¿QUÉ REPRESENTA LA FUNCIÓN DE DISTRIBUCIÓN DE UNA VARIABLE ALEATORIA?

Miguel Ángel García Álvarez

1. Introducción

Si $(\Omega, \Im, P)$ es un espacio de probabilidad, una variable aleatoria real definida sobre ese espacio es una función $X : \Omega \mapsto \mathbb{R}$ con la propiedfad de que, para cualquier $x \in \mathbb{R}$, el conjunto $[X \leq x]$ es un evento, es decir, pertenece a $\Im$. De aquí que podamos asociar a la variable aleatoria X la función $F_X : \mathbb{R} \mapsto [0, 1]$ definida por $F_X(x) = P[X \leq x]$. De esta forma, una variable aleatoria se estudia vía su función de distribución; es decir, todo lo que podamos decir de X lo hacemos mediante F_X; dicho de otra manera, la función de distribución F_X representa a la variable aleatoria X.

El objetivo de este escrito es mostrar que, a su vez, la función de distribución F_X de una variable aleatoria X representa una medida de probabilidad definida sobre los conjuntos borelianos de $\mathbb{R}$. En otras palabras, a partir de una función de distribución podemos generar una medida de probabilidad, la cual, finalmente, es la que representa a la variable aleatoria.

El ver a una función de distribución como una medida es importante ya que esta idea se puede generalizar de tal manera que, a partir de la función de distribución conjunta $F_{X_1,X_2,\ldots,X_n}$ de una familia finita de variables aleatorias podemos generar una medida de probabilidad sobre los conjuntos borelianos de $\mathbb{R}^n$.

Avanzando más con esta idea, cuando trabajamos con un proceso estocástico, lo que tenemos es una familia infinita (numerable o no numerable) de variables aleatorias y, en este caso, la herramienta matemática de que disponemos no nos alcanza para definir una función de distribución conjunta asociada a esa familia infinita de variables aleatorias; sin embargo, lo que sí tenemos a la mano son medidas definidas sobre espacios de dimensión infinita; de manera que es posible asociar al proceso estocástico en consideración una de esas medidas. Así que, el concepto general es el de medida, siendo una función de distribución conjunta $F_{X_1,X_2,\ldots,X_n}$ únicamente un caso particular.

2. Variables aleatorias y funciones de distribución

La teoría de la probabilidad es una teoría matemática axiomática, de manera que los elementos con los que trata son abstractos, independientes de cualquier interpretación práctica que se les dé.

Sin embargo, haremos una pequeña disgresión de tipo práctico con el objeto de clarificar un poco esos elementos con los que trata la probabilidad.

Partimos del concepto de experimento aleatorio, al cual le pedimos dos características: 1. Para cada realización del experimento, podemos especificar de manera precisa cuál es su resultado; 2. El experimento admite diferentes posibles resultados; es decir, el resultado de cada realización del experimento puede ser distinto al de otra de ellas.

Así que, dado un experimento aleatorio, tenemos un conjunto de posibles resultados, el cual suele denotarse con la letra Ω y es llamado el espacio muestral del experimento.

El segundo concepto con el que trata la probabilidad es el de evento, el cual se puede definir (para fines prácticos) como una aseveración que se hace con respecto al resultado del experimento aleatorio antes de que éste se realice, de manera que, cuando se realiza el experimento, esa aseveración que hacemos puede ser verdadera o falsa. Para cada evento hay un conjunto de posibles resultados que hacen que, cuando se realiza el experimento, la aseveración que hicimos resulta verdadera. Esto permite asociar a cada evento un subconjunto del espacio muestral, el que está formado por todos los posibles resultados que hacen que la aseveración sea verdadera.

Habiendo representado los eventos como subconjuntos del espacio muestral, podemos utilizar las operaciones que se tienen definidas entre los conjuntos para definir operaciones entre eventos: unión, intersección, complemento, diferencia, etcétera.

Así que, para modelar el experimento aleatorio que estemos considerando, tenemos dos elementos: su espacio muestral, representado por un conjunto Ω y una familia de subconjuntos de Ω, que denotaremos por la letra $\Im$ y que representan a los eventos relativos al experimento aleatorio.

El tercer elemento es la parte central del modelo, la medida de probabilidad, la cual representaremos con la letra P. Esta función de probabilidad P está definida sobre la familia de eventos $\Im$ y toma valores entre 0 y 1.

El modelo matemático que se utiliza en la teoría de la probabilidad consta entonces de una terna $(\Omega, \Im, P)$ y se requiere que esa terna satisfaga determinadas condiciones.

1. La familia de eventos $\Im$ debe tener a Ω como elemento y debe ser cerrada bajo complementos y uniones numerables. Cuando se cumple esto, decimos que $\Im$ forma una $\sigma-$álgebra de subconjuntos de Ω.

2. La medida de probabilidad P, como dijimos, es una función $P : \Im \mapsto [0, 1]$ y se requiere que cumpla con dos propiedades:

a) $P(\Omega) = 1$.

b) Si $A_1, A_2, \ldots$ es una colección finita o infinita numerable de elementos de $\Im$ (es decir, de eventos), ajenos por parejas, entonces $P(\cup_n A_n) = \sum_n P(A_n)$. A esta propiedad se le conoce como la $\sigma-$aditividad de P.

A una terna $(\Omega, \Im, P)$ con las propiedades mencionadas se le llama un espacio de probabilidad.

Partimos entonces de que tenemos un espacio de probabilidad $(\Omega, \Im, P)$, donde Ω es un conjunto cualquiera, $\Im$ es una σ-algebra de subconjuntos de Ω y P es una función $\sigma-$aditiva definida sobre $\Im$ tal que $P(\Omega) = 1$.

