



Universidad Tecnológica de la Mixteca

Proyecto Final

Apuntes del curso.

**Alumno:** Ceballos Juárez Marco Antonio

**Profesor:** Dr. Ignacio Arroyo Fernández

**Curso:** Probabilidad

**Fecha:** 20 de septiembre de 2024

# 1. Introducción

El origen de la teoría de la probabilidad se encuentra en observaciones físicas asociadas a juegos de azar. Un ejemplo clásico es el lanzamiento de una moneda, donde al repetir el experimento un número muy grande de veces, la frecuencia relativa de obtener cara tiende a  $1/2$ . De manera similar, al extraer una carta de una baraja perfectamente barajada, la frecuencia relativa de obtener un trébol tiende a  $1/4$  después de muchos intentos.

# 2. Definición clásica de probabilidad

La probabilidad de un evento es el número de resultados favorables al evento, divididos por el número total de resultados (o experimentos) de un experimento, donde todos los resultados son iguales en probabilidad.

Por ejemplo, en el caso de una baraja de 52 cartas, la probabilidad de obtener un trébol es:

$$P(\text{Trébol}) = \frac{13}{52} = \frac{1}{4}$$

# 3. Espacio muestral

Un espacio muestral  $\Omega$  es el conjunto de todos los posibles resultados de un experimento aleatorio. Por ejemplo, si se lanza una moneda dos veces, el espacio muestral es:

$$\Omega = \{AB, AA, BA, BB\}$$

Donde, A es el resultado “cara” y B es el resultado “cruz”.

# 4. Medida de probabilidad

Una medida de probabilidad es una función que asigna a cada evento un número real no negativo que refleja la “probabilidad” de que dicho evento ocurra. Formalmente, se define una medida de probabilidad  $P$  en un espacio muestral  $\Omega$ .

# 5. Propiedades de la medida de probabilidad

Para que  $P$  sea una medida de probabilidad, debe cumplir tres propiedades fundamentales:

1. **No negatividad:** Para cualquier evento  $A \in \mathcal{F}$ , se tiene que:

$$P(A) \geq 0$$

2. **Normalización:** La probabilidad del espacio muestral completo es 1, es decir:

$$P(\Omega) = 1$$

3. **Aditividad contable:** Si  $A_1, A_2, A_3, \dots$  son eventos disjuntos en  $\mathcal{F}$  (es decir,  $A_i \cap A_j = \emptyset$  para  $i \neq j$ ), entonces la probabilidad de la unión de estos eventos es igual a la suma de sus probabilidades:

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

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n).$$

Para ver un script de ejemplo de un espacio muestral dirigirse al Ejemplo 1.

## 6. Probabilidad condicional

Si los eventos  $A$  y  $B$  son independientes, la ocurrencia o no ocurrencia de  $A$  no influye en la probabilidad de que ocurra  $B$ . Si no son independientes, es necesario medir cuánto cambia la probabilidad de  $B$  dada la ocurrencia de  $A$ . Esto se define como la **probabilidad condicional**, denotada por  $P(B|A)$ , y se calcula como:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{si } P(A) > 0$$

Para ver un script de ejemplo sobre la union e interseccion de eventos dirigirse al Ejemplo 2.

### 6.1. Ejemplo 1: Lanzamiento de dos dados

Consideremos el lanzamiento de dos dados. Sea  $A$  el evento de que la suma de las caras es 8, y  $B$  el evento de que las caras son iguales. Queremos encontrar  $P(B|A)$ .

El espacio muestral consiste en los 36 posibles pares ordenados de resultados  $(i,j)$  con  $i, j = 1, 2, \dots, 6$ , cada uno con probabilidad  $1/36$ . Las combinaciones que suman 8 son:  $(4,4)$ ,  $(5,3)$ ,  $(3,5)$ ,  $(6,2)$ , y  $(2,6)$ .

- El evento  $A$  ocurre en 5 combinaciones.

- De esas combinaciones,  $B$  ocurre solo en una, que es cuando las caras son iguales  $(4,4)$ .

Por lo tanto, la probabilidad condicional es:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/36}{5/36} = \frac{1}{5}$$

## 7. Teorema de la probabilidad total

Sea  $\{B_1, B_2, \dots\}$  una familia de eventos mutuamente excluyentes y exhaustivos, entonces para cualquier evento  $A$ , se cumple:

$$P(A) = \sum_i P(B_i)P(A|B_i)$$

Este teorema permite calcular la probabilidad de  $A$  dividiendo el espacio muestral en subconjuntos más manejables representados por los  $B_i$ .

## 8. Regla de Bayes

La **Regla de Bayes** es un principio fundamental en la teoría de probabilidad, que describe cómo actualizar la probabilidad de una hipótesis basada en nueva evidencia. Formalmente, la regla se expresa como:

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

Donde:

- $P(A|B)$  es la probabilidad *a posteriori* del evento  $A$  dado el evento  $B$ , es decir, la probabilidad de que ocurra  $A$  después de observar  $B$ .
- $P(A)$  es la probabilidad *a priori* del evento  $A$ , antes de conocer  $B$ .
- $P(B|A)$  es la probabilidad de observar el evento  $B$  si el evento  $A$  es verdadero.
- $P(B)$  es la probabilidad total de observar el evento  $B$ , también conocida como la probabilidad marginal de  $B$ . Se puede calcular usando la ley de la probabilidad total:

$$P(B) = P(B|A)P(A) + P(B|\neg A)P(\neg A)$$

La Regla de Bayes es útil en una variedad de contextos, especialmente en la toma de decisiones bajo incertidumbre, el aprendizaje automático, y el diagnóstico médico, donde se necesita ajustar las probabilidades a medida que se adquiere nueva información.

Para ver un script de ejemplo sobre la regla de Bayes dirigirse al Ejemplo 3.

### 8.1. Ejemplo 3: Monedas

Imaginemos que tenemos dos monedas: una justa y otra con dos caras. Seleccionamos una moneda al azar con probabilidad  $3/4$  de escoger la moneda justa. Si lanzamos la moneda y sale cara, queremos saber la probabilidad de haber escogido la moneda de dos caras.

Utilizamos el teorema de Bayes. Definimos:

-  $B_1$ : evento de escoger la moneda justa. -  $B_2$ : evento de escoger la moneda de dos caras. -  $A$ : evento de que salga cara.

La probabilidad de que se haya escogido la moneda de dos caras dado que salió cara es:

$$P(B_2|A) = \frac{P(B_2 \cap A)}{P(A)} = \frac{(1/4)(1)}{(3/4)(1/2) + (1/4)(1)} = \frac{2}{5}$$

## 9. Regla de la cadena

La **regla de la cadena** se utiliza para calcular la probabilidad conjunta de varios eventos dependientes.

