

INTEGRACIÓN ESTOCÁSTICA

Sesión 02/16

Repaso de la Teoría de la Probabilidad

Introducción

Aunque fue Gerolamo Cardano (1501-1576), en su libro *Liber de ludo aleae* (Libro de los juegos de azar), quien estableció un método para calcular probabilidades en el caso del lanzamiento de varios dados, su trabajo tuvo poca influencia, pues si bien fue escrito en 1564, se publicó en 1663.

Fue en el año 1654 cuando se plantearon algunos problemas cuyas soluciones condujeron al establecimiento de reglas generales para calcular probabilidades. Pareciera, por la referencias que hay, que en esa época los juegos con dados eran muy populares y se conocían algunas reglas que permitían a los jugadores conocer, de manera aproximada y siempre con incertidumbre, qué posibilidades tenían de ganar al realizar una apuesta. En ese año, Antoine Gombaud, conocido como el chevalier de Méré, le planteó a Pascal los siguientes dos problemas:

¿Cuántas veces se requiere lanzar un par de dados para que sea más favorable obtener por lo menos un par de seises?

¿Cómo deben repartirse las apuestas en un juego que se interrumpe? Por ejemplo, suponiendo que dos jugadores, A y B , apuestan 32 pesos cada uno en un juego que consiste de partidas consecutivas, en cada una de las cuales cada jugador tiene la misma posibilidad de ganarla, de tal manera que quien gane una partida acumula un punto y el juego es ganado por quien obtenga primero cuatro puntos, ¿cómo deben de repartirse las apuestas en caso de que el juego se interrumpa cuando el jugador A ha ganado dos puntos y B un punto?

Pascal y Fermat, sin conocer cada uno los métodos del otro, llegaron a las mismas soluciones.

Tres años más tarde, Christiaan Huygens resolvió también esos problemas, en su libro *De ratiociniis in ludo aleae* (Del razonamiento en los juegos de azar), siendo aquí donde se encuentra la primera teorización del Cálculo de Probabilidades.

Huygens tomó como punto de partida la siguiente hipótesis, la cual es básicamente una definición:

En un juego, la posibilidad que se tiene de ganar alguna cosa tiene un valor tal que, si se posee ese valor, se puede uno procurar la misma posibilidad en un juego equitativo.

Por un juego equitativo, Huygens entendía un juego que no va en detrimento de ninguno de los jugadores. Incluye el caso de un juego entre un número cualquiera de jugadores en el cual, o bien todos los jugadores tienen la misma posibilidad de ganar cierta cantidad, o bien cada uno de los jugadores tiene la misma posibilidad de ganar cierta cantidad que de perderla.

Podemos ver que la hipótesis de Huygens es básicamente la definición del concepto de esperanza de lo que se obtiene sobre lo que está en juego; es decir, la esperanza de una variable aleatoria definida como lo que obtiene el jugador.

En el año 1713 se publicó un libro de Jacques Bernoulli (1654-1705), titulado *Ars Conjectandi* (El arte de conjeturar). En ese libro, Bernoulli sentó las bases para el desarrollo posterior del Cálculo de Probabilidades. Formuló métodos generales para resolver problemas de probabilidad, dando así una formulación teórica más sólida que la de Huygens. Esto además de enfrentar el problema de probar que la frecuencia relativa con la que se presenta un evento se aproxima a la probabilidad del mismo a medida que el número de observaciones se hace más grande; en esta búsqueda demostró el primero de los teoremas límite del Cálculo de Probabilidades, el cual daría la pauta para una investigación que se prolongó por más de 200 años, hasta llegar a una formulación general que culminaría hacia el año 1930.

Bernoulli comenzó su libro con un análisis de los problemas resueltos por Huygens en su obra *De ratiociniis in ludo aleae*. Para esto tomó como base la hipótesis de Huygens, a la cual consideró como el **principio fundamental del arte de conjeturar**.

Huygens consideró otro problema con dados, el cual, al ser generalizado por Bernoulli, adquiriría una importancia central en el desarrollo del Cálculo de Probabilidades:

¿Cuántos dados se requieren lanzar para que sea más favorable obtener por lo menos dos seises?

Con relación a este problema, Bernoulli encontró que la probabilidad de obtener exactamente k seises en n lanzamientos de un dado es igual a $\binom{n}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$. Este resultado es de fundamental importancia en su trabajo pues con él se puede calcular la probabilidad de obtener una frecuencia de seises igual a $\frac{k}{n}$ en n lanzamientos de un dado y de aquí encontrar una relación entre la frecuencia de ocurrencia de un evento y su probabilidad, para obtener lo que se llama el Teorema de Bernoulli.

Otro problema de gran importancia que planteó y resolvió Huygens en su libro es el siguiente:

Dos jugadores, P y Q, juegan a lanzar alternadamente un par de dados. El juego comienza lanzando P el par de dados, con la condición de que si obtiene una suma igual a 6, gana el juego; en caso contrario, el juego continúa lanzando Q el par de dados, con la condición de que si obtiene una suma igual a 7, gana el juego; en caso contrario el juego continúa lanzando P el par de dados bajo las condiciones iniciales. ¿Cuáles son las respectivas probabilidades que cada jugador tiene de ganar el juego?

La importancia de ese problema radica en que se refiere a un experimento el cual admite una infinidad de posibles resultados, rebasando el marco de la definición clásica de probabilidad.

Bernoulli resolvió este problema estableciendo una serie geométrica para la probabilidad que cada jugador tiene de ganar el juego.

El método de Bernoulli fue retomado un poco más adelante, en el año 1718, por Abraham de Moivre en su libro *The Doctrine of Chances*, sin embargo, aunque aparentemente era conocido, no se utilizó durante el resto del siglo XVIII y todo el XIX, de manera que la propiedad de σ -aditividad de la función de probabilidad quedó relegada en la sistematización de Laplace, la cual perduró hasta principios del siglo XX.

Las soluciones de Bernoulli a los problemas resueltos por Huygens representaron un avance significativo en el camino de dotar al Cálculo de Probabilidades de una teoría que permitiera ir resolviendo problemas cada vez más complejos. Sin embargo, la aportación central de Bernoulli la encontramos en la última parte de su libro, donde planteó un problema de singular importancia, el cual sería la base para el desarrollo posterior de la teoría durante un periodo de más de 200 años. Fue a partir de ese resultado que el Cálculo de Probabilidades comenzó a ganarse un lugar importante dentro de la Matemática.

Escribió Bernoulli en su libro:

“Parece que, para hacer una hipótesis correcta sobre un hecho cualquiera, sólo es necesario calcular exactamente el número de casos posibles y, entonces, determinar las veces que puede posiblemente ocurrir un caso más que otro. Pero aquí, inmediatamente, surge nuestra mayor dificultad, porque este procedimiento se puede aplicar únicamente a muy pocos fenómenos; de hecho, casi exclusivamente a los relacionados con los juegos de azar ... pero hay otro camino que nos conduce a lo que buscamos, y nos permite, por lo menos, hallar a posteriori lo que no podemos determinar a priori, o sea, averiguando a partir de los resultados observados en numerosos casos similares.

Ha de suponerse, a este respecto, que, bajo condiciones similares, la ocurrencia (o no ocurrencia) de un suceso en el futuro seguirá la misma pauta que se ha observado para sucesos iguales en el pasado ... Lo que aún tiene que ser averiguado es si, cuando se aumenta el número de observaciones, también se sigue aumentando la probabilidad de que la proporción registrada de casos favorables y desfavorables se aproxime a la verdadera relación ... Este es el problema que he decidido publicar aquí, después de haber trabajado sobre él durante veinte años.”

Obsérvese que, en su razonamiento, Bernoulli supone que en los fenómenos aleatorios existe una regularidad, a saber, que la frecuencia relativa con la que se observa que ocurre un evento, se mantiene en el futuro como se observó en el pasado. A esta propiedad la llamaremos **principio de regularidad de la frecuencia relativa con la que ocurre un evento**.

El resultado al que hace referencia Bernoulli en su libro es el ahora llamado teorema de Bernoulli, el cual, utilizando terminología moderna, se puede enunciar como sigue:

Teorema de Bernoulli. Sea \mathcal{E} un experimento aleatorio que admite t posibles resultados equiprobables y A un evento relativo a ese experimento, para el cual hay r resultados que favorecen su ocurrencia. Consideremos un nuevo experimento aleatorio consistente en la repetición indefinida del experimento \mathcal{E} , de tal manera que cada repetición es independiente de las otras. Sea X_{nt} el número de veces que ocurre el evento A en las primeras nt repeticiones del experimento, entonces:

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{X_{nt}}{nt} - \frac{r}{t} \right| > \frac{1}{t} \right] = 0$$

Sin modificar lo esencial del razonamiento de Bernoulli para demostrar este resultado, se puede enunciar de la siguiente manera:

Sea \mathcal{E} un experimento aleatorio que admite t posibles resultados equiprobables y A un evento relativo a ese experimento, para el cual hay r resultados que favorecen su ocurrencia. Consideremos un nuevo experimento aleatorio consistente en la repetición indefinida del experimento \mathcal{E} , de tal manera que cada repetición es independiente de las otras. Sean X_n el número de veces que ocurre el evento A en las primeras n repeticiones del experimento y ε un número positivo arbitrario, entonces:

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{X_n}{n} - \frac{r}{t} \right| > \varepsilon \right] = 0$$

El resultado de Bernoulli hizo patente que en el modelo teórico que se estaba desarrollando se da efectivamente una correspondencia entre las probabilidades y las frecuencias con que se observan los posibles resultados de un suceso azaroso. Este resultado y otros del mismo tipo que le siguieron sentaron las bases teóricas para aplicar el Cálculo de Probabilidades al estudio de datos estadísticos. Muy pronto esta teoría comenzó a aplicarse al tratamiento de datos como los acumulados en tablas de mortalidad y natalidad.