De las propiedades de la medida de probabilidad P, se derivan otras que son importantes para su manejo:

1. Si $A \in \Im$, entonces $P(A^c) = 1 - P(A)$.

2. Si $A, B \in \Im$ y $A \subset B$, entonces $P(A) \leq P(B)$ y $P(B - A) = P(B) - P(A)$.

3. Si $A, B \in \Im$, entonces $P(AUB) = P(A) + P(B) - P(A \cap B)$.

4. Si $(A_n)_{n\in\mathbb{N}}$ es una sucesión no decreciente de elementos de $\Im$, entonces $P(\cup_{n=1}^{\infty} A_n) = \text{lím}_{n\rightsquigarrow\infty} P(A_n)$.

5. Si $(A_n)_{n\in\mathbb{N}}$ es una sucesión no creciente de elementos de $\Im$, entonces $P(\cap_{n=1}^{\infty} A_n) = \text{lím}_{n\rightsquigarrow\infty} P(A_n)$.

Una variable aleatoria real es una función $X : \Omega \mapsto \mathbb{R}$ con la propiedad de que, para cualquier $x \in \mathbb{R}$, el conjunto $\{\omega \in \Omega : X(\omega) \leq x\}$ (el cual, en forma abreviada se denota por $[X \leq x]$) pertenece a la familia de eventos $\Im$.

La propiedad que caracteriza a una variable aleatoria X permite definir lo que llamamos su función de distribución y que denotamos por F_X, la cual es una función $F_X : \mathbb{R} \rightarrow [0, 1]$ definida por:

$$F_X(x) = P[X \leq x]$$

Toda función de distribución F_X tiene las siguientes propiedades:

1. Es no decreciente.

2. Es continua por la derecha.

3. $\text{lím}_{x\rightsquigarrow\infty} F_X(x) = 1$.

4. $\text{lím}_{x\rightsquigarrow-\infty} F_X(x) = 0$.

Aunque quienes lean este escrito ya conocen las demostraciones de estas propiedades, vamos a reproducirlas aquí ya que se trata de analizar lo que representa la función de distribución de una variable aleatoria.

1. Si $x < y$, entonces $[X \leq x] \subset [X \leq y]$, así que $P[X \leq x] \leq P[X \leq y]$; es decir, $F_X(x) \leq F_X(y)$.

2. Si $(x_n)_{n\in\mathbb{N}}$ es una sucesión decreciente de números reales que converge a x, entonces la sucesión de eventos $([X \leq x_n])_{n\in\mathbb{N}}$ es no creciente y $\cap_{n=1}^{\infty} [X \leq x_n] = [X \leq x]$, así que:

$$\text{lím}_{n\rightsquigarrow\infty} F_X(x_n) = \text{lím}_{n\rightsquigarrow\infty} P[X \leq x_n] = P(\cap_{n=1}^{\infty} [X \leq x_n]) = P[X \leq x] = F_X(x)$$

3. Si $(x_n)_{n\in\mathbb{N}}$ es una sucesión creciente de números reales que diverge a ∞, entonces la sucesión de eventos $([X \leq x_n])_{n\in\mathbb{N}}$ es no decreciente y $\cup_{n=1}^{\infty} [X \leq x_n] = \Omega$, así que:

$$\text{lím}_{n\rightsquigarrow\infty} F_X(x_n) = \text{lím}_{n\rightsquigarrow\infty} P[X \leq x_n] = P(\cup_{n=1}^{\infty} [X \leq x_n]) = P(\Omega) = 1$$

4. Si $(x_n)_{n\in\mathbb{N}}$ es una sucesión decreciente de números reales que diverge a $-\infty$, entonces la sucesión de eventos $([X \leq x_n])_{n\in\mathbb{N}}$ es no creciente y $\cap_{n=1}^{\infty}[X \leq x_n] = \emptyset$, así que:

$$\text{lím}_{n\rightsquigarrow\infty} F_X(x_n) = \text{lím}_{n\rightsquigarrow\infty} P[X \leq x_n] = P(\cap_{n=1}^{\infty}[X \leq x_n]) = P(\emptyset) = 1 - P(\Omega) = 0$$

Ahora viene la parte más interesante.

La idea es que cualquier probabilidad del tipo $P[X \in B]$, donde B es un subconjunto de $\mathbb{R}$, se pueda obtener de la función de distribución de X. Sin embargo, en general esto no es posible, únicamente se pueden calcular ese tipo de probabilidades para una clase particular de conjuntos. De manera más específica, se pueden calcular las probabilidades de la forma $P[X \in B]$, cuando B es un elemento de la $\sigma-$álgebra generada por los intervalos; es decir, la más pequeña $\sigma-$álgebra que contenga a todos los intervalos. Esa $\sigma-$álgebra es llamada la $\sigma-$álgebra de los conjuntos borelianos (en honor a Émile Borel) de $\mathbb{R}$.

Ahora bien, aunque la $\sigma-$álgebra de los conjuntos borelianos es bastante grande, no lo es tanto ya que se demuestra que su cardinalidad es la misma que la de $\mathbb{R}$. Sin embargo, se puede agrandar la familia de subconjuntos B de $\mathbb{R}$ para los cuales se puede calcular $P[X \in B]$ a partir de la función de distribución F_X de X; para esto basta con agregar a la $\sigma-$álgebra de los conjuntos borelianos todos los subconjuntos de los conjuntos $B \subset \mathbb{R}$ tales que $P[X \in B] = 0$. Con la $\sigma-$álgebra de los conjuntos borelianos y esa familia de conjuntos que se agregan se genera una nueva $\sigma-$álgebra que podríamos denotar por $\mathcal{L}_X$. Se demuestra que esa $\sigma-$álgebra tiene la misma cardinalidad que la del conjunto potencia de $\mathbb{R}$. Cabe aclarar que todo esto es solamente una extensión de lo que hizo Lebesgue al definir lo que se llama la medida de Lebesgue.