Si  $A$  y  $B$  son independientes entonces:

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

Esto implica que:

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

Por lo tanto, también:

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

Lo anterior puedes interpretarse como que lo que se sabe de  $A$  no se ve afectado por lo que se sabe de  $B$  y también funciona al revés siempre y cuando  $P(A \cap B) = \mathbf{0}$

Para tres eventos  $A$ ,  $B$  y  $C$ , se tiene:

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

Esto extiende la probabilidad condicional a múltiples eventos, permitiendo calcular la probabilidad conjunta descomponiendo el cálculo en términos de probabilidades condicionales sucesivas.

## 10. Variable aleatoria

Una variable aleatoria se define formalmente como una función medible de un espacio muestral  $\Omega$  a los números reales. Intuitivamente, es una variable que asigna un valor numérico al resultado de un experimento aleatorio.

### 10.1. Ejemplos de variables aleatorias

#### Ejemplo 1: Lanzamiento de monedas

Consideremos el experimento de lanzar una moneda 10 veces. Sea  $R$  el número de caras observadas. El espacio muestral consiste en todas las posibles secuencias de caras (C) y cruces (X), y hay  $2^{10} = 1024$  resultados posibles. Un punto muestral podría ser  $CCXCXCCXC$ , donde  $R(\omega) = 6$ , ya que hay 6 caras.

Otra variable aleatoria podría ser el número de veces que una cara es seguida inmediatamente por una cruz. Para el mismo punto muestral, contamos  $R_1(\omega) = 3$  ocurrencias de este tipo.

#### Ejemplo 2: Lanzar dados

Consideremos el experimento de lanzar dos dados. El espacio muestral es el conjunto de todos los pares de enteros  $(x, y)$ , donde  $x, y \in \{1, 2, 3, 4, 5, 6\}$ , por lo que hay 36 resultados posibles.

Definimos las siguientes variables aleatorias:

- $R_1(x, y)$  es el resultado del primer dado, por lo que  $R_1(x, y) = x$ .
- $R_2(x, y)$  es la suma de los dos dados, por lo que  $R_2(x, y) = x + y$ .
- $R_3(x, y)$  vale 1 si al menos uno de los dados muestra un número par, y 0 en caso contrario. Por ejemplo,  $R_3(6, 5) = 1$  y  $R_3(1, 3) = 0$ .

## 10.2. Definición de formal de una variable aleatoria

Una **variable aleatoria** en un espacio de probabilidad  $(\Omega, F, P)$  es una función  $R$ , tal que para cada subconjunto de Borel  $B \subset \mathbb{R}$ , el conjunto  $\{\omega : R(\omega) \in B\} \in F$ . Es suficiente comprobar que  $\{\omega : a < R(\omega) < b\} \in F$  para todo  $a, b \in \mathbb{R}$ .

## 10.3. Clasificación de variables aleatorias

Las variables aleatorias pueden ser:

- **Discretas:** Si sus posibles valores son finitos o contablemente infinitos.
- **Continuas:** Si pueden tomar cualquier valor dentro de un intervalo.

## 11. Función de probabilidad

La probabilidad de un evento relacionado con una variable aleatoria  $X$  puede expresarse de manera más compacta como:

$$P(\omega : X \in B) = P(X \in B),$$

donde  $B = \{a \leq x \leq b\}$  representa un intervalo en el que estamos interesados.

La probabilidad  $P(X \in B)$  se puede expresar como la suma de las probabilidades de los valores discretos que  $X$  puede tomar dentro del conjunto  $B$ :

$$P(X \in B) = P(X = x_1) + P(X = x_2) + \cdots + P(X = x_n) = \sum_{i=1}^n p_X(x_i),$$

donde  $p_X(x)$  es la función de probabilidad de  $X$ , definida como:

$$P(X = x) = p_X(x).$$

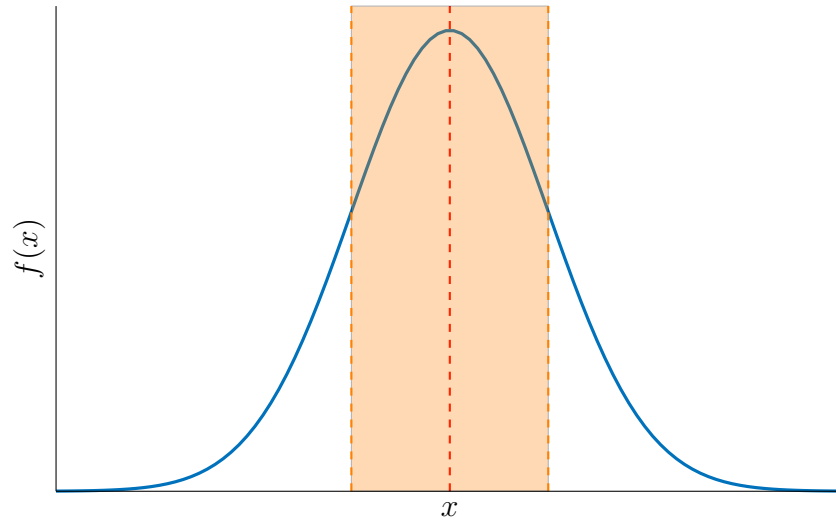
Nótese que la suma anterior se interpreta como la suma de la función de probabilidad sobre el conjunto de puntos que son favorables para el evento en cuestión. Esta expresión es particularmente útil cuando se trabaja con distribuciones discretas, ya que permite calcular la probabilidad de que la variable aleatoria  $X$  tome valores dentro de un intervalo específico. En el caso de distribuciones continuas, la probabilidad de que  $X$  se encuentre dentro del intervalo  $B$  se determina integrando la función de densidad de probabilidad  $f(x)$ :

$$P(X \in B) = \int_a^b f(x) dx.$$

Esta integral da como resultado el área bajo la curva de la función de densidad de probabilidad  $f(x)$  entre los límites  $a$  y  $b$ , lo que corresponde a la probabilidad de que  $X$  esté en el intervalo  $[a, b]$ .

Media  $\mu = 0$

Distribución Normal:  $\mu = 0, \sigma = 1$



Intervalo  $B = [a, b]$

### 11.1. Función de probabilidad binomial

La **distribución binomial** describe el número de éxitos en una secuencia de  $n$  experimentos independientes, donde cada experimento tiene solo dos resultados posibles: éxito (con probabilidad  $p$ ) ó fracaso (con probabilidad  $q = 1 - p$ ). La función de probabilidad de la distribución binomial se expresa como:

$$P(X = k) = \binom{n}{k} p^k q^{n-k}$$

donde:

- $P(X = k)$  es la probabilidad de obtener exactamente  $k$  éxitos en  $n$  ensayos.
- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$  es el coeficiente binomial, que cuenta las formas de elegir  $k$  éxitos de  $n$  ensayos.
- $p$  es la probabilidad de éxito en un solo ensayo.
- $q$  es la probabilidad de fracaso en un solo ensayo.

