(X, Y)| \leq \sigma_X\sigma_Y$

Demostración. $|Cov(X, Y)| = |\mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y))| \leq \mathbb{E}(|(X - \mathbb{E}X)(Y - \mathbb{E}Y)|) \leq \sqrt{(\mathbb{E}((X - \mathbb{E}X)^2))(\mathbb{E}((Y - \mathbb{E}Y)^2))} = \sigma_X\sigma_Y$, usando Cauchy-Schwarz. \square

Esto permite normalizar la covarianza a un valor adimensional comprendido entre -1 y 1 .

Definición 40. El **coeficiente de correlacion de Pearson** de X, Y se define por $Corr(X, Y) = \rho_{XY} \frac{Cov(X, Y)}{\sigma_X\sigma_Y}$.

Proposición 23. *Se tiene:*

1. $\rho_{XY} \in [-1, 1]$.
2. $\rho_{XY} = 0 \iff X, Y$ incorreladas.
3. $\rho_{XY} = 1 \iff Y = aX + b$ para $a > 0$.
4. $\rho_{XY} = -1 \iff Y = aX + b$ para $a < 0$.

Demostración. Para 1, se tiene de la definición y del corolario de Cauchy-Schwarz. Para 2, si son incorreladas $Cov(X, Y) = 0$ luego $\rho_{XY} = 0$ por definición. Para 3, supongamos que $\rho_{XY} = 1$. Entonces, $Var(X^* - Y^*) = Var(X^*) + Var(Y^*) - 2Cov(X^*, Y^*) = 1 + 1 - 2Cov(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}) = 2 - 2Corr(X, Y) = 2 - 2 = 0$, por tanto $X^* - Y^* = c$ para una constante, y despejando, se tiene $Y = aX + b$ para a, b adecuadas, con $a = \frac{\sigma_Y}{\sigma_X} > 0$. Para el recíproco, si suponemos que $Y = aX + b$, entonces $Y^* - X^* = \frac{aX - a\mathbb{E}X}{\sigma_{aX+b}} - \frac{X - \mathbb{E}X}{\sigma_X} = \frac{X - \mathbb{E}X}{\sigma_X} - \frac{X - \mathbb{E}X}{\sigma_X} = 0$, luego usando la identidad anterior, $0 = 2 - 2Corr(X, Y)$ y por tanto $Corr(X, Y) = 1$.

Para 4, basta con llevar a cabo el mismo argumento, utilizando $Y^* + X^*$. \square

Observación 7. Se pueden calcular estos parámetros para dimensiones $d > 2$. Para ello, se obtienen todas las combinaciones posibles 2 a 2, dando lugar a una matriz de covarianzas, de correlaciones...

5. Funciones Características

A continuación estudiaremos la teoría de funciones características, que permite codificar distribuciones de probabilidad en ciertas funciones que gozan de ciertas propiedades que facilitan los cálculos en ciertas situaciones.

Definición 41. Una **variable aleatoria compleja** es $Z : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{C}$, con $Z(\omega) = X(\omega) + iY(\omega)$, para X, Y variables aleatorias reales, y como tal se le pueden aplicar todas las definiciones acerca de variables reales. En concreto, $\mathbb{E}Z = \mathbb{E}X + i\mathbb{E}Y$.

Definición 42. Sea X una variable aleatoria real. Se define su **función característica** por $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$ con $\varphi_X(t) = \mathbb{E}(e^{itX})$.

Proposición 24 (Observaciones sobre la función característica.). *Se tiene:*

1. φ_X está bien definida, es decir, e^{itX} es integrable $\forall t \in \mathbb{R}$.
2. $|\varphi(t)| \leq 1$, es decir, va sobre el disco unidad.
3. $\varphi(0) = 1$.
4. $\varphi(t) = \mathbb{E} \cos(tX) + i\mathbb{E} \sin(tX)$
5. Si $X = Y$ en distribución, entonces $\varphi_X = \varphi_Y$. El recíproco también es cierto.