Hasta ahí se puede llegar; las probabilidades de la forma $P[X \in B]$, donde B es un elemento de $\mathcal{L}_X$ son las únicas que es posible calcular y todas ellas se obtienen de la función de distribución F_X de X. En otras palabras, toda la información (dentro del modelo matemático que estamos considerando) de una variable aleatoria X está condensada en su función de distribución; de ahí su importancia. De hecho, no interesa tanto conocer una variable X como una función $X : \Omega \mapsto \mathbb{R}$; lo que nos interesa es conocer su función de distribución.

¿Cómo calcularíamos, a partir de la función de distribución de X, las probabilidades $P[X < b]$, $P[a \leq X < b]$, $P[a \leq X \leq b]$, $P[a < X < b]$, $P[X \geq b]$, $P[X > b]$ y $P[X = b]$?

Lo haríamos de la siguiente manera:

1. Para calcular $P[X < b]$ sonsideramos una sucesión creciente de números reales $(x_n)_{n\in\mathbb{N}}$ que converja a b. Tenemos entonces:

$$\cup_{n=1}^{\infty}[X \leq x_n] = [X < b]$$

Así que:

$$P[X < b] = P(\cup_{n=1}^{\infty}[X \leq x_n]) = \text{lím}_{n\rightsquigarrow\infty} P[X \leq x_n] = \text{lím}_{n\rightsquigarrow\infty} F_X(x_n) = F_X(b-)$$

donde $F_X(b-)$ denota el límite de $F_x(x)$ cuando x tiende a b por la izquierda.

2. $P[a \leq X < b] = P[X < b] - P[X < a] = F_X(b-) - F_X(a-)$

3. $P[a \leq X \leq b] = P[X \leq b] - P[X < a] = F_X(b) - F_X(a-)$

4. $P[a < X < b] = P[X < b] - P[X \leq a] = F_X(b-) - F_X(a)$

5. $P[X \geq b] = 1 - P[X < b] = 1 - F_X(b-)$

6. $P[X > b] = 1 - P[X \leq b] = 1 - F_X(b)$

7. $P[X = b] = P[X \leq b] - P[X < b] = F_X(b) - F_X(b-)$

Obsérvese que la última igualdad nos dice que $P[X = b]$ es igual a la magnitud del salto de la función F_X en el punto b. Esto implica, en particular, que F_X es continua en b si y sólo si $P[X = b] = 0$. En otras palabras, el conjunto de discontinuidades de F_X coincide con el conjunto de puntos $x \in \mathbb{R}$ tales que $P[X = x] = 0$.

Con lo anterior ya tenemos definida la probabilidad $P(I)$, donde I es un intervalo en $\mathbb{R}$ de cualquier tipo. En efecto, si I es un intervalo, definimos $\mu_X(I) = P[X \in I]$, probabilidad que ya sabemos calcular a partir de F_X.

3. Generación de medidas a partir de funciones de distribución

¿Cómo se obtienen, a partir de la función de distribución F_X de una variable aleatoria X, todas las probabilidades de la forma $P[X \in B]$, donde $B \in \mathcal{L}_X$?

La respuesta es que se obtienen generando una medida de probabilidad definida sobre $\mathcal{L}_X$ a partir de F_X. Y esto se logra siguiendo el mismo camino que siguió Lebesgue para extender el concepto de longitud a todos los conjuntos Lebesgue-medibles.

Vamos a hacerlo.

Si $a, b \in \mathbb{R}$ y $a < b$, definimos:

$$\mu_X((a, b]) = F_X(b) - F_X(a)$$

¿Por qué un intervalo de la forma $(a, b]$? Porque, en general, F_X es únicamente continua por la derecha; puede ser discontinua por la izquierda. Y la idea es que, para cualquier número real x, $\mu_X((-\infty, x])$ coincida con el valor de $F_X(x)$, el cual está dado por $P[X \leq x]$; así que tendríamos:

$$F_X(b) - F_X(a) = P[X \leq b] - P[X \leq a] = P[a < X \leq b] = P[X \in (a, b]]$$

Una vez definida μ_X para los intervalos, de lo que se trata ahora es de extenderla a los conjuntos borelianos. Pero antes, obsérvese que la familia formada por todos los intervalos de números reales es cerrada bajo intersecciones finitas; sin embargo, por ejemplo, el complemento de un intervalo $[a, b]$ no es un intervalo, sino la unión de dos intervalos.

Lo que se ha visto cuando se trata de construir una medida es que resulta más simple hacerlo si se tiene definida esa medida para una familia de conjuntos que formen un álgebra; es decir, que sea cerrada bajo complementos y uniones finitas y que Ω pertenezca a la familia. Así que vamos a hacer eso: como siguiente paso para definir la medida μ_X sobre todos los conjuntos borelianos, vamos a definirla primero para el álgebra generada por los intervalos; ésta está formada por todos los subconjuntos de números reales que son uniones finitas de intervalos ajenos.

Trabajar con todo tipo de intervalos puede ser laborioso, pero la tarea puede simplificarse ya que si consideramos únicamente intervalos de la forma $(a, b]$, (a, ∞) o $(-\infty, b]$, éstos generan la misma $\sigma-$álgebra que todos los intervalos. Denotemos por $\mathcal{H}$ a la familia formada por todos los intervalos de ese tipo.

Definamos entonces $\mathcal{A}$ como la familia formada por los conjuntos de la forma $\bigcup_{j=1}^{n} I_j$ donde $n \in \mathbb{N}$ y $I_1, \ldots, I_n$ son intervalos en $\mathcal{H}$, ajenos por parejas.

Para cada $A = \bigcup_{j=1}^{n} I_j \in \mathcal{A}$, definamos $\mu_X(A) = \sum_{j=1}^{n} \mu_X(I_j)$

Aquí viene una parte técnica, la cual es sencilla, pero requiere de varios pasos. Tenemos que mostrar que μ_X está bien definida sobre $\mathcal{A}$. El problema que se puede presentar es que un elemento de $\mathcal{A}$ tenga varias representaciones como unión finita de intervalos ajenos; así que hay que mostrar que con cualquiera de ellas se llega al mismo resultado.