### 11.2. Ejemplo

Supongamos que lanzamos una moneda justa 10 veces y queremos calcular la probabilidad de obtener exactamente 6 caras. Aquí,  $n = 10$ ,  $k = 6$ , y  $p = 0.5$ . La probabilidad de obtener exactamente 6 caras se calcula como:

$$P(X = 6) = \binom{10}{6} (0.5)^6 (0.5)^{10-6} = \binom{10}{6} (0.5)^{10}$$

Calculando el coeficiente binomial:

$$\binom{10}{6} = \frac{10!}{6!4!} = 210$$

Por lo tanto:

$$P(X = 6) = 210 \cdot (0.5)^{10} = 210 \cdot \frac{1}{1024} \approx 0.205$$

Para ver un script de ejemplo sobre la probabilidad binomial dirigirse al Ejemplo 4.

## 12. Funciones de una variable aleatoria

La **función de una variable aleatoria** se refiere a cómo se transforma una variable aleatoria  $R_1$  mediante una función real  $g$  para generar otra variable aleatoria  $R_2 = g(R_1)$ . En este contexto,  $R_1$  representa el “input” y  $g$  actúa como el “sistema”, generando una salida  $R_2$ . Dado que se conoce la distribución o la función de densidad de  $R_1$ , el objetivo es determinar la distribución o función de densidad de  $R_2$ .

Si  $R_1$  es una variable aleatoria y definimos  $R_2 = g(R_1)$ ,  $R_2$  será también una variable aleatoria siempre que  $g$  sea continua o continua a trozos.

### 12.1. Función de distribución

La función de distribución acumulativa (FDA) de una variable aleatoria  $X$  se define como:

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

Para una variable aleatoria continua, la FDA se puede expresar como una integral de su función de densidad de probabilidad  $f_X(x)$ :

$$F_X(x) = \int_{-\infty}^x f_X(t) dt$$

donde  $f_X(t)$  es la función de densidad de probabilidad de  $X$ . La FDA tiene las siguientes propiedades importantes:

- $F_X(x)$  es una función no decreciente.
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$  y  $\lim_{x \rightarrow \infty} F_X(x) = 1$ .
- Para cualquier  $a < b$ , la probabilidad de que  $X$  esté en el intervalo  $(a, b]$  se puede calcular como:

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

La función de distribución es fundamental para describir el comportamiento de una variable aleatoria y para derivar otras propiedades estadísticas.



## 12.2. Ejemplo 1

Supongamos que  $R_1$  es absolutamente continua, con una función de densidad  $f_1$  dada. Si definimos  $R_2 = R_1^2$ , deseamos encontrar la función de distribución o densidad de  $R_2$ .

### 12.2.1. Método de la función de distribución

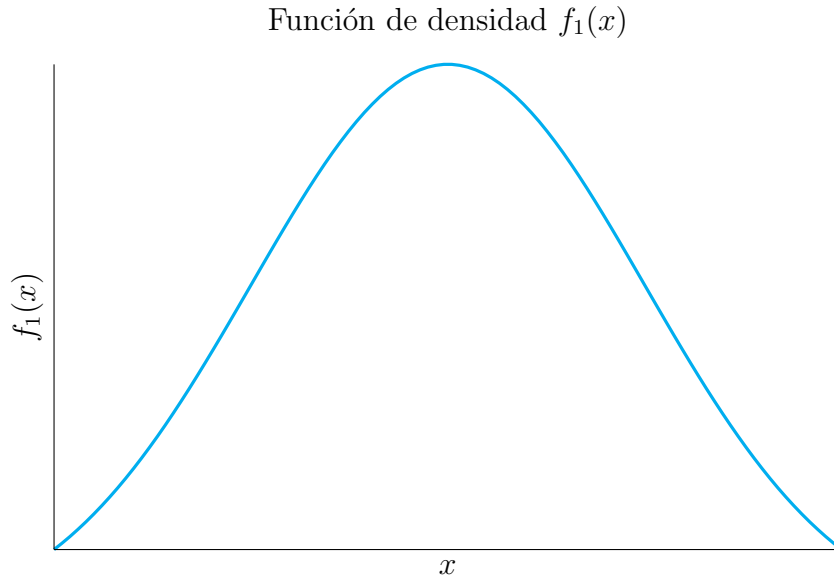
Para determinar la función de distribución  $F_2$  de  $R_2$ , se expresa el evento  $\{R_2 \leq y\}$  en términos de  $R_1$ :

- Para  $y < 0$ ,  $F_2(y) = P\{R_2 \leq y\} = 0$ .
- Si  $y \geq 0$ , entonces  $R_2 \leq y$  equivale a  $-\sqrt{y} \leq R_1 \leq \sqrt{y}$ :

$$P\{R_2 \leq y\} = P\{-\sqrt{y} \leq R_1 \leq \sqrt{y}\} = \int_{-\sqrt{y}}^{\sqrt{y}} f_1(x) dx$$

## 12.3. Gráfica de la función de densidad de probabilidad $f_1(x)$

Esta gráfica muestra la función de densidad de probabilidad  $f_1(x)$  de la variable aleatoria  $R_1$ . Representa cómo se distribuyen las probabilidades en diferentes valores de  $x$ .



## 12.4. Gráfica de la función de distribución acumulativa $F_2(y)$

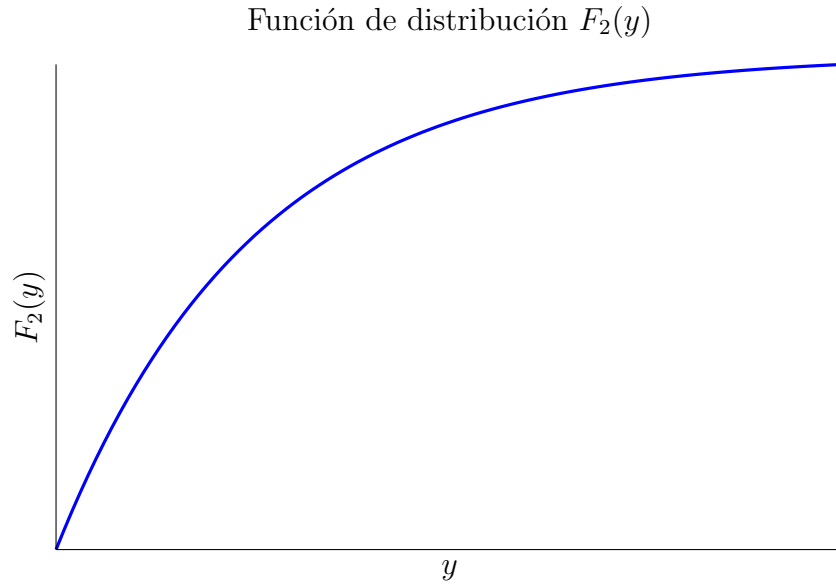
La gráfica de  $F_2(y)$  representa la probabilidad de que la variable aleatoria  $R_2$  sea menor o igual a un valor  $y$ , definida como:

$$F_2(y) = P(R_2 \leq y)$$

- Para  $y < 0$ :  $F_2(y) = 0$ , ya que  $R_2$  es no negativa.