El Cálculo de Probabilidades continuó desarrollándose y **a finales del siglo XIX era ya parte inseparable del cuerpo científico de la época**. Sin embargo, a pesar de su desarrollo, no había una definición satisfactoria de la probabilidad. Eso es lo que afirmaba Henri Poincaré en la primera frase del capítulo I de su libro *Calcul des Probabilités*, publicado en 1896:

“No se puede dar una definición satisfactoria de la probabilidad.”

La invención de la Teoría de la Medida a principios del siglo XX vino a resolver el problema de la fundamentación del Cálculo de Probabilidades, surgiendo así un cuerpo teórico, puramente matemático, el cual constituye lo que ahora podemos llamar la Teoría de la Probabilidad.

Actualmente, para fines de aplicar el Cálculo de Probabilidades, se introduce el concepto de **experimento aleatorio**, donde un experimento se considera como cualquier proceso que conduzca a un resultado. Algunos experimentos pueden consistir simplemente en la observación de un determinado fenómeno natural y en ir anotando algunas de las características que observamos; por ejemplo, la medición día con día del crecimiento de una planta. En otros experimentos están involucradas acciones nuestras, además de la observación, como en el caso del lanzamiento de un dado.

Un experimento consta de dos partes, por un lado tenemos la descripción del proceso que estamos considerando, especificando las condiciones bajo las cuales se realiza, y por otro lado tenemos lo que consideramos como su resultado.

Se dice que un experimento es aleatorio cuando al realizarse, bajo las condiciones que se indiquen, su resultado puede ser cualquiera de un conjunto de resultados posibles. Cabe aclarar que la posibilidad de distintos resultados puede provenir de que las condiciones que se especifican para la realización del experimento incluyen cierta arbitrariedad.

En la formulación moderna, para modelar matemáticamente un experimento aleatorio se busca construir un espacio de medida $(\Omega, \mathfrak{F}, P)$, donde Ω , llamado el espacio muestral del experimento, es el conjunto de sus posibles resultados, \mathfrak{F} es una σ -álgebra de subconjuntos de Ω , los cuales representan a los eventos, y la medida P , definida sobre \mathfrak{F} , tiene la característica de asignarle a Ω el valor 1. Cada elemento de \mathfrak{F} es entonces un subconjunto del total de posibles resultados del experimento aleatorio en consideración y lo que nos interesa obtener es la probabilidad de que ocurra alguno de los posibles resultados que pertenecen a ese subconjunto. En general, la medida P se construye mediante un proceso de extensión: se comienza por asignar probabilidades a algunos eventos y después, utilizando las propiedades de P , se busca extenderla hasta abarcar el mayor número posible de subconjuntos de Ω . De esta forma, el modelo matemático para el estudio del experimento consta de 3 elementos, el primero es el conjunto Ω , el segundo una familia de subconjuntos de Ω y el tercero una medida, que llamaremos medida de probabilidad, la cual asocia a cada elemento A de \mathfrak{F} un número real que designa la probabilidad de que ocurra alguno de los posibles resultados que pertenecen a A .

Este modelo se fue construyendo durante los primeros 33 años del siglo XX, buscando axiomatizar el Cálculo de Probabilidades. Durante esos 33 años se estableció y consolidó su vínculo con la Teoría de la Medida, la cual surgió en los primeros años del siglo con los trabajos de Émile Borel y Henri Lebesgue. **Es con su formulación axiomática que podemos hablar de una teoría matemática de la probabilidad, ya que de esta forma se tiene un cuerpo teórico, independiente de los problemas reales específicos que se tratan con el Cálculo de Probabilidades.**

Espacios de probabilidad

Definición 1 (Espacio de probabilidad). *Llamaremos espacio de probabilidad a una terna $(\Omega, \mathfrak{S}, P)$, donde Ω es un conjunto, \mathfrak{S} una σ -álgebra de subconjuntos de Ω y P una medida sobre \mathfrak{S} tal que $P(\Omega) = 1$, a la cual llamaremos medida de probabilidad. A Ω lo llamaremos el espacio muestral, a los elementos de \mathfrak{S} eventos y a la medida P de un evento A la probabilidad de A .*

Considerando que cualquier espacio de medida se puede completar, asumiremos que cualquier medida de probabilidad con la que trabajemos es completa.

En lo que sigue asumimos que tenemos definido un espacio de probabilidad completo $(\Omega, \mathfrak{S}, P)$.

Definición 2 (Probabilidad condicional). *Sean A y B dos eventos y supongamos $P(A) > 0$, se define la probabilidad condicional de B , dada la ocurrencia de A , $P(B|A)$, mediante la fórmula:*

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

La probabilidad condicional dado un evento A es una nueva medida de probabilidad, la cual asigna el valor 1 a A y el valor 0 a A^c ; es decir, se trata de una medida de probabilidad concentrada en A . Se puede decir que al tomar probabilidades condicionales dado un evento A , el espacio muestral se reduce, convirtiéndose A en un nuevo espacio muestral.

Definición 3 (Independencia de eventos). *Diremos que los eventos de una familia no vacía cualquiera $\{A_\gamma\}$, finita o infinita, son estocásticamente independientes si dada cualquier subcolección finita de ellos, $A_{\gamma_1}, \dots, A_{\gamma_n}$, donde $n \in \mathbb{N}$, se tiene:*

$$P(A_{\gamma_1} \cap \dots \cap A_{\gamma_m}) = P(A_{\gamma_1}) \cdots P(A_{\gamma_m})$$

Definición 4 (Eventos mutuamente excluyentes). *Diremos que los eventos de una familia no vacía cualquiera $\{A_\gamma\}$, finita o infinita, son mutuamente excluyentes si cualquier par de ellos son conjuntos ajenos.*

Proposición 1. *Sea $\{A_1, \dots, A_n\}$ una familia de eventos independientes, donde $n \geq 2$, y reemplacemos uno de ellos, cualquiera, por su complemento, entonces los eventos de la nueva familia siguen siendo independientes.*

Demostración

Reordenemos los eventos A_1, \dots, A_n de tal manera que el evento que reemplazamos por su complemento sea A_n .

Sea $\{n_1, \dots, n_k\}$ cualquier subconjunto no vacío del conjunto $\{1, \dots, n\}$, donde los elementos n_1, \dots, n_k están ordenados del menor al mayor y $k \geq 2$. Entonces:

Si $n \notin \{n_1, \dots, n_k\}$, como los eventos de la familia $\{A_1, \dots, A_n\}$ son independientes, se tiene:

$$P(A_{n_1} \cap \dots \cap A_{n_j}) = P(A_{n_1}) \dots P(A_{n_j})$$

Si $n \in \{n_1, \dots, n_k\}$, entonces $n_k = n$ y se tiene:

$$\begin{aligned} & P(A_{n_1} \cap \dots \cap A_{n_{k-1}} \cap A_n^c) \\ &= P(A_{n_1} \cap \dots \cap A_{n_{k-1}} - A_{n_1} \cap \dots \cap A_{n_{k-1}} \cap A_n) \\ &= P(A_{n_1} \cap \dots \cap A_{n_{k-1}}) - P(A_{n_1} \cap \dots \cap A_{n_{k-1}} \cap A_n) \\ &= P(A_{n_1}) \dots P(A_{n_{k-1}}) - P(A_{n_1}) \dots P(A_{n_{k-1}}) P(A_n) \\ &= P(A_{n_1}) \dots P(A_{n_{k-1}}) [1 - P(A_n)] = P(A_{n_1}) \dots P(A_{n_{k-1}}) P(A_n^c). \end{aligned}$$

Así que los eventos de la familia $\{A_1, \dots, A_{n-1}, A_n^c\}$ son independientes. ■

Corolario 1. Sea $\{A_1, \dots, A_n\}$ una familia de eventos independientes y U un subconjunto del conjunto $\{1, \dots, n\}$. Para cada $j \in \{1, \dots, n\}$, definamos:

$$B_j = \begin{cases} A_j & \text{si } j \in U \\ A_j^c & \text{si } j \notin U \end{cases}$$

Entonces, los eventos de la familia $\{B_1, \dots, B_n\}$ son independientes.

Corolario 2. Sean $\{A_\gamma\}_{\gamma \in \Gamma}$ una familia de eventos independientes y T un subconjunto de Γ . Para cada $\gamma \in \Gamma$, definamos:

$$B_\gamma = \begin{cases} A_\gamma & \text{si } \gamma \in T \\ A_\gamma^c & \text{si } \gamma \notin T \end{cases}$$

Entonces, los eventos de la familia $\{B_\gamma\}_{\gamma \in \Gamma}$ son independientes.