Con el objeto de no tener que tratar el problema por casos, demos las siguientes definiciones:

$F_X(\infty) = 1$

$F_X(-\infty) = 0$

Si $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$ y $a < b$, definamos $(a, b|$ de la siguiente manera:

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

$\mathcal{H}$ es entonces la familia formada por todos los intervalos de este tipo. Para fines prácticos, agreguemos a $\mathcal{H}$ el conjunto vacío y definamos $\mu_X(\emptyset) = 0$.

Obsérvese que si $I = (a, b| \in \mathcal{H}$, entonces $\mu_X(I) = F_X(b) - F_X(a)$. Además, $\mu_X(\mathbb{R}) = \mu_X[(-\infty, \infty)] = 1$.

Lema 1. *Sea* $I = (a, b| \in \mathcal{H}$ *y* $\left(a^{(1)}, b^{(1)}\right|, \left(a^{(2)}, b^{(2)}\right|, \ldots, \left(a^{(m)}, b^{(m)}\right|$, *una colección finita de intervalos en* $\mathcal{H}$, *ajenos por parejas, tal que* $I = \bigcup_{j=1}^{m} \left(a^{(j)}, b^{(j)}\right|$, *entonces:*

$\mu_X(I) = \sum_{j=1}^{m} \mu_X\left(\left(a^{(j)}, b^{(j)}\right|\right)$

Demostración

Como los intervalos $\left(a^{(1)}, b^{(1)}\right|, \left(a^{(2)}, b^{(2)}\right|, \ldots, \left(a^{(m)}, b^{(m)}\right|$ son ajenos por parejas y su unión es $(a, b|$, podemos ordenarlos para obtener una colección de intervalos ajenos por parejas, $\left(x^{(1)}, y^{(1)}\right|, \left(x^{(2)}, y^{(2)}\right|, \ldots, \left(x^{(m)}, y^{(m)}\right|$, de tal forma que:

$a = x^{(1)} < y^{(1)} = x^{(2)} < y^{(2)} = x^{(2)} < \cdots = x^{(m)} < y^{(m)} = b$

Así que:

$\sum_{j=1}^{m} \mu_X\left(\left(a^{(j)}, b^{(j)}\right|\right) = \sum_{j=1}^{m} \mu_X\left(\left(y^{(j)}, z^{(j)}\right|\right) = \sum_{j=1}^{m} \left[F_X\left(y^{(j)}\right) - F_X\left(x^{(j)}\right)\right]$

$= F_X(b) - F_X(a) = \mu_X(I)$

■

Lema 2. *Sea* $I = (a, b| \in \mathcal{I}$ *y* $I^{(1)} = \left(a^{(1)}, b^{(1)}\right|, \ldots, I^{(m)} = \left(a^{(m)}, b^{(m)}\right|$ *una colección finita de intervalos en* $\mathcal{H}$ *tales que* $I \subset \bigcup_{j=1}^{m} I^{(j)}$*, entonces:*

$\mu_X(I) \leq \sum_{j=1}^{m} \mu_X\left(I^{(j)}\right)$.

Demostración

Los puntos $a, b, a^{(1)}, b^{(1)}, \ldots, a^{(m)}, b^{(m)}$ constituyen una partición de un intervalo $(c, d|$ que contiene al intervalo $(a, b|$.

Esta partición parte cada intervalo $I^{(j)}$, con $j \in \{1, 2, \ldots, m\}$, en subintervalos ajenos por parejas, $I_1^{(j)}, \ldots, I_{n_j}^{(j)}$. Así que:

$\mu_X\left(I^{(j)}\right) = \sum_{k=1}^{n_j} \mu_X\left(I_k^{(j)}\right)$.

La partición definida antes también parte el intervalo I en subintervalos ajenos por parejas, $I_1, \ldots, I_n$. Así que:

$\mu_X(I) = \sum_{k=1}^{n} \mu_X(I_k)$.

Por otra parte, como $I \subset \bigcup_{j=1}^{m} I^{(j)}$, cada intervalo I_k, con $k \in \{1, 2, \ldots, n\}$, coincide con un intervalo $I_{k'}^{(j)}$ para alguna $j \in \{1, 2, \ldots, m\}$ y alguna $k' \in \{1, 2, \ldots, n_j\}$, por lo tanto:

$\mu_X(I) = \sum_{k=1}^{n} \mu_X(I_k) \leq \sum_{j=1}^{m} \sum_{k=1}^{n_j} \mu_X\left(I_k^{(j)}\right) = \sum_{j=1}^{m} \mu_X\left(I^{(j)}\right)$.

■

Lema 3. *Sean* $I_1, \ldots, I_k$ *y* $I^{(1)}, \ldots, I^{(m)}$ *dos colecciones finitas de intervalos en* $\mathcal{H}$ *tales que* $I_1, \ldots, I_k$ *son ajenos por parejas,* $I^{(1)}, \ldots, I^{(m)}$ *son ajenos por parejas y* $\bigcup_{i=1}^{k} I_i = \bigcup_{j=1}^{m} I^{(j)}$*, entonces:*

$\sum_{i=1}^{k} \mu_X (I_i) = \sum_{j=1}^{m} \mu_X \left(I^{(j)}\right)$.

Demostración

Para cada $i \in \{1, \ldots, k\}$ y $j \in \{1, \ldots, m\}$, definamos $I_i^{(j)} = I_i \cap I^{(j)}$. Entonces, como $\bigcup_{i=1}^{k} I_i = \bigcup_{j=1}^{m} I^{(j)}$, se tiene $I_i = \bigcup_{j=1}^{m} I_i^{(j)}$ y $I^{(j)} = \bigcup_{i=1}^{k} I_i^{(j)}$, así que:

$\mu_X (I_i) = \sum_{j=1}^{m} \mu_X \left(I_i^{(j)}\right)$,

$\mu_X \left(I^{(j)}\right) = \sum_{i=1}^{k} \mu_X \left(I_i^{(j)}\right)$.