- Para  $y \geq 0$ :  $F_2(y)$  aumenta de 0 a 1, indicando que a medida que  $y$  crece, hay más probabilidades de que  $R_2$  sea menor o igual a  $y$ .

Ambas gráficas son esenciales para entender la transformación y distribución de variables aleatorias en relación con funciones aplicadas.



## 12.5. Ejemplo 2

Hallar la función de distribución si la función de densidad de probabilidad de la variable aleatoria  $X$  está dada por:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0 \\ 0, & \text{si } x < 0 \end{cases}$$

Donde  $\lambda$  es una constante.

Función de distribución acumulativa.

La función de distribución acumulativa  $F_X(x)$  está dada por:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt.$$

**Caso 1:**  $x < 0$

Para  $x < 0$ , la función de densidad  $f_X(x) = 0$ , por lo que:

$$F_X(x) = 0, \quad \text{para } x < 0.$$

**Caso 2:**  $x \geq 0$

Para  $x \geq 0$ , calculamos la integral:

$$F_X(x) = \int_0^x \lambda e^{-\lambda t} dt.$$

$$\int \lambda e^{-\lambda t} dt = -e^{-\lambda t}.$$

Evaluando la integral en  $t = 0$  y  $t = x$ , se tiene:

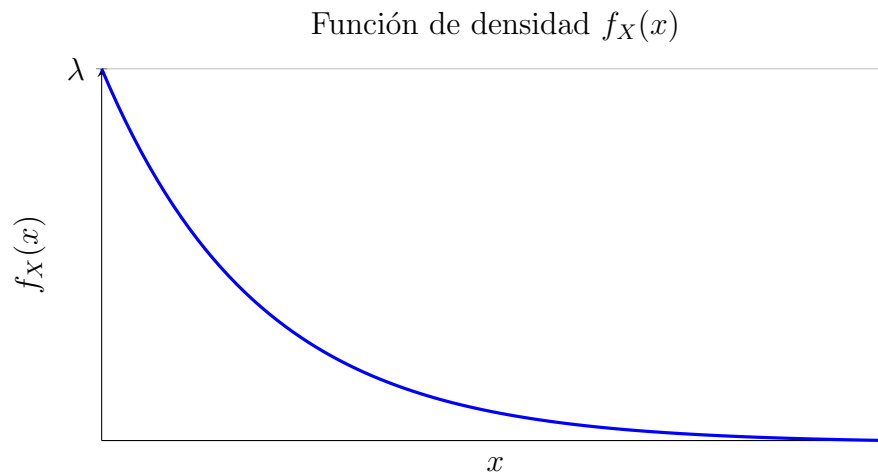
$$F_X(x) = [-e^{-\lambda t}]_0^x = -e^{-\lambda x} + 1 = 1 - e^{-\lambda x}, \quad \text{para } x \geq 0.$$

## 12.6. Resultado final

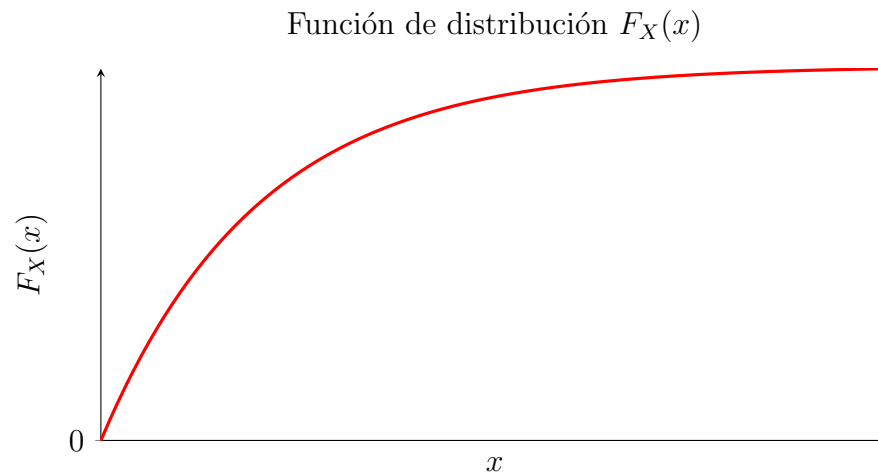
La función de distribución acumulativa es:

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

## 12.7. Gráfica de la función de densidad $f_X(x)$



## 12.8. Gráfica de la función de distribución $F_X(x)$



## 13. Funciones de densidad conjunta

### 13.1. Función de distribución conjunta

La función de distribución conjunta de dos variables aleatorias  $R_1$  y  $R_2$  se define como:

$$F_{12}(x, y) = P(R_1 \leq x, R_2 \leq y).$$

Si las variables son absolutamente continuas, existe una función  $f_{12}(x, y)$ , llamada **densidad conjunta** de  $R_1$  y  $R_2$ , tal que:

$$F_{12}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{12}(u, v) dv du.$$

### 13.2. Extensión a $n$ dimensiones

En el caso general de  $n$  variables aleatorias  $R_1, R_2, \dots, R_n$  definidas en un mismo espacio de probabilidad, la función de distribución conjunta se expresa como:

$$F_{12\dots n}(x_1, x_2, \dots, x_n) = P(R_1 \leq x_1, R_2 \leq x_2, \dots, R_n \leq x_n).$$

Si estas variables son absolutamente continuas, existe una función de densidad conjunta  $f_{12\dots n}(x_1, x_2, \dots, x_n)$  tal que:

$$F_{12\dots n}(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{12\dots n}(u_1, u_2, \dots, u_n) du_1 \dots du_n.$$

De manera análoga al caso de dos dimensiones, la probabilidad de un conjunto  $B$  en el espacio  $n$ -dimensional se calcula integrando la densidad conjunta sobre  $B$ :

$$P\{(R_1, R_2, \dots, R_n) \in B\} = \int_B f_{12\dots n}(x_1, x_2, \dots, x_n) dV.$$

### 13.3. Ejemplo

Consideremos la función de densidad conjunta  $f_{XY}(x, y)$  definida como:

$$f_{XY}(x, y) = \begin{cases} e^{-(x+y)} & \text{si } x \geq y \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

Calcular la probabilidad  $P(X \geq Y \geq 2)$ .