Proposición 2 (Regla de la probabilidad total). Sean B un evento cualquiera y A_1, A_2, \dots una colección finita o infinita numerable de eventos de probabilidad positiva, mutuamente excluyentes y tales que $P(\bigcup_n A_n) = 1$, entonces:

$$P(B) = \sum_n P(B | A_n) P(A_n)$$

Teorema 1 (Lema de Borel-Cantelli - 1a. parte). *Sea A_1, A_2, \dots una sucesión de eventos tales que $\sum_{n=1}^{\infty} P(A_n) < \infty$ y sea:*

$$A = \{\omega \in \Omega : \omega \in A_n \text{ para una infinidad de valores de } n\}$$

Entonces $P(A) = 0$.

Demostración

Para cada $m \in \mathbb{N}$, sea $B_m = \bigcup_{n=m}^{\infty} A_n$. Entonces la sucesión de eventos B_m es decreciente y $A = \bigcap_{m=1}^{\infty} B_m$, así que:

$$P(A) = P\left[\bigcap_{m=1}^{\infty} B_m\right] = \lim_{m \rightarrow \infty} P[B_m]$$

Pero, $P(B_m) = P\left(\bigcup_{n=m}^{\infty} A_n\right) \leq \sum_{n=m}^{\infty} P(A_n)$.

Por lo tanto, $P(A) \leq \lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} P(A_n) = 0$.

■

Lema 1. Sea $(p_n)_{n \in \mathbb{N}}$ una sucesión de números reales en el intervalo $[0, 1]$ tal que la serie $\sum_{n=1}^{\infty} p_n$ es divergente, entonces $\lim_{n \rightarrow \infty} \prod_{j=1}^n (1 - p_j) = 0$.

Por el teorema del valor medio, se tiene, para cada $x \in [0, 1)$:

$$\ln(1 - x) = -\frac{x}{1 - \theta x}, \text{ donde } \theta \in (0, 1)$$

De manera que, para cualquier $x \in [0, 1)$, $\ln(1 - x) \leq -x$, es decir, $1 - x \leq e^{-x}$, resultado que también es válido para $x = 1$.

En particular, $\prod_{j=1}^n (1 - p_j) < e^{-\sum_{j=1}^n p_j}$, así que:

$$\lim_{n \rightarrow \infty} \prod_{j=1}^n (1 - p_j) \leq \lim_{n \rightarrow \infty} e^{-\sum_{j=1}^n p_j} = 0$$

■

Teorema 2 (Lema de Borel-Cantelli - 2a. parte). Sea A_1, A_2, \dots una sucesión de eventos independientes tales que $\sum_{n=1}^{\infty} P(A_n) = \infty$ y sea:

$$A = \{\omega \in \Omega : \omega \in A_n \text{ para una infinidad de valores de } n\}.$$

Entonces $P(A) = 1$.

Demostración

Para cada $m \in \mathbb{N}$, sea $B_m = \bigcup_{n=m}^{\infty} A_n$. Entonces la sucesión de eventos B_m^c es decreciente y $A^c = \bigcap_{m=1}^{\infty} B_m^c$, así que:

$$P(A^c) = P\left[\bigcap_{m=1}^{\infty} B_m^c\right] = \lim_{m \rightarrow \infty} P[B_m^c]$$

Pero, $B_m^c = \bigcap_{n=m}^{\infty} A_n^c \subset \bigcap_{n=m}^{m+k} A_n^c$ para cualquier $k \in \mathbb{N}$, así que, $P(B_m^c) \leq \prod_{n=m}^{m+k} [1 - P(A_n)]$ para cualquier $k \in \mathbb{N}$. Por lo tanto:

$$P(B_m^c) \leq \lim_{k \rightarrow \infty} \prod_{n=m}^{m+k} [1 - P(A_n)] = 0$$

Se concluye entonces que:

$$P(A) = 1 - P(A^c) = 1 - \lim_{m \rightarrow \infty} P[B_m^c] = 1$$

■

Variables Aleatorias

Definición 5. *Llamaremos variable aleatoria real a cualquier función medible de (Ω, \mathfrak{S}) en $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. Una variable aleatoria con valores en el conjunto de números reales extendido será una función medible de (Ω, \mathfrak{S}) en $(\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$. Un vector aleatorio real será cualquier función medible de (Ω, \mathfrak{S}) en $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$, donde $n \in \{2, 3, \dots\}$.*

Obviamente, una variable aleatoria real puede considerarse también como una función de (Ω, \mathfrak{S}) en $(\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$, y esta función es medible.

Dado un conjunto finito de variables aleatorias, X_1, X_2, \dots, X_n , con valores en $\overline{\mathbb{R}}^n$, la función $(X_1, X_2, \dots, X_n) : \Omega \rightarrow \overline{\mathbb{R}}^n$ definida por:

$$(X_1, X_2, \dots, X_n)(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$$

es medible. A una función así definida la llamaremos vector aleatorio con valores en $\overline{\mathbb{R}}^n$. También usaremos la notación \overline{X} para un vector aleatorio (X_1, X_2, \dots, X_n) .

A menos que se indique otra cosa, una variable aleatoria (resp. vector aleatorio con n componentes) será considerada con valores en $\overline{\mathbb{R}}$ (resp. con valores en $\overline{\mathbb{R}}^n$).

Dada una variable aleatoria $X : \Omega \rightarrow \overline{\mathbb{R}}$ y $B \in \mathfrak{B}(\overline{\mathbb{R}})$, la notación $[X \in B]$ será una manera abreviada de representar al conjunto $\{\omega \in \Omega : X(\omega) \in B\}$. Si $a, b \in \overline{\mathbb{R}}$, las notaciones $[a < X < b]$, $[a \leq X \leq b]$, $[a \leq X < b]$, $[a < X \leq b]$, $[X < b]$, $[X \leq b]$, $[X > a]$ y $[X \geq a]$ se entenderán en un sentido similar. Para el caso de un vector aleatorio utilizaremos una notación análoga.

Si (X_1, X_2, \dots, X_n) es un vector aleatorio con valores en $\overline{\mathbb{R}}^n$ y B_1, B_2, \dots, B_n son conjuntos borelianos en $\overline{\mathbb{R}}$, denotaremos por $[X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n]$ al conjunto $\bigcap_{k=1}^n [X_k \in B_k]$.

Definición 6. *Si $(\mathbb{E}, \mathcal{E})$ es un espacio medible y $Z : \Omega \rightarrow \mathbb{E}$ es una función medible, la función $\mu_Z : \mathcal{E} \rightarrow \mathbb{R}$ definida por $\mu_Z(E) = P[Z \in E]$ será llamada la proyección de P bajo Z .*

Definición 7. *Sea X una variable aleatoria. La proyección de P bajo X será llamada la distribución de la variable aleatoria X . Si (X_1, X_2, \dots, X_n) es un vector aleatorio, la proyección de P bajo (X_1, X_2, \dots, X_n) será denotada por $\mu_{X_1, X_2, \dots, X_n}$ y la llamaremos la distribución del vector aleatorio (X_1, X_2, \dots, X_n) .*

RESULTADOS

1. Si $(\mathbb{E}, \mathcal{E})$ es un espacio medible y $Z : \Omega \rightarrow \mathbb{E}$ es una función medible, la proyección de P bajo Z es una medida de probabilidad sobre $(\mathbb{E}, \mathcal{E})$.

2. Si X es una variable aleatoria, μ_X su distribución y $f : (\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}), \mu_X) \mapsto (\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$ una función medible, no negativa o integrable, se tiene:

$$\int_{\mathfrak{F}} f(X) dP = \int_{\overline{\mathbb{R}}} f d\mu_X$$

3. Si X es una variable aleatoria real y la consideramos como una función de (Ω, \mathfrak{S}) en $(\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$, entonces X es una variable aleatoria con valores en el conjunto de números reales extendido y su distribución μ_X , aunque está definida sobre $\mathfrak{B}(\overline{\mathbb{R}})$, está concentrada en \mathbb{R} ya que $\mu_X(\{-\infty, \infty\}) = 0$.

4. Si (X_1, X_2, \dots, X_n) es un vector aleatorio con valores en $\overline{\mathbb{R}}^n$, $\mu_{X_1, X_2, \dots, X_n}$ su distribución y $f : (\overline{\mathbb{R}}^n, \mathfrak{B}(\overline{\mathbb{R}}^n), \mu_{X_1, X_2, \dots, X_n}) \mapsto (\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$ una función medible, no negativa o integrable, se tiene:

$$\int_{\mathfrak{F}} f(X_1, X_2, \dots, X_n) dP = \int_{\overline{\mathbb{R}}^n} f d\mu_{X_1, X_2, \dots, X_n}$$

Definición 8 (σ álgebra generada por una familia de funciones). Sea $(\mathbb{E}, \mathcal{E})$ un espacio medible y \mathbb{F} un conjunto cualquiera. Dada una colección de funciones:

$$\mathcal{H} = \{f_\gamma : \mathbb{F} \rightarrow (\mathbb{E}, \mathcal{E}) : \gamma \in \Gamma\}$$

donde Γ es un conjunto de índices cualquiera, se define la σ -álgebra generada por \mathcal{H} como la más pequeña σ -álgebra de subconjuntos de \mathbb{F} tal que toda función $f \in \mathcal{H}$ es medible. Denotaremos a esta σ -álgebra por $\sigma(\mathcal{H})$ o por $\sigma(\{f_\gamma : \gamma \in \Gamma\})$.