Por lo tanto:

$\sum_{i=1}^{k} \mu_X (I_i) = \sum_{i=1}^{k} \sum_{j=1}^{m} \mu_X \left(I_i^{(j)}\right)$

$= \sum_{j=1}^{m} \sum_{i=1}^{k} \mu_X \left(I_i^{(j)}\right) = \sum_{j=1}^{m} \mu_X \left(I^{(j)}\right)$.

■

Por los lemas anteriores, μ_X está bien definida.

Además, $\mathcal{A}$ es un álgebra de subconjuntos de $\mathbb{R}$ y la función $\mu_X : \mathcal{A} \mapsto \mathbb{R}$ es no negativa y finitamente aditiva; es decir, si $A_1, A_2, \ldots, A_n$ es una colección finita de elementos de $\mathcal{A}$, ajenos por parejas, entonces $\mu_X (\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} \mu_X (A_k)$.

3.1. El resultado central.

En la demostración del siguiente teorema vamos a utilizar el teorema de Heine-Borel, el cual asegura que si un intervalo cerrado $[a, b]$ es cubierto por una unión cualquiera de intervalos abiertos, entonces existe una familia finita de esos intervalos abiertos cuya unión también cubre al intervalo $[a, b]$.

Teorema 1. *Sea $(a, b| \in \mathcal{H}$ y $(a_1, b_1| , (a_2, b_2| , \ldots$ una colección infinita de intervalos en $\mathcal{H}$ tales que $(a, b| \subset \bigcup_{k=1}^{\infty} (a_k, b_k|$. Entonces:*

$F_X(b) - F_X(a) \leq \sum_{k=1}^{\infty} [F_X(b_k) - F_X(a_k)]$.

Demostración

Tomemos $\varepsilon > 0$ y $\delta > 0$, arbitrarios.

Como F_X es continua por la derecha, para cada cada $k \in \mathbb{N}$, existe $\delta_k > 0$ tal que:

$F_X (d_k) - F_X (b_k) < \frac{\varepsilon}{2^k}$,

donde:

$d_k = \begin{cases} b_k + \delta_k & \text{si } b_k \in \mathbb{R} \\ b_k & \text{si } b_k = \infty \end{cases}$

Definamos:

$c_\delta = \begin{cases} a + \delta & \text{si } a \in \mathbb{R} \\ -\frac{1}{\delta} & \text{si } a = -\infty \end{cases}$

$d_\delta = \begin{cases} b & \text{si } b \in \mathbb{R} \\ \frac{1}{\delta} & \text{si } b = \infty \end{cases}$

Entonces:

$\lim_{\delta \to 0} \left[F_X\left(d_\delta\right) - F_X\left(c_\delta\right)\right] = F_X\left(b\right) - F_X\left(a\right)$.

Además:

$(c_\delta, d_\delta| \subset [c_\delta, d_\delta] \subset (a, b| \subset \bigcup_{k=1}^{\infty} (a_k, b_k| \subset \bigcup_{k=1}^{\infty} (a_k, d_k)$.

Así que, por el teorema de Heine-Borel, existe una colección finita, $(a_{k_1}, d_{k_1}), \ldots, (a_{k_m}, d_{k_m})$, tal que:

$[c_\delta, d_\delta] \subset \bigcup_{j=1}^{m} \left(a_{k_j}, d_{k_j}\right)$

Por lo tanto:

$(c_\delta, d_\delta| \subset [c_\delta, d_\delta] \subset \bigcup_{j=1}^{m} \left(a_{k_j}, d_{k_j}\right) \subset \bigcup_{j=1}^{m} \left(a_{k_j}, d_{k_j}\right|$.

Así que:

$F_X\left(d_\delta\right) - F_X\left(c_\delta\right) \leq \sum_{j=1}^{m} \left[F_X\left(d_{k_j}\right) - F_X\left(a_{k_j}\right)\right] \leq \sum_{k=1}^{\infty} \left[F_X\left(d_k\right) - F_X\left(a_k\right)\right]$

$\leq \sum_{k=1}^{\infty} \left[F_X\left(b_k\right) - F_X\left(a_k\right)\right] + \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \sum_{k=1}^{\infty} \left[F_X\left(b_k\right) - F_X\left(a_k\right)\right] + \varepsilon$.

Y, como $\varepsilon > 0$ es arbitraria:

$F_X\left(d_\delta\right) - F_X\left(c_\delta\right) \leq \sum_{k=1}^{\infty} \left[F_X\left(b_k\right) - F_X\left(a_k\right)\right]$.

Finalmente, tomando límites cuando $\delta \to 0$, se obtiene:

$F_X\left(b\right) - F_X\left(a\right) \leq \sum_{k=1}^{\infty} \left[F_X\left(b_k\right) - F_X\left(a_k\right)\right]$.

■

Teorema 2. *Sea $A_1, A_2, \ldots$ una colección infinita numerable de elementos de $\mathcal{A}$, ajenos por parejas, no vacíos y tales que $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. Entonces:*

$\mu_X\left(A\right) = \sum_{i=1}^{\infty} \mu_X\left(A_i\right)$

Demostración

Como para cada $i \in \mathbb{N}$, $A_i \in \mathcal{A}$, A_i es una unión finita de intervalos en $\mathcal{I}$ ajenos por parejas. Además, como $A \in \mathcal{A}$, A también es una unión finita de intervalos en $\mathcal{H}$ ajenos por parejas.

Sean $A = \bigcup_{j=1}^{m} I^{(j)}$ y, para cada $i \in \mathbb{N}$, $A_i = \bigcup_{k=1}^{m_i} I_{(i,k)}$. Entonces:

$\bigcup_{j=1}^{m} I^{(j)} = A = \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \bigcup_{k=1}^{m_i} I_{(i,k)}$

Tenemos dos colecciones de intervalos, por un lado la familia $\left\{I_{(i,k)} : i \in \mathbb{N}, k \in \{1, \ldots, m_i\}\right\}$ y por el otro la familia $\left\{I^{(j)} : j \in \{1, \ldots, m\}\right\}$. Tanto los intervalos de la primera familia como los de la segunda son ajenos por parejas y A es igual tanto a la unión de los intervalos de la primera familia como de la segunda. Por otra parte, una pareja de intervalos, uno de la primera familia y otro de la segunda, podrían no ser ajenos.