Demostración. Para 1 y 2, se tiene $|\varphi(t)| = |\mathbb{E}e^{itX}| \leq \mathbb{E}|e^{itX}| = \mathbb{E}1 = 1$. 3, 4 y 5 siguen de la definición inmediatamente. El recíproco de 5 se dará posteriormente. \square

Proposición 25 (Propiedades de la función característica.). *Se tiene:*

1. $\varphi(-t) = \overline{\varphi(t)}$.
2. Si $Y = aX + b$, entonces $\varphi_Y(t) = e^{itb}\varphi_X(at)$.
3. Si X es simétrica, es decir, $X = -X$ en cuanto a distribución, entonces $\varphi_X(t)$ es real, con $\varphi_X(t) = \mathbb{E} \cos(tx)$, y, por 1, es par. (va sobre $[-1, 1]$).
4. Si X_1, \dots, X_n son independientes, entonces $\varphi_{\sum X_i}(t) = \prod \varphi_{X_i}(t)$.
5. φ_X es uniformemente continua.

Demostración. Para 1, $\varphi(-t) = \mathbb{E}e^{-itX} = \mathbb{E} \cos(-tX) + i\mathbb{E} \sin(-tX) = \mathbb{E} \cos(tX) - i\mathbb{E} \sin(tX) = \overline{\varphi(t)}$. Para 2, se tiene $\mathbb{E}e^{it(aX+b)} = e^{itb}\mathbb{E}e^{itaX} = e^{itb}\varphi_X(at)$. Para 3, se tiene que $i\mathbb{E} \sin(tX) = i \int_{-\infty}^{\infty} \sin(tX)f(x)dx$. Si X es simétrica, $f(x)$ es par, luego el integrando es impar y esa esperanza es nula, dando lugar a que $\varphi(t) = \mathbb{E} \cos(tX)$. Para 4, veamos que $\mathbb{E}e^{it\sum X_i} = \mathbb{E} \prod e^{itX_i} = \prod \mathbb{E}e^{itX_i} = \prod \varphi_{X_i}(t)$. La 5 se presenta sin prueba. \square

Observación 8 (Funciones características de distribuciones típicas). Siguen de la definición y las propiedades de independencia:

1. **Bernoulli parámetro p .** $\varphi(x) = pe^{it} + (1-p)$.
2. **Binomial parámetros n, p .** $\varphi(x) = (pe^{it} + (1-p))^n$
3. **Poisson parámetro λ .** $\varphi(x) = e^{\lambda(e^{it}-1)}$
4. **Geométrica parámetro p .** $\varphi(x) = \frac{p}{1-(1-p)e^{it}}$

5. **Binomial negativa** parámetros t, p . $\varphi(x) = \left(\frac{p}{1-(1-p)e^{it}}\right)^n$
6. **Uniforme** (a, b) . $\varphi(x) = \frac{e^{itb} - e^{ita}}{it(b-a)}$
7. **Normal** $(0, 1)$. $\varphi(x) = e^{-\frac{t^2}{2}}$
8. **Exponencial** parámetro λ . $\varphi(x) = \left(1 - \frac{it}{\lambda}\right)^{-1}$
9. **Gamma** parámetros α, β . $\varphi(x) = \left(1 - \frac{it}{\beta}\right)^{-\alpha}$

Podemos observar que, según derivamos con respecto a t la generatriz $\mathbb{E}e^{itX}$, si es posible, podemos ir obteniendo los momentos de X :

Teorema 6. Si $\mathbb{E}|X|^n < \infty$ con $n \in \mathbb{N}$, entonces $\varphi_X \in \mathcal{C}^n(\mathcal{D}_1)$ (disco unidad) y genera los momentos de X :

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}X^k$$

Esto nos permite, además, poner en desarrollo de Taylor: $\varphi_X(y) = \sum_{k=0}^n \frac{i^k \mathbb{E}X^k}{k!} t^k + o(t^n)$.