Obsérvese que si $f : \mathbb{F} \rightarrow (\mathbb{E}, \mathcal{E})$ es cualquier función, la familia de conjuntos $\{f^{-1}(B) : B \in \mathcal{E}\}$ es una σ -álgebra de subconjuntos de \mathbb{F} . Sin embargo, si $\{f_\gamma : \mathbb{F} \rightarrow (\mathbb{E}, \mathcal{E}) : \gamma \in \Gamma\}$ es una colección de funciones, la familia de conjuntos $\{f_\gamma^{-1}(B) : \gamma \in \Gamma \text{ y } B \in \mathcal{E}\}$ no es, en general, una σ -álgebra, pero la σ -álgebra generada por esa familia de conjuntos es la σ -álgebra generada por la familia de funciones $\{f_\gamma : \mathbb{F} \rightarrow (\mathbb{E}, \mathcal{E}) : \gamma \in \Gamma\}$.

Por otra parte, si f_1, f_2, \dots, f_n son funciones de \mathbb{F} en $(\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$ y definimos $f : \mathbb{F} \rightarrow (\overline{\mathbb{R}}^n, \mathfrak{B}(\overline{\mathbb{R}}^n))$ mediante la relación:

$$f(x) = (f_1(x), f_2(x), \dots, f_n(x))$$

sabemos que si \mathfrak{F} es una σ -álgebra de subconjuntos de \mathbb{F} , entonces f es medible si y sólo si f_k es medible para cualquier $k \in \{1, 2, \dots, n\}$. Por lo tanto:

$$\{f^{-1}(B) : B \in \mathfrak{B}(\overline{\mathbb{R}}^n)\} = \sigma(f) = \sigma(\{f_1, f_2, \dots, f_n\})$$

Proposición 3. Sean Y_1, \dots, Y_n n variables aleatorias y $Z : \Omega \mapsto \overline{\mathbb{R}}$ una función $\sigma(Y_1, \dots, Y_n)$ -medible. Entonces, existe una función boreliana $h : \overline{\mathbb{R}}^n \mapsto \overline{\mathbb{R}}$ tal que $Z = h(Y_1, \dots, Y_n)$.

Demostración

Si $Z = I_E$, donde $E \in \sigma(Y_1, \dots, Y_n)$, entonces existe un boreliano $B \subset \overline{\mathbb{R}}^n$ tal que $Z = I_B(Y_1, \dots, Y_n)$. Por lo tanto, se tiene el resultado para el caso de una función simple Z $\sigma(Y_1, \dots, Y_n)$ -medible.

Si Z es una variable aleatoria no negativa, sea $(Z_n)_{n \in \mathbb{N}}$ una sucesión no decreciente de funciones simples no negativas tales que $Z = \lim_{n \rightarrow \infty} Z_n$ y, para cada $n \in \mathbb{N}$, sea $h_n : \overline{\mathbb{R}}^n \mapsto \overline{\mathbb{R}}$ una función boreliana no negativa tal que $Z_n = h_n(Y_1, \dots, Y_n)$.

Sea $D = \{x \in \overline{\mathbb{R}}^n : \lim_{n \rightarrow \infty} h_n(x) \text{ existe}\}$. Entonces D es un conjunto boreliano de $\overline{\mathbb{R}}^n$ y contiene a la imagen de Ω bajo la función (Y_1, \dots, Y_n) . Definamos la función $h : \overline{\mathbb{R}}^n \mapsto \overline{\mathbb{R}}$ de la siguiente manera:

$$h(x) = \begin{cases} \lim_{n \rightarrow \infty} h_n(x) & \text{si } x \in D \\ 0 & \text{en otro caso} \end{cases}$$

h es entonces una función boreliana y $Z = h(Y_1, \dots, Y_n)$. ■

Corolario 3. Sean Y_1, \dots, Y_n n variables aleatorias reales y $Z : \Omega \mapsto \mathbb{R}$ una función $\sigma(Y_1, \dots, Y_n)$ -medible. Entonces, existe una función boreliana $h : \mathbb{R}^n \mapsto \mathbb{R}$ tal que $Z = h(Y_1, \dots, Y_n)$.

Independencia de variables aleatorias

Definición 9. Diremos que las variables aleatorias de una familia no vacía cualquiera $\{X_\gamma\}$, finita o infinita, son independientes, si dada cualquier subcolección finita de ellas, $X_{\gamma_1}, \dots, X_{\gamma_n}$ y cualquier colección de subconjuntos borelianos de $\overline{\mathbb{R}}$, A_1, \dots, A_n , donde $n \in \mathbb{N}$, se tiene:

$$P [X_{\gamma_1} \in A_1, \dots, X_{\gamma_n} \in A_n] = P [X_{\gamma_1} \in A_1] \cdots P [X_{\gamma_n} \in A_n]$$

RESULTADOS

1. n variables aleatorias X_1, \dots, X_n , son independientes si y sólo si para cualquier colección de subconjuntos borelianos de $\overline{\mathbb{R}}$, A_1, \dots, A_n , se tiene:

$$P [X_1 \in A_1, \dots, X_n \in A_n] = P [X_1 \in A_1] \cdots P [X_n \in A_n]$$

2. Las variables aleatorias de una familia infinita numerable, $\{X_1, X_2, \dots\}$, son independientes si y sólo si para cualquier $n \in \mathbb{N}$ y cualquier colección de subconjuntos borelianos en $\overline{\mathbb{R}}$, A_1, \dots, A_n , se tiene:

$$P [X_1 \in A_1, \dots, X_n \in A_n] = P [X_1 \in A_1] \cdots P [X_n \in A_n]$$

3. Sean X_1, \dots, X_n n variables aleatorias independientes y f_1, \dots, f_n n funciones borelianas de $\overline{\mathbb{R}}$ en $\overline{\mathbb{R}}$. Entonces las variables aleatorias $f_1(X_1), \dots, f_n(X_n)$ son independientes.

4. Sean $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m}$ $n + m$ variables aleatorias independientes y $f : \overline{\mathbb{R}}^n \mapsto \overline{\mathbb{R}}$ y $g : \overline{\mathbb{R}}^m \mapsto \overline{\mathbb{R}}$ dos funciones borelianas. Entonces, las variables aleatorias $f(X_1, \dots, X_n)$ y $g(X_{n+1}, \dots, X_{n+m})$ son independientes.

Convergencia de variables aleatorias

En lo que se refiere a la convergencia de una sucesión de variables aleatorias, hay algunos cambios en la terminología. En el contexto de la Teoría de la Probabilidad, la convergencia casi en todas partes será denominada **convergencia casi segura** y a la convergencia en medida la llamaremos **convergencia en probabilidad**.

Es decir, tenemos las siguientes definiciones:

Definición 10 (Convergencia casi segura). Diremos que una sucesión de variables aleatorias $(X_n)_{n \in \mathbb{N}}$ converge casi seguramente a una variable aleatoria X si $P[\lim_{n \rightarrow \infty} X_n = X] = 1$. Si éste es el caso, se escribirá $X_n \xrightarrow{c.s.} X$.

Definición 11 (Convergencia en probabilidad). Diremos que una sucesión de variables aleatorias $(X_n)_{n \in \mathbb{N}}$ converge en probabilidad a una variable aleatoria X si:

$$\lim_{n \rightarrow \infty} P[|X_n - X| > \varepsilon] = 0$$

para cualquier $\varepsilon > 0$. Si éste es el caso, se escribirá $X_n \xrightarrow{P} X$.

Las siguientes propiedades se siguen inmediatamente de las correspondientes propiedades para las sucesiones de números reales:

- i. Si una sucesión de variables aleatorias $(X_n)_{n \in \mathbb{N}}$ converge casi seguramente a X , entonces cualquier subsucesión de $(X_n)_{n \in \mathbb{N}}$ también converge casi seguramente a X .
- ii. Si $(X_n)_{n \in \mathbb{N}}$ es una sucesión de variables aleatorias tal que $X_n \xrightarrow{c.s.} X$ y $X_n \xrightarrow{c.s.} Y$, entonces $P[X = Y] = 1$.
- iii. Si c es una constante y $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias tal que $X_n \xrightarrow{c.s.} X$, entonces $cX_n \xrightarrow{c.s.} cX$.
- iv. Si $(X_n)_{n \in \mathbb{N}}$ y $(Y_n)_{n \in \mathbb{N}}$ son dos sucesiones de variables aleatorias tales que $X_n \xrightarrow{c.s.} X$ y $Y_n \xrightarrow{c.s.} Y$, entonces $X_n + Y_n \xrightarrow{c.s.} X + Y$ y $X_n Y_n \xrightarrow{c.s.} XY$.

Los resultados que siguen, con relación a la convergencia casi en todas partes, se entienden mejor si se tienen en mente los siguientes conceptos:

Dada una sucesión A_1, A_2, \dots de subconjuntos de un conjunto \mathbb{F} , se define el límite inferior (lím inf) y el límite superior (lím sup) de esa sucesión de la siguiente manera:

$$\begin{aligned}\text{lím inf } A_n &= \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m \\ \text{lím sup } A_n &= \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\end{aligned}$$

Obsérvese que $\text{lím inf } A_n$ está formado por todos los elementos $x \in \mathbb{F}$ para los cuales existe $N \in \mathbb{N}$ tal que $x \in A_n$ para cualquier $n \geq N$, mientras que $\text{lím sup } A_n$ está formado por todos los elementos $x \in \mathbb{F}$ que pertenecen a una infinidad de conjuntos de la sucesión. Así que se tiene siempre $\text{lím inf } A_n \subset \text{lím sup } A_n$.