La idea ahora es partir cada intervalo $I^{(j)}$ en intervalos ajenos por parejas, utilizando los intervalos $I_{(i,k)}$. Para esto, definamos, para cada $j \in \{1, \ldots, m\}$, $i \in \mathbb{N}$ y $k \in \{1, \ldots, m_i\}$:

$I_{(i,k)}^{(j)} = I_{(i,k)} \cap I^{(j)}$

Entonces, los intervalos de la familia $\left\{I_{(i,k)}^{(j)} : j \in \{1, \ldots, m\}, i \in \mathbb{N}, k \in \{1, \ldots, m_i\}\right\}$ son ajenos por parejas y:

$I_{(i,k)} = \bigcup_{j=1}^{m} I_{(i,k)}^{(j)}$ para cualesquiera $i \in \mathbb{N}$ y $k \in \{1, \ldots, m_i\}$.

$I^{(j)} = \bigcup_{i=1}^{\infty} \bigcup_{k=1}^{m_i} I_{(i,k)}^{(j)}$ para cualquier $j \in \{1, \ldots, m\}$.

Además, por el teorema 1, se tiene:

$\mu_X\left(I^{(j)}\right) \leq \sum_{i=1}^{\infty} \sum_{k=1}^{m_i} \mu_X\left(I_{(i,k)}^{(j)}\right)$ para cualquier $j \in \{1, \ldots, m\}$.

Así que:

$\mu_X(A) = \sum_{j=1}^{m} \mu_X\left(I^{(j)}\right) \leq \sum_{j=1}^{m} \sum_{i=1}^{\infty} \sum_{k=1}^{m_i} \mu_X\left(I_{(i,k)}^{(j)}\right)$

$= \sum_{i=1}^{\infty} \sum_{k=1}^{m_i} \sum_{j=1}^{m} \mu_X\left(I_{(i,k)}^{(j)}\right) = \sum_{i=1}^{\infty} \sum_{k=1}^{m_i} \mu_X\left(I_{(i,k)}\right)$

$= \sum_{i=1}^{\infty} \mu_X(A_i)$

Además, como μ_X es finitamente aditiva y $A \supset \bigcup_{i=1}^{n} A_i$ para cualquier $n \in \mathbb{N}$, se tiene $\mu_X(A) \geq \sum_{i=1}^{n} \mu_X(A_i)$ para cualquier $n \in \mathbb{N}$, así que:

$\mu_X(A) \geq \sum_{i=1}^{\infty} \mu_X(A_i)$.

Por lo tanto:

$$\mu_X\left(A\right)=\sum_{i=1}^{\infty}\mu_X\left(A_i\right)$$

■

A la propiedad enunciada en el resultado anterior se le conoce como la $\sigma-$aditividad de μ_X.

Teorema 3. *Sea $A_1, A_2, \ldots$ una colección infinita numerable de elementos de $\mathcal{A}$ tales que $A=\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. Entonces:*

$\mu_X\left(A\right)\leq\sum_{i=1}^{\infty}\mu_X\left(A_i\right)$

Demostración

Para cada $n\in\mathbb{N}$, definamos $B_n=A_n-\bigcup_{j=1}^{n-1}A_j$, entonces $\bigcup_j A_j=\bigcup_j B_j$, así que:

$\mu_X(\bigcup_j A_j)=\mu_X(\bigcup_j B_j)=\sum_{k=1}^{\infty}\mu_X(B_j)\leq\sum_{k=1}^{\infty}\mu_X(A_j).$

■

A la propiedad enunciada en el resultado anterior se le conoce como la $\sigma-$subaditividad de μ_X.

3.2. Extensión de la medida.

Vamos a ver cómo, siguiendo el método de Lebesgue, se puede extender μ_X a la σ-álgebra, $\sigma(\mathcal{A})$, generada por $\mathcal{A}$. Primero definiremos la medida exterior de cualquier subconjunto de $\mathbb{R}$; después definiremos la medibilidad de un conjunto. Una vez hecho esto, mostraremos que la familia de conjuntos medibles forma una σ-álgebra, la cual contiene a los elementos de $\mathcal{A}$ y a los conjuntos de medida exterior cero. La medida de un conjunto medible la definiremos como su medida exterior y mostraremos que la medida así definida, restringida a $\mathcal{A}$, coincide con μ_X.

Definición 1. *Diremos que una colección finita o infinita numerable $A_1, A_2, \ldots$ de elementos de $\mathcal{A}$ es una cubierta del conjunto A si $A\subset\bigcup_n A_n$.*

Definición 2. *Se define la medida exterior, $\mu_e(A)$, de un conjunto A, mediante la relación*

$\mu_e(A)=\text{ínf}\left\{\sum_j\mu_X(A_j) : A_1, A_2, \ldots \text{ es cubierta de } A\right\}.$

Proposición 1. *Si A y B son dos conjuntos tales que $A\subset B$ entonces $\mu_e(A)\leq\mu_e(B)$.*

Gracias a la σ-subaditividad de μ_X se tiene el siguiente resultado:

Proposición 2. *Si $A\in\mathcal{A}$ entonces $\mu_e(A)=\mu_X(A)$.*

Demostración

Sea $A\in\mathcal{A}$ y $A_1, A_2, \ldots$ una cubierta de A, entonces $A_n\cap A\in\mathcal{A}$ para cualquier $n\in\mathbb{N}$ y $A=\bigcup_n\left(A_n\cap A\right)$; así que, como μ_X es σ-subaditiva:

$\mu_X(A) \leq \sum_n \mu_X(A_n \cap A) \leq \sum_n \mu_X(A_n)$

Por lo tanto, como esto ocurre para cualquier cubierta de A, $\mu_X(A) \leq \mu_e(A)$.