Teorema 7. Si $\exists \rho > 0$ con $\mathbb{E}e^{\rho|X|} < \infty$, entonces $\varphi_X = \sum_{k=0}^{\infty} \frac{i^k \mathbb{E}X^k}{k!} t^k$, siempre que $|t| \leq \rho$.

Esto permite, por ejemplo, obtener el desarrollo en serie de $\varphi_X(t)$ para la normal $(0, 1)$. Si lo comparamos, coeficiente a coeficiente, con el desarrollo en serie de $e^{-\frac{t^2}{2}}$, veremos que $\mathbb{E}X^k = 0$ si k es impar (lógico: es simétrica), o bien $\mathbb{E}X^{2k} = (2k-1)!!$.

Anteriormente vimos que φ_X determina unívocamente la distribución. Para mostrar eso, así como obtener una manera de hallar la distribución a partir de la función característica, se idean las **fórmulas de inversión**, que deshacen la función característica.

Proposición 26 (Inversión en el caso discreto). Si X es aleatoria con soporte en \mathbb{Z} , se tiene:

$$P(X = n) = \frac{1}{2\pi} \int_0^{2\pi} e^{-itn} \varphi_X(t) dt$$

Los límites de integración pueden ser valores cualesquiera que difieran en 2π .

Demostración. $\frac{1}{2\pi} \int_0^{2\pi} e^{-itn} \varphi_X(t) dt = \frac{1}{2\pi} \int_0^{2\pi} e^{-itn} \mathbb{E}e^{itX} dt = \frac{1}{2\pi} \int_0^{2\pi} e^{-itn} (\sum_{k \in \mathbb{Z}} e^{itk} P(X = k)) dt = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} P(X = k) \int_0^{2\pi} e^{it(k-n)} dt$. Para resolver esta integral, veamos que si $k = n$, entonces $\int_0^{2\pi} e^{it(k-n)} dt = \int_0^{2\pi} dt = 2\pi$, y, por otro lado, si $k \neq n$, se tiene $\int_0^{2\pi} e^{it(k-n)} dt = \frac{e^{it(k-n)}}{i(k-n)} \Big|_0^{2\pi} = 0$, luego, finalmente: $\frac{1}{2\pi} \int_0^{2\pi} e^{-itn} \varphi_X(t) dt = \frac{1}{2\pi} P(X = n) 2\pi = P(X = n)$ como se quería. \square

Teorema 8 (Teorema de Inversión de Lévy). Sea X una variable aleatoria.

$$\frac{P(X = a) + P(X = b)}{2} + P(a < X < b) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt$$

Teorema 9. Si $\varphi_X(t)$ es integrable, es decir, $\int_{-\infty}^{\infty} |\varphi_X(t)| dt < \infty$, entonces X es continua y

$$f(x) = \int_{-\infty}^{\infty} e^{-itx} |\varphi_X(t)| dt$$

De estos teoremas se deduce lo que ya habíamos mencionado:

Observación 9. Dos variables X, Y , con $\varphi_X = \varphi_Y$, son iguales en distribución.

Con esto se pueden obtener distribuciones de sumas de variables independientes, por ejemplo, multiplicando sus características e identificando la función resultante.

6. Teoremas límite

Queremos comprobar a qué convergen distintas variables aleatorias. En concreto, si repetimos numerosas veces un experimento, siendo cada resultado el dado por la variable X_i , nos gustaría que, si $S_n = \sum_1^n X_i$, el valor $\frac{S_n}{n}$ convergiera a un valor fijo. De esta manera se podrían calcular probabilidades de forma experimental, siguiendo el planteamiento frecuentista que se comentó en la primera sección.