Teorema 3. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias. Entonces $X_n \xrightarrow{c.s.} 0$ si y sólo si para cualquier $\varepsilon > 0$, se tiene:*

$$P[\text{lím sup } \{\omega \in \Omega : |X_n(\omega)| > \varepsilon\}] = 0$$

Demostración

Supongamos primero que $\text{lím}_{n \rightarrow \infty} X_n = 0$ casi seguramente. Entonces existe un conjunto $\Omega_0 \subset \Omega$ de probabilidad 0 tal que si $\omega \in \Omega_0^c$ entonces $\text{lím}_{n \rightarrow \infty} X_n(\omega) = 0$. Así que, dado $\omega_0 \in \Omega_0^c$ y $\varepsilon > 0$, existe $N \in \mathbb{N}$ tal que $|X_n(\omega_0)| \leq \varepsilon$ para cualquier $n \geq N$; esto significa que:

$$\omega_0 \in \bigcap_{n=N}^{\infty} \{\omega \in \Omega : |X_n(\omega)| \leq \varepsilon\}$$

Dicho de otra forma, si $\omega_0 \in \Omega_0^c$, entonces, dada cualquier $\varepsilon > 0$:

$$\omega_0 \in \bigcup_{m=1}^{\infty} [\bigcap_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)| \leq \varepsilon\}]$$

Así que:

$$\Omega_0^c \subset \bigcup_{m=1}^{\infty} [\bigcap_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)| \leq \varepsilon\}]$$

Por lo tanto:

$$P(\bigcap_{m=1}^{\infty} [\bigcup_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)| > \varepsilon\}]) = 0$$

Inversamente, supongamos que, para cualquier $\varepsilon > 0$, se tiene:

$$P(\bigcap_{m=1}^{\infty} [\bigcup_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)| > \varepsilon\}]) = 0$$

Para cada $r \in \mathbb{N}$, sea:

$$B_r = \bigcap_{m=1}^{\infty} \left[\bigcup_{n=m}^{\infty} \left\{ \omega \in \Omega : |X_n(\omega)| > \frac{1}{r} \right\} \right]$$

Se tiene $P(B_r) = 0$ para cualquier $r \in \mathbb{N}$ y la sucesión de conjuntos B_1, B_2, \dots es creciente, así que:

$$P\left(\bigcup_{r=1}^{\infty} B_r\right) = \lim_{r \rightarrow \infty} P(B_r) = 0$$

Pero:

$$B_r^c = \left\{ \omega \in \Omega : \text{Existe } N(\omega) \in \mathbb{N} \text{ tal que } |X_n(\omega)| \leq \frac{1}{r} \text{ para cualquier } n \geq N(\omega) \right\}$$

De manera que si $\omega \in \bigcap_{r=1}^{\infty} B_r^c$, entonces para cualquier $r \in \mathbb{N}$ existe $N(\omega) \in \mathbb{N}$ tal que $|X_n(\omega)| \leq \frac{1}{r}$ para cualquier $n \geq N(\omega)$. En particular, dada $\varepsilon > 0$ sea $r \in \mathbb{N}$ tal que $\frac{1}{r} < \varepsilon$ y $N(\omega) \in \mathbb{N}$ tal que $|X_n(\omega)| \leq \frac{1}{r}$ para cualquier $n \geq N(\omega)$, entonces $|X_n(\omega)| < \varepsilon$ para cualquier $n \geq N(\omega)$, lo cual significa que $\lim_{n \rightarrow \infty} X_n(\omega) = 0$. Es decir:

$$\bigcap_{r=1}^{\infty} B_r^c \subset [\lim_{n \rightarrow \infty} X_n = 0]$$

Sea $\Omega_0 = \bigcup_{r=1}^{\infty} B_r$. Entonces $P(\Omega_0) = 0$ y si $\omega \in \Omega_0^c = \bigcap_{r=1}^{\infty} B_r^c$, entonces $\lim_{n \rightarrow \infty} X_n(\omega) = 0$. Así que $\lim_{n \rightarrow \infty} X_n = 0$ casi seguramente. ■

Corolario 4. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias. Entonces $X_n \xrightarrow{c.s.} X$ si y sólo si para cualquier $\varepsilon > 0$, se tiene:*

$$P\left[\limsup \left\{ \omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon \right\}\right] = 0$$

Proposición 4. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias tales que:*

$$\sum_{n=1}^{\infty} P[|X_n| > \varepsilon] < \infty \text{ para cualquier } \varepsilon > 0$$

Entonces $X_n \xrightarrow{c.s.} 0$.

Demostración

Dada $\varepsilon > 0$, sea $A(\varepsilon) = \limsup \left\{ \omega \in \Omega : |X_n(\omega)| > \varepsilon \right\}$.

Por el lema de Borel-Cantelli, $P[A(\varepsilon)] = 0$ para cualquier $\varepsilon > 0$. Así que el resultado se sigue aplicando el corolario 4. ■

Corolario 5. *Sea X una variable aleatoria y $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias tales que $\sum_{n=1}^{\infty} P[|X_n - X| > \varepsilon] < \infty$ para cualquier $\varepsilon > 0$. Entonces $X_n \xrightarrow{c.s.} X$.*

RESULTADOS PARA LA CONVERGENCIA EN PROBABILIDAD

1. Sea $\{f_n\}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{\mu} f$ y $f_n \xrightarrow{\mu} g$, entonces $f = g$ casi en todas partes.
2. Sea c una constante y $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{\mu} f$, entonces $cf_n \xrightarrow{\mu} cf$.
3. Sean $(f_n)_{n \in \mathbb{N}}$ y $(g_n)_{n \in \mathbb{N}}$ dos sucesiones de funciones medibles tales que $f_n \xrightarrow{\mu} f$ y $g_n \xrightarrow{\mu} g$, entonces $f_n + g_n \xrightarrow{\mu} f + g$.
4. Sean f una función medible, $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{\mu} f$ y $g : \mathbb{R} \rightarrow \mathbb{R}$ una función uniformemente continua, entonces $g \circ f_n \xrightarrow{\mu} g \circ f$.
5. Sean f una función medible, $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles tales que $f_n \xrightarrow{\mu} f$ y $g : \mathbb{R} \rightarrow \mathbb{R}$ una función continua nula fuera de un intervalo $[a, b]$, entonces $g \circ f_n \xrightarrow{\mu} g \circ f$.
6. Supongamos que μ es finita y sean f una función medible, $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{\mu} f$ y $g : \mathbb{R} \rightarrow \mathbb{R}$ una función continua, entonces $g \circ f_n \xrightarrow{\mu} g \circ f$.
7. Supongamos que μ es finita y sean f y g dos funciones medibles, y $(f_n)_{n \in \mathbb{N}}$ y $(g_n)_{n \in \mathbb{N}}$ dos sucesiones de funciones medibles tales que $f_n \xrightarrow{\mu} f$ y $g_n \xrightarrow{\mu} g$, entonces $f_n g_n \xrightarrow{\mu} fg$.
8. Supongamos que μ es finita y sea $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles tal que $f_n \xrightarrow{c.t.p.} f$, entonces $f_n \xrightarrow{\mu} f$.
9. Sea $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles que converge en medida, entonces $(f_n)_{n \in \mathbb{N}}$ es de Cauchy en medida.
10. Sea $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles que converge en medida, entonces $(f_n)_{n \in \mathbb{N}}$ contiene una subsucesión $(f_{n_k})_{k \in \mathbb{N}}$ que converge casi en todas partes a una función medible f y tal que, dada $\delta > 0$, existe un conjunto medible A tal que $\mu(A) < \delta$ y la sucesión $(f_{n_k})_{k \in \mathbb{N}}$ converge uniformemente a f sobre A^c .
11. Si μ es finita y $(f_n)_{n \in \mathbb{N}}$ es una sucesión de funciones medibles que converge casi en todas partes a la función medible f , entonces, dada $\delta > 0$, existe un conjunto medible A tal que $\mu(A) < \delta$ y la sucesión $(f_n)_{n \in \mathbb{N}}$ converge uniformemente a f sobre A^c .
12. Sean g una función no negativa e integrable, f una función medible y $(f_n)_{n \in \mathbb{N}}$ una sucesión de funciones medibles tales que $|f_n| \leq g$ para cualquier $n \in \mathbb{N}$ y $f_n \xrightarrow{\mu} f$, entonces $\lim_{n \rightarrow \infty} \int_{\mathbb{E}} |f_n - f| d\mu = 0$.

Funciones de distribución

Definición 12 (Función de distribución). Si X es una variable aleatoria real, la función $F_X : \mathbb{R} \mapsto \mathbb{R}$, definida por $F_X(x) = P[X \leq x]$, es llamada la función de distribución de X .

Definición 13 (Función de distribución conjunta). Sean X_1, \dots, X_n n variables aleatorias reales. La función $F_{X_1, \dots, X_n} : \mathbb{R}^n \mapsto \mathbb{R}$, definida por:

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P[X_1 \leq x_1, \dots, X_n \leq x_n]$$

será llamada la función de distribución conjunta de X_1, \dots, X_n .