### 13.4. Región de integración

La condición  $X \geq Y \geq 2$  implica que:

$Y$  debe estar en el rango  $[2, X]$

y

$X$  debe ser mayor o igual a 2.

Esto significa que la región de integración se encuentra en el primer cuadrante del plano  $XY$ , y se restringe a  $X$  que varía desde 2 hasta  $\infty$ , y para cada  $X$ ,  $Y$  varía desde 2 hasta  $X$ .

### 13.5. Cálculo de la probabilidad

La probabilidad se calcula como:

$$P(X \geq Y \geq 2) = \int_2^\infty \int_2^x f_{XY}(x, y) dy dx.$$

Sustituyendo  $f_{XY}(x, y)$ :

$$P(X \geq Y \geq 2) = \int_2^\infty \int_2^x e^{-(x+y)} dy dx.$$

### 13.6. Integración interna

Primero, realizamos la integración con respecto a  $y$ :

$$\int_2^x e^{-(x+y)} dy = e^{-x} \int_2^x e^{-y} dy.$$

Calculamos la integral:

$$\int e^{-y} dy = -e^{-y} + C.$$

Evalando desde 2 hasta  $x$ :

$$\int_2^x e^{-y} dy = [-e^{-y}]_2^x = -e^{-x} + e^{-2} = e^{-2} - e^{-x}.$$

Entonces, la integral interna se convierte en:

$$\int_2^x e^{-(x+y)} dy = e^{-x}(e^{-2} - e^{-x}) = e^{-x-2} - e^{-2x}.$$

### 13.7. Integración externa

Ahora sustituimos la integral interna en la integral externa:

$$P(X \geq Y \geq 2) = \int_2^\infty (e^{-x-2} - e^{-2x}) dx.$$

Calculamos ambas integrales:

1. Para  $\int_2^\infty e^{-x-2} dx$ :

$$\int e^{-x-2} dx = e^{-2} \int e^{-x} dx = e^{-2}(-e^{-x}) = -e^{-x-2}.$$

Evaluando de 2 a  $\infty$ :

$$\left[-e^{-x-2}\right]_2^\infty = 0 - (-e^{-4}) = e^{-4}.$$

2. Para  $\int_2^\infty e^{-2x} dx$ :

$$\int e^{-2x} dx = -\frac{1}{2}e^{-2x}.$$

Evaluando de 2 a  $\infty$ :

$$\left[-\frac{1}{2}e^{-2x}\right]_2^\infty = 0 - \left(-\frac{1}{2}e^{-4}\right) = \frac{1}{2}e^{-4}.$$

### 13.8. Resultado final

Sumando ambas partes, obtenemos:

$$P(X \geq Y \geq 2) = e^{-4} - \frac{1}{2}e^{-4} = \frac{1}{2}e^{-4}.$$

Por lo tanto, la probabilidad buscada es:

$$P(X \geq Y \geq 2) = \frac{1}{2}e^{-4}.$$

## 14. Función de probabilidad de un evento conjunto.

Si  $R_1$  y  $R_2$  son dos variables aleatorias definidas en el mismo espacio de probabilidad, es importante entender la relación entre la caracterización de las variables aleatorias de manera individual y su caracterización simultánea. A continuación, se plantean dos problemas clave:

1. Si  $(R_1, R_2)$  es absolutamente continua, ¿son  $R_1$  y  $R_2$  absolutamente continuas? Si es así, ¿cómo se pueden encontrar las densidades individuales de  $R_1$  y  $R_2$  en términos de las densidades conjuntas?
2. Dados  $R_1$  y  $R_2$  (individualmente) absolutamente continuas, ¿es  $(R_1, R_2)$  absolutamente continua? Si es así, ¿se puede derivar la densidad conjunta a partir de la densidad individual?

### 14.1. Problema 1: De la información simultánea a la individual

Para obtener información individual a partir de información conjunta, es necesario sumar las probabilidades a lo largo de una fila o columna. Por ejemplo, en un grupo de 14 personas con una distribución de edad y peso, la cantidad de personas entre 20 y 25 años se obtiene sumando los números en la columna correspondiente.

Si  $R_1$  y  $R_2$  son discretas, la función de probabilidad conjunta de  $R_1$  y  $R_2$  se define como:

$$P_{12}(x, y) = P(R_1 = x, R_2 = y)$$

Si los posibles valores de  $R_1$  son  $1, 2, \dots$ , entonces la probabilidad de  $R_1$  se puede expresar como:

$$P(R_1 = x) = \sum_y P_{12}(x, y)$$

Análogamente, la probabilidad de  $R_2$  es:

$$P(R_2 = y) = \sum_x P_{12}(x, y)$$

## 14.2. Caso absolutamente continuo

Si  $(R_1, R_2)$  es absolutamente continuo con densidad conjunta  $f_{12}$ , se puede demostrar que  $R_1$  y  $R_2$  son absolutamente continuas y se pueden encontrar sus densidades marginales  $f_1$  y  $f_2$  en términos de  $f_{12}$ :

$$f_1(x) = \int_{-\infty}^{\infty} f_{12}(x, y) dy$$

$$f_2(y) = \int_{-\infty}^{\infty} f_{12}(x, y) dx$$

La existencia de estas integrales garantiza que las variables aleatorias individuales también son absolutamente continuas.

## 14.3. Densidades marginales y cálculo

El proceso de obtener las densidades marginales a partir de la densidad conjunta se conoce como cálculo de densidades marginales, ya que implica sumar (integrar) sobre una dimensión para obtener información sobre la otra. Para la densidad conjunta  $f_{12}(x, y)$ , las densidades marginales se definen como:

$$f_1(x) = \int f_{12}(x, y) dy$$

$$f_2(y) = \int f_{12}(x, y) dx$$

Esto es útil en aplicaciones donde se requiere entender la distribución de una variable aleatoria en función de la otra.