Definición 43 (Convergencia en media cuadrática). Sean $\{X_i\}_1^\infty$ variables aleatorias. Se dice que **convergen en media cuadrática** a la variable X si $\lim_{n \rightarrow \infty} \mathbb{E}((X_n - X)^2) = 0$, y se denota $X_n \xrightarrow{m-2} X$. Intuitivamente, cada variable se va acercando más a X en tanto que difiere menos de X .

Teorema 10 (Ley de Grandes Números en Media Cuadrática). Sean $\{X_i\}_1^\infty$ variables incorreladas de igual media μ y varianza σ^2 . Entonces:

$$\frac{S_n}{n} \xrightarrow{m-2} \mu$$

Demostración. $\mathbb{E}(\frac{S_n}{n} - \mu)^2 = \frac{1}{n^2} \mathbb{E}(S_n - n\mu)^2 = \frac{1}{n^2} \mathbb{E}(S_n - \mathbb{E}S_n)^2 = \frac{n\sigma^2}{n^2} = \frac{1}{n}$ por incorrelación, que tiende a 0, como queríamos. \square

Las hipótesis se pueden rebajar a que $\sum_1^n \text{Var}(X_i) = O(n^\alpha)$, con $\alpha < 2$, sin necesidad de que sean las mismas, para que el último paso funcione.

Definición 44 (Convergencia en probabilidad). Sean $\{X_i\}_1^\infty$ variables aleatorias. Se dice que **convergen en probabilidad** a la variable X si $\forall \epsilon > 0$, $\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$, y se denota $X_n \xrightarrow{P} X$. Intuitivamente, la probabilidad de que X_n esté alejada de X una cantidad dada se va anulando si aumenta n .

Proposición 27 (Desigualdad de Markov). $\forall \epsilon > 0$, $\alpha > 0$, se tiene $P(|X| \geq \epsilon) \leq \frac{\mathbb{E}|X|^\alpha}{\epsilon^\alpha}$.

Demostración. Para el caso discreto, sea X discreta de soporte S . Sea $T = \{x : |x| \geq \epsilon\}$. Entonces, $\mathbb{E}|X|^\alpha = \sum_{x \in S} |x|^\alpha p(x) = \sum_{x \in S \cap T} |x|^\alpha p(x) + \sum_{x \notin S \cap T} |x|^\alpha p(x) \geq \sum_{x \in S \cap T} |x|^\alpha p(x) \geq \epsilon^\alpha \sum_{x \in S \cap T} p(x) = \epsilon^\alpha P(|X| \geq \epsilon)$, como se quería.

Para el caso continuo, sea X continua, entonces $\mathbb{E}|X|^\alpha = \int_{\mathbb{R}} |x|^\alpha f(x) dx = \int_T |x|^\alpha f(x) dx + \int_{T^c} |x|^\alpha f(x) dx \geq \int_T |x|^\alpha f(x) dx \geq \epsilon^\alpha \int_T f(x) dx = \epsilon^\alpha P(|X| \geq \epsilon)$, como se quería. \square

Observación 10 (Corolario. Desigualdad de Chebyshev). Si aplicamos la desigualdad de Markov a $\alpha = 2$, y $Y = X - \mathbb{E}X$, tenemos que: $P(|X - \mathbb{E}X| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$.

Teorema 11 (Ley débil de Grandes Números). Sean $\{X_i\}_1^\infty$ variables incorreladas de igual media μ y varianza σ^2 . Entonces:

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

Demostración. Es suficiente con ver que **convergencia en media cuadrática implica convergencia en probabilidad**. Para ello, veamos que $P(|X_n - X| \geq \epsilon) \leq \frac{\mathbb{E}(X_n - X)^2}{\epsilon^2} \rightarrow 0$ por la desigualdad de Markov y la hipótesis. \square

Definición 45 (Convergencia casi segura). Sean $\{X_i\}_1^\infty$ variables aleatorias. Se dice que **convergen casi seguramente** a la variable X si $P(X_n \rightarrow X) = P(\{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\}) = 1$, y se denota $X_n \xrightarrow{c.s.} X$. Intuitivamente, el valor de X_n en cada punto del espacio muestral converge al de X con n , salvo en un conjunto de probabilidad 0.