RESULTADOS

1. Sea X una variable aleatoria real y F_X su función de distribución, entonces:

i. F_X es una función no decreciente y continua por la derecha.

ii. $\lim_{x \rightarrow \infty} F_X(x) = 1$

iii. $\lim_{x \rightarrow -\infty} F_X(x) = 0$

iv. $F_X(x-) = P[X < x]$ para cualquier $x \in \mathbb{R}$.

2. Sean X_1, \dots, X_n n variables aleatorias y sea F_{X_1, \dots, X_n} su función de distribución conjunta, entonces, para cada:

$(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) \in \mathbb{R}^{n-1}$, se tiene:

i. La función $x \mapsto F_{X_1, \dots, X_n}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n)$, definida sobre \mathbb{R} , es no decreciente y continua por la derecha.

ii. $\lim_{x \rightarrow \infty} F_{X_1, \dots, X_n}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n)$
 $= F_{X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n)$.

iii. $\lim_{x \rightarrow -\infty} F_{X_1, \dots, X_n}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n) = 0$.

Una familia de variables aleatorias X_1, \dots, X_n puede verse como la función de Ω en \mathbb{R}^n que asigna a cada $\omega \in \Omega$ el vector $(X_1(\omega), \dots, X_n(\omega))$; de esta forma, podemos decir que las variables aleatorias forman un **vector aleatorio** (X_1, \dots, X_n) .

Recordemos que un rectángulo en \mathbb{R}^n es un conjunto de la forma $I_1 \times \dots \times I_n$, en donde I_1, \dots, I_n son intervalos en \mathbb{R} .

Si $R = I_1 \times \cdots \times I_n$ es un rectángulo en \mathbb{R}^n y $a_1, b_1, \dots, a_n, b_n$ son los extremos de I_1, \dots, I_n , respectivamente, los intervalos I_k serán llamados los lados del rectángulo y los puntos del conjunto $V_{(a_1, b_1, \dots, a_n, b_n)} = \{(x_1, \dots, x_n) : x_k \in \{a_k, b_k\} \text{ para toda } k\}$ serán llamados los vértices del rectángulo.

El rectángulo $(a_1, b_1] \times \cdots \times (a_n, b_n]$ será denotado por $R_{(a_1, b_1, \dots, a_n, b_n)}$ y $S_{(a_1, b_1, \dots, a_n, b_n)}^{(k)}$ denotará al conjunto:

$$\{(x_1, \dots, x_n) : x_i = a_i \text{ para } k \text{ índices } i \text{ y } x_i = b_i \text{ para el resto de índices}\}$$

Obviamente, $V_{(a_1, b_1, \dots, a_n, b_n)} = \bigcup_{k=0}^n S_{(a_1, b_1, \dots, a_n, b_n)}^{(k)}$.

3. Sean X_1, \dots, X_n n variables aleatorias y $(a_1, b_1] \times \cdots \times (a_n, b_n]$ un rectángulo de \mathbb{R}^n .

Entonces, para cualquier evento A se tiene:

$$\begin{aligned} & P([a_1 < X_1 \leq b_1, \dots, a_n < X_n \leq b_n] \cap A) \\ &= \sum_{k=0}^n (-1)^k \sum_{\{(x_1, \dots, x_n) \in S_{(a_1, b_1, \dots, a_n, b_n)}^{(k)}\}} P([X_1 \leq x_1, \dots, X_n \leq x_n] \cap A) \end{aligned}$$

4. Sean X_1, \dots, X_n n variables aleatorias y $(a_1, b_1] \times \cdots \times (a_n, b_n]$ un rectángulo de \mathbb{R}^n . Entonces:

$$\begin{aligned} & P([a_1 < X_1 \leq b_1, \dots, a_n < X_n \leq b_n]) \\ &= \sum_{k=0}^n (-1)^k \sum_{\{(x_1, \dots, x_n) \in S_{(a_1, b_1, \dots, a_n, b_n)}^{(k)}\}} F_{X_1, \dots, X_n}(x_1, \dots, x_n) \end{aligned}$$

5. Sean X_1, \dots, X_n n variables aleatorias y F_{X_1, \dots, X_n} su función de distribución conjunta, entonces:

$$\sum_{k=0}^n (-1)^k \sum_{\{(x_1, \dots, x_n) \in S_{(a_1, b_1, \dots, a_n, b_n)}^{(k)}\}} F_{X_1, \dots, X_n}(x_1, \dots, x_n) \geq 0$$

para cualquier rectángulo $(a_1, b_1] \times \cdots \times (a_n, b_n]$.

6. Sean X_1, \dots, X_n n variables aleatorias y F_{X_1, \dots, X_n} su función de distribución conjunta, entonces:

$$\lim_{m \rightarrow 0} F_{X_1, \dots, X_n}(x_1 + \delta_1^{(m)}, \dots, x_n + \delta_n^{(m)}) = F_{X_1, \dots, X_n}(x_1, \dots, x_n)$$

para cualquier vector $(x_1, \dots, x_n) \in \mathbb{R}^n$ y cualquier sucesión $\left((\delta_1^{(m)}, \dots, \delta_n^{(m)}) \right)_{m \in \mathbb{N}}$ que converja al vector $\bar{0} \in \mathbb{R}^n$ y tal que $\delta_1^{(m)}, \dots, \delta_n^{(m)}$ sean números reales positivos.

7. n variables aleatorias reales, X_1, \dots, X_n , son independientes si y sólo si:

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n)$$

para cualquier vector $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.

Sea X una variable aleatoria real y definamos:

$$D = \{x \in \mathbb{R} : P[X = x] > 0\},$$

$$C = \{x \in \mathbb{R} : P[X = x] = 0\},$$

$$p = P[X \in D].$$

Clasificaremos a las variables aleatorias de acuerdo al valor de p . Si $p = 0$, diremos que la variable aleatoria es continua, si $p = 1$ diremos que es discreta. Cuando $0 < p < 1$, se tiene $P[X \in D] > 0$ y $P[X \in C] > 0$, de manera que se puede decir que, en ese caso, la variable aleatoria tiene una parte discreta (la que corresponde al conjunto D) y una parte continua (la que corresponde al conjunto C).

Un subconjunto importante del conjunto de variables aleatorias continuas está formado por las variables aleatorias reales X cuya distribución μ_X es absolutamente continua con respecto a la medida de Lebesgue restringida a $\mathfrak{B}(\overline{\mathbb{R}})$. En este caso μ_X se puede extender de manera única a $\mathfrak{L}(\overline{\mathbb{R}})$ ya que si $E \in \mathfrak{L}(\overline{\mathbb{R}})$, consideremos $B \in \mathfrak{B}(\overline{\mathbb{R}})$ y $C \in \mathfrak{L}(\overline{\mathbb{R}})$, de medida de Lebesgue cero, tales que $E = B \cup C$; tomemos entonces $F \in \mathfrak{B}(\overline{\mathbb{R}})$ de medida de Lebesgue cero tal que $C \subset F$. Como μ_X es absolutamente continua con respecto a la medida de Lebesgue restringida a $\mathfrak{B}(\overline{\mathbb{R}})$, entonces, $\mu_X(F) = 0$. Así que podemos extender, de manera única, μ_X a $\mathfrak{L}(\overline{\mathbb{R}})$ definiendo $\mu_X(G) = 0$ para cualquier conjunto $G \in \mathfrak{L}(\overline{\mathbb{R}})$ de medida de Lebesgue cero. Diremos que una variable aleatoria de este tipo es absolutamente continua.

Definición 14. Si X es una variable aleatoria discreta, llamaremos función de densidad de X , y la denotaremos por f_X , a la función $f_X : \mathbb{R} \rightarrow [0, 1]$ definida mediante la relación:

$$f_X(x) = P[X = x]$$

Definición 15. Si X es una variable aleatoria absolutamente continua, llamaremos función de densidad de X , y la denotaremos por f_X , a cualquier función medible $f : (\mathbb{R}, \mathfrak{L}(\overline{\mathbb{R}})) \rightarrow \mathbb{R}^+$ tal que:

$$\mu_X(B) = \int_B f d\lambda$$

para cualquier conjunto $B \in \mathfrak{L}(\overline{\mathbb{R}})$.

Algunos ejemplos de distribuciones

Cualquier medida de probabilidad sobre $(\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$ es una distribución de probabilidad, es decir, una distribución de alguna variable aleatoria. La demostración de esta afirmación es inmediata, ya que si μ es una medida de probabilidad sobre $(\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$, entonces podemos tomar como espacio de probabilidad a la terna $(\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}), \mu)$. Entonces la variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ definida por $X(\omega) = \omega$ tiene como distribución a la medida μ . Sin embargo, es importante explicitar algunas de ellas, ya sea por su interés histórico o por presentarse con frecuencia en diferentes problemas. Así que a continuación haremos un listado de algunas distribuciones discretas y algunas absolutamente continuas, dando el nombre de la distribución y la función de densidad que la determina.