Por otra parte, como A es una cubierta de él mismo, se tiene $\mu_e(A) \leq \mu_X(A)$.

■

Proposición 3. *Si* $(A_n)_{n\in\mathbb{N}}$ *es una sucesión de conjuntos, entonces:*

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

Demostración

Si $\mu_e(A_n) = \infty$ para alguna n el resultado es trivial.

Supongamos entonces que $\mu_e(A_n) < \infty$ para toda n. Dada $\varepsilon > 0$, para cada conjunto A_n sea $A_{n1}, A_{n2}, \ldots$ una cubierta de A_n tal que $\sum_m \mu_X(A_{nm}) < \mu_e(A_n) + \frac{\varepsilon}{2^n}$. La familia de conjuntos A_{nm} forman una cubierta de $\bigcup_n A_n$, así que:

$\mu_e\left(\bigcup_n A_n\right) \leq \sum_n \sum_m \mu_X(A_{nm}) \leq \sum_n \left[\mu_e(A_n) + \frac{\varepsilon}{2^n}\right] \leq \sum_n \mu_e(A_n) + \varepsilon$;

es decir, $\mu_e\left(\bigcup_n A_n\right) \leq \sum_n \mu_e(A_n) + \varepsilon$ para cualquier $\varepsilon > 0$. Por lo tanto:

$\mu_e\left(\bigcup_n A_n\right) \leq \sum_n \mu_e(A_n)$.

■

Definición 3. *Diremos que un conjunto E es medible si $\mu_e(A) = \mu_e(A \cap E) + \mu_e(A \cap E^c)$ para cualquier conjunto A. Además, en este caso, se define la medida de E, $\mu(E)$, como la medida exterior de E.*

Obsérvese que, por la σ-subaditividad de la medida exterior, se tiene:

$$\mu_e(A) \leq \mu_e(A \cap E) + \mu_e(A \cap E^c)$$

para cualquier par de conjuntos E y A, de manera que para demostrar la medibilidad de un conjunto E únicamente es necesario probar la otra desigualdad.

Además, si A es cualquier subconjunto de $\mathbb{R}$, se tiene:

$\mu_e(A \cap \mathbb{R}) + \mu_e(A \cap \mathbb{R}^c) = \mu_e(A) + \mu_e(\emptyset) = \mu_e(A) + \mu_X(\emptyset) = \mu_e(A)$

Así que $\mathbb{R}$ es medible y $\mu(\mathbb{R}) = \mu_e(\mathbb{R}) = \mu_X(\mathbb{R}) = 1$

Proposición 4. *La familia de conjuntos medibles forma un álgebra de subconjuntos de $\mathbb{R}$.*

Demostración

Que $\mathbb{R}$ es medible, así como que el complemento de un conjunto medible es medible, son resultados obvios.

Sean E_1 y E_2 dos conjuntos medibles y A cualquier conjunto. Se tiene entonces:

$\mu_e\left(A \cap (E_1 \cup E_2)\right) + \mu_e\left(A \cap (E_1 \cup E_2)^c\right)$

$= \mu_e\left((A \cap E_1) \cup (A \cap E_1^c \cap E_2)\right) + \mu_e\left(A \cap E_1^c \cap E_2^c\right)$

$\leq \mu_e(A \cap E_1) + \mu_e(A \cap E_1^c \cap E_2) + \mu_e\left(A \cap E_1^c \cap E_2^c\right)$

$= \mu_e(A \cap E_1) + \mu_e(A \cap E_1^c) = \mu_e(A).$

Así que, $E_1 \cup E_2$ es medible.

■

Proposición 5. *La función que asigna a cada conjunto medible E su medida, $\mu(E)$, es una función finitamente aditiva.*

Demostración

Sean E_1 y E_2 dos conjuntos medibles ajenos, entonces, como $E_1 \cup E_2$ es medible, se tiene:

$\mu\left(E_1 \cup E_2)\right) = \mu\left((E_1 \cup E_2)) \cap E_1\right) + \mu\left((E_1 \cup E_2) \cap E_1^c\right)$

$= \mu\left(E_1\right) + \mu\left(E_2\right).$

■

Proposición 6. *La familia de conjuntos medibles forma una σ-álgebra de subconjuntos de $\mathbb{R}$.*

Demostración

Sea $E_1, E_2, \ldots$ una colección infinita numerable de conjuntos medibles ajenos por parejas y A cualquier subconjunto de $\mathbb{R}$.

Demostremos que $\mu_e\left(A \cap (\bigcup_{j=1}^{n} E_j)\right) = \sum_{j=1}^{n} \mu_e(A \cap E_j)$ para cualquier $n \in \mathbb{N}$.

Para $n = 1$ la igualdad es obvia.

Supongamos ahora que la igualdad es válida para $n = k$, entonces, como E_{k+1} es medible, se tiene:

$\mu_e\left(A \cap (\bigcup_{j=1}^{k+1} E_j)\right) = \mu_e\left(A \cap (\bigcup_{j=1}^{k+1} E_j) \cap E_{k+1}\right) + \mu_e\left(A \cap (\bigcup_{j=1}^{k+1} E_j) \cap E_{k+1}^c\right)$

$= \mu_e\left(A \cap E_{k+1}\right) + \mu_e\left(A \cap (\bigcup_{j=1}^{k} E_j)\right) = \mu_e\left(A \cap E_{k+1}\right) + \sum_{j=1}^{k} \mu_e(A \cap E_j)$

$= \sum_{j=1}^{k+1} \mu_e(A \cap E_j).$

Por lo tanto, la igualdad es válida para $n = k + 1$, así que, por el principio de inducción, lo es para cualquier $n \in \mathbb{N}$.