Teorema 12 (Ley fuerte de Grandes Números de Kolmogorov.). Sean $\{X_i\}_1^\infty$ variables independientes, de igual distribución y de media μ , entonces:

$$\frac{S_n}{n} \xrightarrow{c.s.} \mu$$

Este resultado indica que la media muestral de repetir muchas veces un experimento converge a su valor esperado.

Definición 46 (Convergencia en distribución). Sean $\{X_i\}_1^\infty$ variables aleatorias, de funciones de distribución F_1, F_2, \dots . Se dice que **convergen en distribución** a la variable X con función F si $\lim_{n \rightarrow \infty} F_n(x) = F(x) \forall x \in C_F = \{x \in \mathbb{R} \text{ con } F \text{ continua en } x\}$, y se denota $X_n \xrightarrow{D} X$. Intuitivamente, la probabilidad de que $X_n \leq a$ se parece cada vez más a la de que $X \leq a$.

Este modo es el más débil: convergencia en probabilidad implica este.

Veremos un resultado previo para demostrar el Teorema Central del Límite.

Teorema 13 (Teorema de continuidad de Lévy-Cramer). Si X_1, X_2, \dots son variables aleatorias con funciones características $\varphi_1, \varphi_2, \dots$, y $\exists \lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$, con φ continua en 0, entonces φ es a su vez función característica, de una variable X tal que $X_n \xrightarrow{D} X$.

Teorema 14 (Teorema Central Del Límite). Sean X_1, X_2, \dots independientes y equidistribuidas. Sean $\mathbb{E}X_i = \mu$, $\text{Var}(X_i) = \sigma^2 > 0$ (es decir, no constantes). Entonces, si $S_n = \sum_{i=1}^n X_i$, se tiene:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} Z \sim N(0, 1)$$

Es decir, $S_n^* \xrightarrow{D} Z \sim N(0, 1)$.

Demostración. Si las X_i son tipificadas, veamos que $\varphi_{\frac{S_n}{\sqrt{n}}} = (\varphi(t\sqrt{n}))^n$ al ser independientes e iguales, con φ la función característica de cada X_i . Ahora bien, $\varphi(t) = \varphi(0) + \varphi'(0)t + \frac{\varphi''(0)}{2}t^2 + o(t^2) = 1 - \frac{t^2}{2} + o(t^2)$, sabiendo que $\mathbb{E}X_i = 0$ y $\text{Var}(X_i) = 1$. Por tanto:

$$\varphi_{\frac{S_n}{\sqrt{n}}} = \left(1 - \frac{t^2}{2} + o\left(\frac{t^2}{n}\right)\right)^n = \left(1 + \frac{-\frac{t^2}{2} - \frac{o(\frac{t^2}{n})}{\frac{t^2}{n}}t^2}{n}\right)^n \rightarrow e^{-\frac{t^2}{2}}$$

luego por el teorema de Lévy-Cramer, sigue que $\frac{S_n}{\sqrt{n}} \xrightarrow{D} Z \sim N(0, 1)$. Si las variables no son tipificadas, veamos que $\frac{\sum^n X_i - n\mu}{\sqrt{n}\sigma} = \frac{\sum^n X_i^*}{\sqrt{n}}$ y estamos en el caso anterior. \square

El teorema indica que si alguna variable aleatoria es suma de muchas otras iguales, se aproxima a una normal. De hecho, si $X = \sum X_i$, entonces se aproxima a una $N(n\mu, \sigma\sqrt{n})$. Esto permite aproximar binomiales, poisson, gammas... que son sumas de muchas variables más pequeñas, siempre que sean suma de muchas (la velocidad de la convergencia depende de la distribución).