1. Distribución Bernoulli con parámetro p , donde $p \in [0, 1]$.

$$f(x) = \begin{cases} p & \text{si } x = 1 \\ 1 - p & \text{si } x = 0 \\ 0 & \text{en otro caso} \end{cases}$$

2. Distribución binomial con parámetros n y p , donde $n \in \mathbb{N}$ y $p \in [0, 1]$.

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x \in \{0, 1, \dots, n\} \\ 0 & \text{en otro caso} \end{cases}$$

3. Distribución geométrica con parámetro p , donde $p \in [0, 1]$.

$$f(x) = \begin{cases} p(1-p)^x & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{en otro caso} \end{cases}$$

4. Distribución binomial negativa con parámetros n y p , donde $n \in \mathbb{N}$ y $p \in [0, 1]$.

$$f(x) = \begin{cases} \binom{n+x-1}{x} p^n (1-p)^x & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{en otro caso} \end{cases}$$

5. Distribución Poisson con parámetro λ , donde λ es un número real positivo.

$$f(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{en otro caso} \end{cases}$$

6. Distribución hipergeométrica con parámetros r , s y n , donde $r, s, n \in \mathbb{N}$ y $n \leq r + s$.

$$f(x) = \begin{cases} \frac{\binom{r}{x}\binom{s}{n-x}}{\binom{r+s}{n}} & \text{si } x \in \{0, 1, \dots, n\} \\ 0 & \text{en otro caso} \end{cases}$$

7. Distribución uniforme discreta en el conjunto A , donde A es un conjunto finito de números reales.

$$f_X(x) = \begin{cases} \frac{1}{N} & \text{si } x \in A \\ 0 & \text{en otro caso} \end{cases}$$

donde N es el número de elementos de A .

8. Distribución uniforme en el conjunto A , donde A es un conjunto Lebesgue medible de medida positiva.

$$f_X(x) = \begin{cases} \frac{1}{\lambda(A)} & \text{si } x \in A \\ 0 & \text{en otro caso} \end{cases}$$

9. Distribución normal con parámetros μ y σ^2 , donde $\mu, \sigma \in \mathbb{R}$ y $\sigma > 0$.

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

10. Distribución normal estándar.

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

11. Distribución exponencial con parámetro λ , donde λ es un número real positivo.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

12. Distribución gama con parámetros α y λ , donde α y λ son números reales positivos.

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

donde $\Gamma : (0, \infty) \rightarrow \mathbb{R}$ es la función gama, la cual está definida por: $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.

Si X es una variable aleatoria real, μ_X su distribución y F_X su función de distribución, definamos $d_X : (0, 1) \mapsto \mathbb{R}$ y $c_X : (0, 1) \mapsto \mathbb{R}$ mediante las siguientes relaciones:

$$c_X(t) = \inf \{x \in \mathbb{R} : F_X(x) > t\}$$

$$d_X(t) = \inf \{x \in \mathbb{R} : F_X(x) \geq t\}$$

Teorema 4. *Sean X una variable aleatoria real con función de distribución F_X y U una variable aleatoria con distribución uniforme en el intervalo $(0, 1)$. Entonces la función de distribución de la variable aleatoria $d_X(U)$ es F_X .*

Teorema 5. *Sea X una variable aleatoria con función de distribución F_X y U una variable aleatoria con distribución uniforme en el intervalo $(0, 1)$. Entonces la función de distribución de la variable aleatoria $c_X(U)$ es F_X .*

Funciones de distribución como medidas

Definición 16. Diremos que una función $F : \mathbb{R} \mapsto \mathbb{R}$ es una función de distribución finita en 1 variable si satisface las siguientes propiedades:

i. F es una función no decreciente y continua por la derecha.

ii. $\lim_{x \rightarrow \infty} F_X(x) < \infty$

iii. $\lim_{x \rightarrow -\infty} F_X(x) = 0$

Definición 17. Para $n \in \{2, 3, \dots\}$, diremos que una función $F : \mathbb{R}^n \mapsto \mathbb{R}$ es una función de distribución finita en n variables si satisface las siguientes propiedades:

i. $\sum_{k=0}^n (-1)^k \sum_{\{x_1, \dots, x_n\} \in S_{(a_1, b_1, \dots, a_n, b_n)}^{(k)}} F(x_1, \dots, x_n) \geq 0$

para cualquier rectángulo $(a_1, b_1] \times \dots \times (a_n, b_n]$.

ii. $\lim_{m \rightarrow 0} F(x_1 + \delta_1^{(m)}, \dots, x_n + \delta_n^{(m)}) = F(x_1, \dots, x_n)$

para cualquier vector $(x_1, \dots, x_n) \in \mathbb{R}^n$ y cualquier sucesión $\left((\delta_1^{(m)}, \dots, \delta_n^{(m)}) \right)_{m \in \mathbb{N}}$ que converja al vector $\bar{0} \in \mathbb{R}^n$ y tal que $\delta_1^{(m)}, \dots, \delta_n^{(m)}$ sean números reales positivos.

iii. $\lim_{x \rightarrow -\infty} F(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n) = 0$

para cualquier $(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) \in \mathbb{R}^{n-1}$.

iv. Para cada $(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) \in \mathbb{R}^{n-1}$, el límite

$$\lim_{x \rightarrow \infty} F(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n)$$

existe y la función $G : \mathbb{R}^{n-1} \mapsto \mathbb{R}$ definida por:

$$G(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) = \lim_{x \rightarrow \infty} F(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n)$$

es una función de distribución finita en $n - 1$ variables.

Cuando $\lim_{(x_1, \dots, x_n) \rightarrow (\infty, \dots, \infty)} F(x_1, \dots, x_n) = 1$, diremos simplemente que F es una función de distribución en n variables.

Si $a, b \in \overline{\mathbb{R}}$ y $a \leq b$, entonces definimos $(a, b|$ de la siguiente manera:

$$(a, b| = \begin{cases} (a, b] & \text{si } b \in \mathbb{R} \\ (a, b) & \text{si } b = \infty \end{cases}$$

Si $F : \mathbb{R}^n \mapsto \mathbb{R}$ es una función de distribución finita y

$(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) \in \mathbb{R}^{n-1}$, definimos:

$$F(x_1, \dots, x_{j-1}, \infty, x_{j+1}, \dots, x_n) = \lim_{x \rightarrow \infty} F(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n).$$

$$F(x_1, \dots, x_{j-1}, -\infty, x_{j+1}, \dots, x_n) = 0.$$

Con estas convenciones, se tiene que:

$$\sum_{k=0}^n (-1)^k \sum_{\{(x_1, \dots, x_n) \in S_{(a_1, b_1, \dots, a_n, b_n)}^{(k)}\}} F_{X_1, \dots, X_n}(x_1, \dots, x_n) \geq 0$$

para cualquier rectángulo $(a_1, b_1| \times \dots \times (a_n, b_n|$.

Definición 18. Si $F : \mathbb{R}^n \mapsto \mathbb{R}$ es una función de distribución finita y $R = (a_1, b_1| \times \dots \times (a_n, b_n|$ es un rectángulo en \mathbb{R}^n , definimos $\mu_F(R)$ de la siguiente manera:

$$\mu_F(R) = \sum_{k=0}^n (-1)^k \sum_{\{(x_1, \dots, x_n) \in S_{(a_1, b_1, \dots, a_n, b_n)}^{(k)}\}} F(x_1, \dots, x_n)$$

RESULTADOS

1. Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ una función de distribución finita y $R = (a_1, b_1| \times \dots \times (a_n, b_n|$ un rectángulo en \mathbb{R}^n . Para cada intervalo $(a_i, b_i|$ consideremos una partición:

$$P_i = \left\{ a_i = c_0^{(i)} < c_1^{(i)} < \dots < c_{m_i}^{(i)} = b_i \right\}$$

Entonces:

$$\mu_F(R) = \sum_{j_i \in \{1, \dots, m_i\}} \mu_F \left(R_{(c_{j_1-1}^{(1)}, c_{j_1}^{(1)}, \dots, c_{j_n-1}^{(n)}, c_{j_n}^{(n)})} \right)$$

2. Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ una función de distribución finita, $R = (a_1, b_1| \times \dots \times (a_n, b_n|$ un rectángulo en \mathbb{R}^n y $R^{(j)} = (a_1^{(j)}, b_1^{(j)}| \times \dots \times (a_n^{(j)}, b_n^{(j)}|$ una colección finita de rectángulos en \mathbb{R}^n , ajenos por parejas, tal que $R = \bigcup_{j=1}^m R^{(j)}$, entonces:

$$\mu_F(R) = \sum_{j=1}^m \mu_F(R^{(j)})$$

3. Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ es una función de distribución finita, $R = (a_1, b_1| \times \dots \times (a_n, b_n|$ un rectángulo en \mathbb{R}^n y $R^{(j)} = (a_1^{(j)}, b_1^{(j)}| \times \dots \times (a_n^{(j)}, b_n^{(j)}|$ una colección finita de rectángulos en \mathbb{R}^n tal que $R \subset \bigcup_{j=1}^m R^{(j)}$, entonces:

$$\mu_F(R) \leq \sum_{j=1}^m \mu_F(R^{(j)})$$

4. Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ una función de distribución finita y R_1, \dots, R_k y $R^{(1)}, \dots, R^{(m)}$ dos colecciones finitas de rectángulos en \mathbb{R}^n , todos de la forma:

$$\left(a_1^{(1)}, b_1^{(1)} \mid \times \cdots \times \left(a_n^{(1)}, b_n^{(1)} \mid \right.$$

y tales que R_1, \dots, R_k son ajenos por parejas, $R^{(1)}, \dots, R^{(m)}$ son ajenos por parejas y $\bigcup_{i=1}^k R_i = \bigcup_{j=1}^m R^{(j)}$.