## 14.4. Ejemplo

Consideremos la densidad conjunta:

$$f_{12}(x, y) = \begin{cases} 8xy & \text{si } 0 \leq y \leq x \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

### 14.4.1. Densidad marginal de $R_1$

La densidad marginal de  $R_1$  se calcula como:

$$f_1(x) = \int_0^x f_{12}(x, y) dy = \int_0^x 8xy dy$$

Resolviendo la integral:

$$f_1(x) = 8x \left[ \frac{y^2}{2} \right]_0^x = 8x \cdot \frac{x^2}{2} = 4x^3 \quad \text{para } 0 \leq x \leq 1.$$

Por lo tanto, la densidad marginal de  $R_1$  es:

$$f_1(x) = 4x^3.$$

### 14.4.2. Densidad marginal de $R_2$

La densidad marginal de  $R_2$  se calcula como:

$$f_2(y) = \int_y^1 f_{12}(x, y) dx = \int_y^1 8xy dx$$

Resolviendo esta integral:

$$f_2(y) = 8y \left[ \frac{x^2}{2} \right]_y^1 = 8y \left( \frac{1}{2} - \frac{y^2}{2} \right) = 4y(1 - y^2) \quad \text{para } 0 \leq y \leq 1.$$

Por lo tanto, la densidad marginal de  $R_2$  es:

$$f_2(y) = 4y(1 - y^2).$$

## 14.5. Gráficas

Consideramos la función de densidad conjunta  $f_{12}(x, y) = 8xy$  para  $0 \leq y \leq x \leq 1$ . A partir de esta función, podemos calcular las densidades marginales.

### Densidad marginal de $R_1$

La densidad marginal de  $R_1$  se calcula como:

$$f_1(x) = \int_0^x f_{12}(x, y) dy = \int_0^x 8xy dy$$

Resolviendo la integral:

$$f_1(x) = 8x \left[ \frac{y^2}{2} \right]_0^x = 8x \cdot \frac{x^2}{2} = 4x^3$$

Para  $0 \leq x \leq 1$ .



## Densidad marginal de $R_2$

La densidad marginal de  $R_2$  se calcula como:

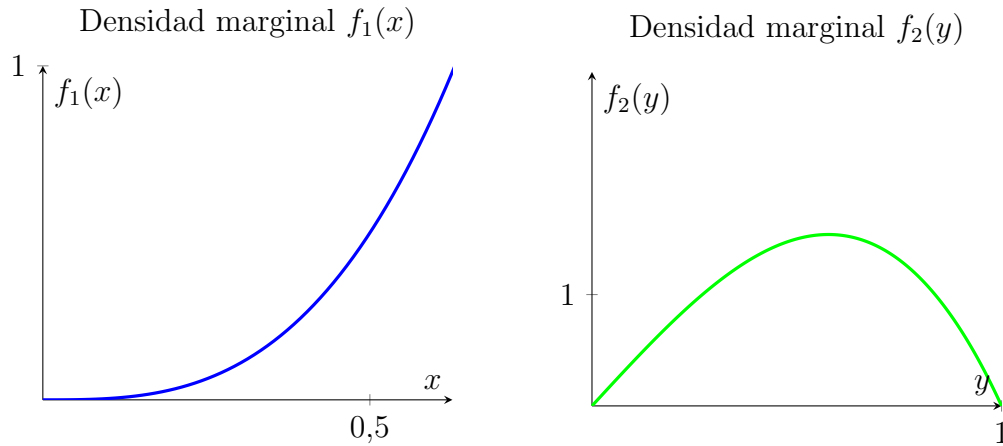
$$f_2(y) = \int_y^1 f_{12}(x, y) dx = \int_y^1 8xy dx$$

Resolviendo la integral:

$$f_2(y) = 8y \left[ \frac{x^2}{2} \right]_y^1 = 8y \left( \frac{1}{2} - \frac{y^2}{2} \right) = 4y(1 - y^2)$$

Para  $0 \leq y \leq 1$ .

## Gráficas de las densidades marginales



Para ver un script de ejemplo de densidad marginal dirigirse al Ejemplo 5.

## 15. Independencia de variables aleatorias

Hasta ahora hemos considerado la independencia entre eventos. Esta analogía se puede aplicar para definir la independencia entre variables aleatorias. Sean  $(X_1, X_2, \dots, X_n)$  variables aleatorias, estas se consideran independientes si el conocimiento que se tiene sobre cualquier  $X_i$  no cambia el que se tiene sobre otras  $X_j$  (con  $j \neq i$ ).

Formalmente,  $X_1, X_2, \dots, X_n$  son independientes si para todos los conjuntos  $B_1, B_2, \dots, B_n$  contenidos en  $\mathbb{R}$  se cumple que:

$$P\{X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n\} = P\{X_1 \in B_1\} \cdot P\{X_2 \in B_2\} \cdots P\{X_n \in B_n\}$$

Bajo el supuesto de independencia, como cada  $X_i$  está asociada a una función de densidad  $f_i(x_i)$ , se cumple que:

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdots f_n(x_n)$$

Por lo tanto, la probabilidad de que el vector  $(X_1, \dots, X_n)$  esté en el conjunto  $B$  se puede expresar como:

$$P\{(X_1, X_2, \dots, X_n) \in B\} = \int \cdots \int_B f_1(x_1) f_2(x_2) \cdots f_n(x_n) dx_1 dx_2 \cdots dx_n$$

donde  $B = B_1 \cap B_2 \cap \dots \cap B_n$  es el conjunto de interés y se asume que su medida es nula.

## 16. Valor esperado

Estamos considerando un promedio ponderado de los valores posibles de la variable aleatoria  $R$ , donde los pesos son las probabilidades de que  $R$  asuma esos valores. Esto sugiere la siguiente definición.

Sea  $R$  una variable aleatoria discreta simple, es decir, una variable aleatoria que toma solo un número finito de valores posibles. Definimos la expectativa (también llamada valor esperado, promedio o media) de  $R$  como:

$$E(R) = \sum xP\{R = x\} \quad (3.1.1)$$

Dado que  $R$  es simple, esta es una suma finita y no hay problemas de convergencia. En particular, si  $R$  es constantemente igual a  $c$ , entonces:

$$E(R) = cP\{R = c\} = c \quad (3.1.2)$$

Nota que si  $R$  toma los valores  $x_1, x_2, \dots, x_n$ , cada uno con probabilidad  $\frac{1}{n}$ , entonces:

$$E(R) = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

como intuitivamente esperaríamos. En este caso, cada  $x$  tiene el mismo peso, es decir,  $\frac{1}{n}$ .

Ahora tenemos el problema de extender la definición a variables aleatorias más generales.

Si  $R$  es una variable aleatoria discreta arbitraria, la elección natural para  $E(R)$  es de nuevo  $\sum xP\{R = x\}$ , siempre que la suma tenga sentido (el Teorema 1 lo precisará).