Ahora bien, como la familia de conjuntos medibles forma un álgebra de subconjuntos de $\mathbb{R}$, para cada $n \in \mathbb{N}$ el conjunto $\bigcup_{j=1}^{n} E_j$ es medible, así que:

$$\mu_e(A) = \mu_e\left(A \cap (\bigcup_{j=1}^{n} E_j)\right) + \mu_e\left(A \cap (\bigcup_{j=1}^{n} E_j)^c\right)$$

$$= \sum_{j=1}^{n} \mu_e(A \cap E_j) + \mu_e\left(A \cap (\bigcup_{j=1}^{n} E_j)^c\right)$$

$$\geq \sum_{j=1}^{n} \mu_e(A \cap E_j) + \mu_e\left(A \cap (\bigcup_{j=1}^{\infty} E_j)^c\right).$$

Tomando límite cuando $n \rightsquigarrow \infty$ y utilizando la σ-subaditividad de la medida exterior,se obtiene entonces:

$$\mu_e(A) \geq \sum_{j=1}^{\infty} \mu_e(A \cap E_j) + \mu_e\left(A \cap (\bigcup_{j=1}^{\infty} E_j)^c\right)$$

$$\geq \mu_e\left(A \cap (\bigcup_{j=1}^{\infty} E_j)\right) + \mu_e\left(A \cap (\bigcup_{j=1}^{\infty} E_j)^c\right).$$

Por lo tanto, $\bigcup_{j=1}^{\infty} E_j$ es medible. ■

Proposición 7. *La función que asigna a cada conjunto medible E su medida, $\mu(E)$, es una función $\sigma-$aditiva.*

Demostración

Sea $E_1, E_2, \ldots$ una colección infinita numerable de conjuntos medibles ajenos por parejas. Por la σ-subaditividad de la medida exterior, se tiene $\mu(\bigcup_{j=1}^{\infty} E_j) \leq \sum_{j=1}^{\infty} \mu(E_j)$. Por otra parte, por la aditividad finita de la función que asigna a cada conjunto medible su medida, se tiene, para cualquier $n \in \mathbb{N}$:

$$\mu(\bigcup_{j=1}^{\infty} E_j) \geq \mu(\bigcup_{j=1}^{n} E_j) = \sum_{j=1}^{n} \mu(E_j).$$

Así que tomando límite cuando $n \rightsquigarrow \infty$, se Aiene:

$$\mu(\bigcup_{j=1}^{\infty} E_j) \geq \sum_{j=1}^{\infty} \mu(E_j).$$

■

Si denotamos por $\mathcal{M}$ a la familia de los conjuntos medibles, sabemos que $\mathcal{M}$ forma una σ-álgebra de subconjuntos de $\mathbb{R}$. Además, la función $\mu : \mathcal{M} \mapsto \mathbb{R}$ es no negativa, $\sigma-$aditiva y $\mu(\mathbb{R}) = 1$. Así que μ es una medida de probabilidad definida sobre $\mathcal{M}$. Lo que resta probar es que μ es una extensión de μ_X. Vamos a demostrar que efectivamente esto es así.

Proposición 8. *Todo conjunto de medida exterior cero es medible.*

Demostración

Sea E un conjunto de medida exterior cero y A cualquier conjunto, entonces $A \cap E$ tiene medida exterior cero, así que:

$\mu_e(A) \geq \mu_e(A \cap E^c) = \mu_e(A \cap E^c) + \mu_e(A \cap E)$.

■

Proposición 9. *Todo elemento de $\mathcal{A}$ es medible.*

Demostración

Sea $E \in \mathcal{A}$, A cualquier conjunto y $A_1, A_2, \ldots$ una cubierta de A, entonces, para cada A_n, los conjuntos $A_n \cap E$ y $A_n \cap E^c$ pertenecen a $\mathcal{A}$ y se tiene:

$\mu_e(A \cap E) \leq \mu_e\left(\left(\bigcup_n A_n\right) \cap E\right) = \mu_e\left(\bigcup_n (A_n \cap E)\right)$

$\leq \sum_n \mu_e(A_n \cap E) = \sum_n \mu_X(A_n \cap E)$,

$\mu_e(A \cap E^c) \leq \mu_e\left(\left(\bigcup_n A_n\right) \cap E^c\right) = \mu_e\left(\bigcup_n (A_n \cap E^c)\right)$

$\leq \sum_n \mu_e(A_n \cap E^c) = \sum_n \mu_X(A_n \cap E^c)$.

Así que:

$\mu_e(A \cap E) + \mu_e(A \cap E^c) \leq \sum_n \mu_X(A_n \cap E) + \sum_n \mu_X(A_n \cap E^c) = \sum_n \mu_X(A_n)$.

Finalmente, como lo anterior es válido para cualquier cubierta de A, se puede concluir que:

$\mu_e(A \cap E) + \mu_e(A \cap E^c) \leq \mu_e(A)$.

■

Proposición 10. *Todo elemento de $\sigma(\mathcal{A})$ es medible.*

Demostración

El resultado es inmediato pues la familia de conjuntos medibles forma una σ-álgebra que contiene a los elementos de $\mathcal{A}$.

■

Corolario 1. *μ es una medida de probabilidad definida sobre una $\sigma-$álgebra de subconjuntos de $\mathbb{R}$, la cual contiene a la $\sigma-$álgebra generada por $\mathcal{A}$ y los conjuntos de medida exterior cero. Además, μ restringida a $\mathcal{A}$ coincide con μ_X*

Podemos renombrar a μ como μ_X, de manera que, en resumen, tenemos el siguiente resultado:

Teorema 4. *Existe una medida de probabilidad μ_X, definida sobre los conjuntos borelianos de $\mathbb{R}$, tal que:*

$\mu_X((a, b]) = F_X(b) - F_X(a)$,

para cualquier pareja de números reales, a y b, tales que $a < b$.

Esta medida μ_X es la que representa a la variable aleatoria X en el sentido de que, si A es un conjunto boreliano de $\mathbb{R}$, entonces:

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