Entonces:

$$\sum_{i=1}^k \mu_F(R_i) = \sum_{j=1}^m \mu_F(R^{(j)})$$

5. Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ una función de distribución finita, $R = (a_1, b_1 \mid \times \cdots \times (a_n, b_n \mid$ un rectángulo en \mathbb{R}^n y $R^{(i)} = \left(a_1^{(i)}, b_1^{(i)} \mid \times \cdots \times \left(a_n^{(i)}, b_n^{(i)} \mid \right.$ una colección infinita de rectángulos en \mathbb{R}^n tal que $R \subset \bigcup_{i=1}^{\infty} R^{(i)}$, entonces:

$$\mu_F(R) \leq \sum_{i=1}^{\infty} \mu_F(R^{(i)})$$

6. Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ una función de distribución finita. Entonces existe una única medida finita μ_F , definida sobre los conjuntos borelianos de \mathbb{R}^n , tal que:

$$\mu_F((-\infty, x_1] \times \cdots \times (-\infty, x_n]) = F(x_1, \dots, x_n)$$

para cualquier $(x_1, \dots, x_n) \in \mathbb{R}^n$.

7. Sea $F : \mathbb{R}^n \mapsto \mathbb{R}$ una función de distribución. Entonces existe un espacio de probabilidad (Ω, \mathcal{F}, P) y una familia X_1, \dots, X_n de variables aleatorias reales definidas sobre Ω tal que F es la función de distribución conjunta de X_1, \dots, X_n .

8. Si μ es una medida finita definida sobre los conjuntos borelianos de \mathbb{R}^n , la función $F : \mathbb{R}^n \mapsto \mathbb{R}$ definida por:

$$F(x_1, \dots, x_n) = \mu((-\infty, x_1] \times \cdots \times (-\infty, x_n])$$

es una función de distribución finita y la medida μ_F que genera sobre los borelianos, por ser única, coincide con μ . De esta forma, toda medida finita μ definida sobre los conjuntos borelianos de \mathbb{R}^n está generada por una función de distribución finita en n variables.

Esperanza de variables aleatorias

Definición 19. Si X es una variable aleatoria no negativa, definimos la esperanza de X , $E[X]$, de la siguiente manera:

$$E[X] = \int_{\Omega} X dP$$

Definición 20. Diremos que la variable aleatoria X tiene esperanza finita si $E[|X|] < \infty$. En ese caso definimos su esperanza de la siguiente manera:

$$E[X] = E[X^+] - E[X^-]$$

La Esperanza de una variable aleatoria tiene las mismas propiedades que la integral, es decir, se tiene la linealidad, el teorema de la convergencia monótona, el lema de Fatou, etcétera.

Sea $\mu_X : \mathfrak{B}(\overline{\mathbb{R}}) \mapsto \mathbb{R}$ la medida de probabilidad definida por:

$$\mu_X(B) = P[X \in B]$$

$(\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}), \mu_X)$ es un espacio de probabilidad, el cual podemos completar, obteniendo una nueva σ -álgebra, la cual denotaremos por $\mathfrak{B}_X(\overline{\mathbb{R}})$, y una medida definida sobre $\mathfrak{B}_X(\overline{\mathbb{R}})$, para la cual no cambiaremos la notación; es decir, la denotaremos por μ_X .

Tenemos así un espacio de probabilidad completo $(\overline{\mathbb{R}}, \mathfrak{B}_X(\overline{\mathbb{R}}), \mu_X)$, de manera que podemos definir la integral de las funciones medibles $f : (\overline{\mathbb{R}}, \mathfrak{B}_X(\overline{\mathbb{R}}), \mu_X) \mapsto (\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$ como con cualquier otra medida de probabilidad y los resultados expuestos acerca de la integral son válidos también en este caso.

Para evitar confusiones, utilizaremos el término variable aleatoria únicamente para las funciones medibles $X : (\Omega, \mathfrak{F}, P) \rightarrow (\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$.

RESULTADOS

1. Si X es una variable aleatoria, entonces $E[f(X)] = \int_{\overline{\mathbb{R}}} f d\mu_X$ para cualquier función medible no negativa $f : (\overline{\mathbb{R}}, \mathfrak{B}_X(\overline{\mathbb{R}})) \mapsto (\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$.

2. Si X es una variable aleatoria, entonces $E[f(X)] = \int_{\overline{\mathbb{R}}} f d\mu_X$ para cualquier función integrable $f : (\overline{\mathbb{R}}, \mathfrak{B}_X(\overline{\mathbb{R}}), \mu_X) \mapsto (\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$.

3. Para cualquier variable aleatoria X , el conjunto $\{x \in \mathbb{R} : P[X = x] > 0\}$ es a lo más infinito numerable.

4. Sean X y Y dos variables aleatorias independientes de esperanza finita, entonces XY también tiene esperanza finita y $E[XY] = E[X]E[Y]$.

5. Sean X_1, \dots, X_n n variables aleatorias independientes de esperanza finita, entonces $\prod_{k=1}^n X_k$ también tiene esperanza finita y:

$$E\left[\prod_{k=1}^n X_k\right] = \prod_{k=1}^n E[X_k]$$

6. Sea X una variable aleatoria, entonces X tiene esperanza finita si y sólo si:

$$(R) \int_0^\infty [1 - F_X(x)] dx < \infty \text{ y } (R) \int_{-\infty}^0 F_X(x) dx < \infty$$

y, en ese caso, se tiene:

$$E[X] = (R) \int_0^\infty [1 - F_X(x)] dx - (R) \int_{-\infty}^0 F_X(x) dx$$

Donde (R) significa que las integrales se pueden tomar en el sentido de Riemann.

Cuando se tiene $(R) \int_0^\infty P[X > x] dx = \infty$ y $(R) \int_{-\infty}^0 P[X < x] dx < \infty$, se define $E[X] = \infty$, mientras que cuando $(R) \int_0^\infty P[X > x] dx < \infty$ y $(R) \int_{-\infty}^0 P[X < x] dx = \infty$, se define $E[X] = -\infty$. Cuando ambas integrales sean divergentes, entonces la esperanza de X no está definida.

Varianza y covarianza

Definición 21. Sea X una variable aleatoria de esperanza finita. Se define la **varianza** de X , $Var(X)$, mediante la relación:

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

Definición 22. A la raíz cuadrada no negativa de la varianza se le llama la **desviación estándar** de X .

Definición 23. Diremos que una variable aleatoria X tiene varianza finita si se cumplen las siguientes dos condiciones:

- i. X tiene esperanza finita.
- ii. $(X - E[X])^2$ tiene esperanza finita.

Definición 24. Sean X y Y dos variables aleatorias de varianza finita. Se define la **covarianza** de X y Y , $Cov(X, Y)$, mediante la relación:

$$Cov(X, Y) = E [(X - E[X])(Y - E[Y])] = E [XY] - E[X] E[Y]$$

RESULTADOS

1. Una variable aleatoria X tiene varianza finita si y sólo si X^2 tiene esperanza finita.
2. Sea X una variable aleatoria de esperanza finita, entonces:

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

3. Sean X y Y dos variables aleatorias de varianza finita. Entonces, XY tiene esperanza finita.
4. Si X y Y son dos variables aleatorias de varianza finita y a y b son dos números reales cualesquiera, entonces $aX + bY$ tiene varianza finita.
5. Sean X y Y dos variables aleatorias independientes de varianza finita, entonces $Cov(X, Y) = 0$.

6. Sean X y Y dos variables aleatorias de varianza finita y a y b dos números reales cualesquiera. Entonces $Cov(aX, bY) = abCov(X, Y)$.

7. Sean X, X_1, \dots, X_n $n + 1$ variables aleatorias de esperanza finita. Entonces:

a) $Var(X) = 0$ si y sólo si existe una constante c tal que $P[X = c] = 1$.

b) $Var(aX + b) = a^2Var(X)$ para cualquier pareja $a, b \in \mathbb{R}$.

c) Si X_1, \dots, X_n tienen varianza finita, entonces $\sum_{i=1}^n X_i$ también tiene varianza finita y:

$$Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + 2 \sum_{\{i,j \in \{1, \dots, n\}; i < j\}} Cov(X_i, X_j)$$

8. Sean X_1, \dots, X_n n variables aleatorias independientes y de varianza finita, entonces $\sum_{i=1}^n X_i$ también tiene varianza finita y:

$$Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i)$$

9. Desigualdad de Cauchy-Schwarz: Sean X y Y dos variables aleatorias cualesquiera, entonces:

$$E[|XY|] \leq \sqrt{E[X^2]} \sqrt{E[Y^2]}$$

Además, si X y Y tienen varianza finita, entonces $|E[XY]| = \sqrt{E[X^2]} \sqrt{E[Y^2]}$ si y sólo si existen constantes a y b tales que por lo menos una de ellas es distinta de cero y $P[aX + bY = 0] = 1$.

10. Sean X y Y dos variables aleatorias de varianza finita. Entonces:

$$|Cov(X, Y)| \leq \sqrt{Var(X)} \sqrt{Var(Y)}$$

Además, la igualdad se cumple si y sólo si existen constantes a, b y c tales que a y b no son ambas cero y $P[aX + bY = c] = 1$.