Similarmente, si  $R_1$  es discreta y  $R_2 = g(R_1)$ , como  $R_2$  también es discreta, tenemos:

$$E(R_2) = \sum yP\{R_2 = y\} \quad (3.1.3)$$

Sin embargo, si  $x_1, x_2, \dots$  son los valores de  $R_1$ , entonces con probabilidad  $P\{R_1 = x\}$ , tenemos  $R_2 = g(x)$ . Por lo tanto, si nuestra definición de expectativa es válida, deberíamos tener la siguiente expresión alternativa para  $E(R_2)$ :

$$E(R_2) = E[g(R_1)] = \sum g(x)P\{R_1 = x\} \quad (3.1.4)$$

Si  $R_1$  es absolutamente continua, este enfoque falla completamente, ya que  $P\{R_1 = x\} = 0$  para todos los  $x$ . Sin embargo, podemos obtener una idea de cómo calcular  $E(R)$  y  $E[g(R_1)]$  cuando  $R_1$  es absolutamente continua, haciendo una aproximación discreta. Si dividimos la recta real en intervalos  $(x, x + dx]$ , entonces, aproximadamente, la probabilidad de que  $x < R_1 \leq x + dx$  es  $f_R(x)dx$ .

Si  $R_1$  cae en este intervalo,  $g(R_1)$  es aproximadamente  $g(x)$ , al menos si  $g$  es continua. Así, una aproximación a  $E(R)$  debería ser:

$$E[g(R_1)] \approx \int g(x)f_R(x) dx$$

## 17. Momento de una variable aleatoria

Si  $R$  es una variable aleatoria en un espacio de probabilidad dado, el  $k$ -ésimo momento de  $R$  ( $k > 0$ , no necesariamente un entero) se define como:

$$M_k = E(R^k) \quad \text{si la esperanza existe}$$

Así que:

$$M_k = \sum x^k P\{R = x\} \quad \text{si } R \text{ es discreta}$$

$$M_k = \int x^k f_R(x) dx \quad \text{si } R \text{ es absolutamente continua}$$

El primer momento  $M_1$  es simplemente  $E(R)$ , la esperanza de  $R$ , a menudo escrita como  $m$  y llamada la media de  $R$ . Si  $R$  tiene densidad  $f_R$ ,  $m$  puede ser considerado como la abscisa del centroide de la región en el plano entre el eje  $x$  y el gráfico de  $f_R$  (ver Figura 3.2.1).

Para ver esto, observa que el momento total de la región sobre la línea  $x = m$  es:

$$\int (x - m)f_R(x) dx = 0$$

La expectativa de  $R$  es una medida de tendencia central en el sentido de que el promedio aritmético de  $n$  observaciones independientes de  $R$  converge (en un sentido que será precisado más adelante) a  $E(R)$ .

El  $k$ -ésimo momento central de  $R$  ( $k > 0$ ) se define como:

$$\mu_k = E[(R - m)^k] = \sum (x - m)^k P\{R = x\} \quad \text{si } m \text{ es finito y la esperanza existe}$$

$$\mu_k = \int (x - m)^k f_R(x) dx \quad \text{si } R \text{ es absolutamente continua}$$

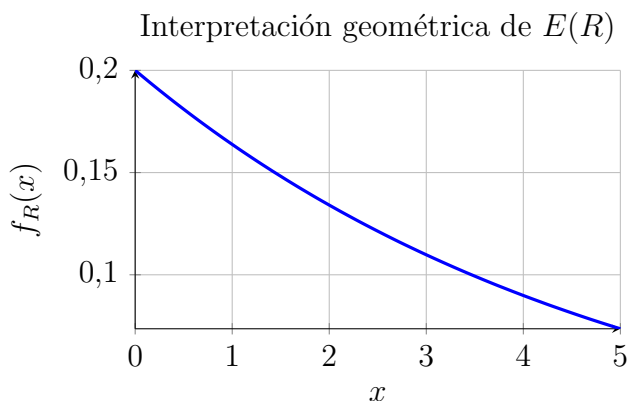
Observa que:

$$\mu_1 = E(R - m) = 0$$

El segundo momento central  $\mu_2 = E[(R - m)^2]$  se llama la varianza de  $R$ , escrita como  $\sigma^2$ ,  $\sigma^2(R)$  o  $\text{Var}(R)$ .  $\sigma$  (la raíz cuadrada positiva de  $\sigma^2$ ) se llama la desviación estándar de  $R$ . Nota que si  $R$  tiene media finita, entonces, dado que  $(R - m)^2 \geq 0$ ,  $\text{Var}(R)$  siempre existe; puede ser infinita.

Si  $R$  tiene densidad  $f_R$ , la varianza de  $R$  puede ser considerada como el momento de inercia de la región en el plano entre el eje  $x$  y el gráfico de  $f_R$  respecto al eje  $x = m$ .

La varianza puede ser interpretada como una medida de dispersión. Una varianza grande corresponde a una alta probabilidad de que  $R$  caiga lejos de su media, mientras que una varianza pequeña indica que  $R$  probablemente estará cerca de su media (ver Figura 3.2.2). Haremos una declaración cuantitativa sobre esto (la desigualdad de Chebyshev) en la Sección 3.7.



## 18. Algunas propiedades de $\mathbb{E}$

A continuación, se presentan algunas propiedades importantes del valor esperado:

### 1. Existencia del valor esperado:

Si  $\mathbb{E}[X] = \infty$  o  $\mathbb{E}[X] = -\infty$ , entonces el valor esperado de  $X$  no existe.

### 2. Minimización del error cuadrático:

Si  $\mathbb{E}[X] = b$ , entonces  $\mathbb{E}[(X - b)^2]$  es mínimo.

**3. Propiedad de linealidad:** Sean  $X$  y  $Y$  variables aleatorias con expectativas finitas. Entonces, se cumple que:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

**3.1 Desigualdad de esperanza:** Si  $g(x) \geq h(x)$ , entonces:

$$\mathbb{E}[g(X)] \geq \mathbb{E}[h(X)]$$

**3.2 Transformaciones lineales:** Si  $a, b, c \in \mathbb{R}$ , se cumple que:

$$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$$

### 4. Independencia de variables aleatorias

Sean  $X$  y  $Y$  variables aleatorias que cumplan  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$  o bien  $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$ . Esto implica que, si  $X$  y  $Y$  son independientes, el valor esperado del producto es igual al producto de los valores esperados.

Para ver un script de ejemplo sobre el cálculo del valor esperado y la verificación de algunas propiedades del mismo dirigirse al Ejemplo 6.

## 19. Teorema binomial

El teorema binomial proporciona una herramienta fundamental para el análisis de momentos y centralización de variables aleatorias.

**1. Relación entre momentos centrales y momentos ordinarios:** Los momentos centrales  $\beta_1, \dots, \beta_n$  (con  $n \geq 2$ ) pueden ser obtenidos a partir de los momentos  $\alpha_1, \dots, \alpha_n$ , siempre que  $\alpha_1, \dots, \alpha_{n-1}$  sean finitos y  $\alpha_n$  exista.

Para ver esto, expandimos  $(R - m)^n$  utilizando el teorema binomial:

$$(R - m)^n = \sum_{k=0}^n \binom{n}{k} R^k (-m)^{n-k}$$

De aquí, podemos escribir la expectativa del momento central:

$$\beta_n = \mathbb{E}[(R - m)^n] = \sum_{k=0}^n \binom{n}{k} (-m)^{n-k} \mathbb{E}[R^k]$$

Notamos que, dado que  $\alpha_1, \dots, \alpha_{n-1}$  son finitos, no pueden aparecer términos de la forma  $+\infty$  en la suma. Por lo tanto, podemos tomar la expectativa término a término, gracias a la propiedad de linealidad del valor esperado.

Este resultado se aplica con mayor frecuencia cuando  $n = 2$ . Si  $R$  tiene media finita, entonces  $\mathbb{E}[R^2]$  siempre existe, dado que  $R^2 > 0$ . En este caso, se puede escribir:

$$(R - m)^2 = R^2 - 2mR + m^2$$

Por lo tanto, la varianza de  $R$  se expresa como:

$$\text{Var}(R) = \mathbb{E}[R^2] - 2m\mathbb{E}[R] + m^2$$

Finalmente, esto se simplifica a:

$$\sigma^2 = \mathbb{E}[R^2] - [\mathbb{E}[R]]^2$$

donde  $\sigma^2$  es la varianza de  $R$ , que se interpreta como la media de los cuadrados menos el cuadrado de la media.

## 20. Correlación

Si  $X_1$  y  $X_2$  son variables aleatorias en un espacio de probabilidad dado, podemos definir momentos conjuntos asociados con  $X_1$  y  $X_2$ :

$$\alpha_{jk} = \mathbb{E}(X_1^j X_2^k), \quad \forall j, k > 0$$

y momentos centrales conjuntos:

$$\beta_{jk} = \mathbb{E}[(X_1 - m_1)^j (X_2 - m_2)^k], \quad m_1 = \mathbb{E}(X_1), m_2 = \mathbb{E}(X_2)$$

Para el caso específico de  $\beta_{11}$ :

$$\beta_{11} = \mathbb{E}[(X_1 - m_1)(X_2 - m_2)] = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2),$$

lo cual se llama la covarianza de  $X_1$  y  $X_2$ , escrita como  $\text{Cov}(X_1, X_2)$ .

Se asume que  $\mathbb{E}(X_1)$  y  $\mathbb{E}(X_2)$  son finitas, y que  $\mathbb{E}(X_1 X_2)$  existe; por lo tanto, la covarianza de  $X_1$  y  $X_2$  está bien definida.

**Teorema 1:** Si  $X_1$  y  $X_2$  son independientes, entonces  $\text{Cov}(X_1, X_2) = 0$ , pero no al revés.

**Teorema 2 (Desigualdad de Schwarz):** Supongamos que  $\mathbb{E}(X_1)$  y  $\mathbb{E}(X_2^2)$  son finitas. Entonces  $\mathbb{E}(X_1 X_2)$  es finita, y

$$|\mathbb{E}(X_1 X_2)|^2 \leq \mathbb{E}(X_1^2)\mathbb{E}(X_2^2).$$

**Definición (Coeficiente de correlación):**

Asumamos que  $\mathbb{E}(X_1^2)$  y  $\mathbb{E}(X_2^2)$  son finitas y, además, que las varianzas  $\sigma_1^2$  y  $\sigma_2^2$  de  $X_1$  y  $X_2$  son  $> 0$ . Definimos el coeficiente de correlación de  $X_1$  y  $X_2$  como:

$$\rho(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sigma_1 \sigma_2}.$$

Nota: Si  $x_1, x_2$  son independientes, entonces  $\rho(x_1, x_2) = 0$ , lo contrario no siempre se cumple.

**Teorema 3:**  $-1 \leq \rho(X_1, X_2) \leq 1$ .

Para ver un script de ejemplo sobre el cálculo del valor esperado y la verificación de algunas propiedades del mismo dirigirse al Ejemplo 7.

Para ver un script de ejemplo sobre el cálculo del coeficiente de correlación del mismo dirigirse al Ejemplo 8.

## 21. Funciones de densidad condicional

Consideramos dos variables aleatorias  $R_1$  y  $R_2$  con una densidad conjunta  $f(x, y)$ . La probabilidad condicional  $P\{R_2 \in B \mid R_1 = x\}$  se define como:

$$P\{R_2 \in B \mid R_1 = x\} = \int_B h(y|x) dy$$

donde la densidad condicional  $h(y|x)$  está dada por:

$$h(y|x) = \frac{f(x, y)}{f(x)}$$

La densidad condicional solo está definida si  $f(x) > 0$ . Se puede verificar que se cumple el teorema de la probabilidad total:

$$P\{R_1 \in A, R_2 \in B\} = \int f(x, y) dx dy = \int f_1(x) P(B) dx$$

La fórmula  $f(x, y) = f(x)h(y|x)$  se interpreta de dos maneras:

1. Si  $(R_1, R_2)$  tiene densidad  $f(x, y)$ , la probabilidad condicional es:

$$P(B) = \int h(y|x) dy$$

2. Si  $R_1$  tiene densidad  $f(x)$  y  $R_2$  tiene densidad condicional  $h(y|x)$ , entonces  $(R_1, R_2)$  tiene densidad:

$$f(x, y) = f(x)h(y|x)$$

## 22. Valor esperado condicional

Dada una densidad conjunta de las variables aleatorias  $R_1$  y  $R_2$ , la probabilidad condicional se puede expresar como:

$$P\{R_2 \in B \mid R_1 = x\} = \int_B h(y|x) dy$$

Para cada  $x$ , definimos una medida de probabilidad  $P_2$  sobre los subconjuntos de Borel de  $E_1$ . La expectativa condicional de  $R_2$  dado que  $R_1 = x$  se define como:

$$E(R_2 \mid R_1 = x) = \int_{-\infty}^{\infty} y h(y|x) dy$$

Si  $g$  es una función continua en  $E_1$ , la expectativa condicional también se puede expresar como:

$$E[g(R_2) \mid R_1 = x] = \int_{-\infty}^{\infty} g(y) h(y|x) dy$$

## 23. Ejemplos

- Ejemplo 1.
- Ejemplo 2.
- Ejemplo 3.
- Ejemplo 4.
- Ejemplo 5.
- Ejemplo 6.
- Ejemplo 7.
- Ejemplo 8.

### 23.1. Repositorio en GitHub

Codigos de ejemplo.

## 24. Referencias

[1] R. B. Ash, *Basic Probability Theory*, Courier Corporation, 2008.