

14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca

Clase 3

Variables Aleatorias y Distribuciones Conjuntas

Teoría

Módulos oficiales del MicroMaster:

M2 – Joint Distributions; M3 – Joint, Marginal & Conditional Distributions; Summarizing & Describing Data

B.Sc. Cristian Lazo Quispe

✉ clazoq@uni.pe

[in linkedin.com/in/cristian-lazo-quispe](https://www.linkedin.com/in/cristian-lazo-quispe) · github.com/CristianLazoQuispe

Material del *Advanced Program in Data Science & Global Skills*, diseñado y desarrollado por **BREIT (Brescia Institute of Technology)**, iniciativa de **Aporta** —plataforma de innovación e impacto social del **Grupo Breca**— en alianza con el **MIT IDSS** (Institute for Data, Systems, and Society).

Basado en MIT 14.310x (E. Duflo & S. Ellison) y en diversas fuentes abiertas (ver bibliografía al final), bajo licencia CC BY-NC-SA 4.0.

11 de agosto de 2026

Índice

1. De qué trata esta clase	2
2. Variable aleatoria	2
3. Variables discretas: la función de probabilidad (pmf)	3
4. Variables continuas: la función de densidad (pdf)	3
4.1. Una densidad muy usada: la exponencial	4
5. La función de distribución acumulada (cdf)	4
6. Distribuciones conjuntas	5
6.1. La región de soporte R_{XY}	6
7. Distribuciones marginales	7
8. Distribuciones condicionales	7
9. Funciones de variables aleatorias	8
10. En R: muestrear y visualizar una densidad	9
Resumen de la clase	9

1. De qué trata esta clase

En la Clase 1 aprendimos a asignar probabilidades a eventos. Pero la mayoría de las veces, lo que de verdad nos interesa de un experimento no es el evento crudo, sino una **característica numérica** de él: los años de educación de una persona, el ingreso mensual de un hogar, el gasto en salud, el tamaño del hogar. Esa característica numérica es una **variable aleatoria (random variable)**.

Una vez que tenemos un número asociado al azar, queremos describir *cómo se reparte* ese número: ¿qué valores toma y con qué probabilidad? Esa descripción es la **distribución**. La estudiaremos en tres formatos equivalentes (pmf, pdf, cdf) y luego daremos el salto a **dos** variables a la vez (distribución conjunta), de donde nacerán las **marginales** y las **condicionales**.

Una variable aleatoria es simplemente una *regla que traduce resultados del azar en números*. La distribución es el “mapa” de cómo esos números se reparten la probabilidad total, que siempre suma (o integra) 1.

2. Variable aleatoria

Definición 2.1 (Variable aleatoria). Una **variable aleatoria (random variable)** X es una función de valor real cuyo dominio es el espacio muestral S . Es decir, a cada resultado $s \in S$ le asigna un número $X(s) \in \mathbb{R}$.

La probabilidad definida sobre S “induce” una distribución sobre los valores de X : escribimos $\mathbb{P}(X = x)$, $\mathbb{P}(a \leq X \leq b)$, etc., heredando el sentido de la probabilidad original.

Definición 2.2 (Discreta vs. continua). X es **discreta** si toma un número finito o infinito numerable de valores (p.ej. $0, 1, 2, \dots$). X es **continua** si puede tomar cualquier valor en algún intervalo de la recta real (acotado o no).

Ejemplo 2.1 (Discreta vs. continua en ciencias sociales). ■ *Tamaño del hogar* (número de personas): $0, 1, 2, 3, \dots \Rightarrow$ **discreta**.

- *Número de hijos* de una mujer: $0, 1, 2, \dots \Rightarrow$ **discreta**.
- *Ingreso mensual* de un hogar (en soles): cualquier valor $\geq 0 \Rightarrow$ **continua** (la modelamos así aunque el dinero tenga céntimos).
- *Tiempo* hasta que un desempleado encuentra trabajo: $\geq 0 \Rightarrow$ **continua**.

“Variable aleatoria” es un nombre engañoso: no es una variable ni es aleatoria, es una *función*. Lo aleatorio es el resultado s ; X solo lo traduce a un número.

3. Variables discretas: la función de probabilidad (pmf)

Definición 3.1 (Función de probabilidad (pmf)). Si X es discreta, su **función de probabilidad (probability mass function, pmf)** es la función $p(x) = f_X(x)$ definida por

$$f_X(x) = \mathbb{P}(X = x) \quad \text{para todo } x \in \mathbb{R}.$$

Propiedades de una pmf

Para que f_X sea una pmf válida debe cumplir, sobre los valores posibles $x_1, x_2, \dots$:

$$(i) \quad 0 \leq f_X(x_i) \leq 1, \quad (ii) \quad \sum_i f_X(x_i) = 1.$$

Y la probabilidad de un evento A se obtiene sumando: $\mathbb{P}(X \in A) = \sum_{x_i \in A} f_X(x_i)$.

Ejemplo 3.1 (pmf del número de adultos que trabajan). En cierta zona, el número X de adultos que trabajan en un hogar (de a lo más 3 adultos) tiene la pmf:

x	0	1	2	3
$f_X(x)$	0.10	0.35	0.40	0.15

(a) ¿Es una pmf válida? Sumamos: $0.10 + 0.35 + 0.40 + 0.15 = 1.00$ y todos los valores están en $[0, 1]$. Sí lo es.

(b) ¿Probabilidad de que trabaje al menos un adulto?

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - 0.10 = 0.90.$$

(c) ¿Probabilidad de que trabajen 1 o 2? $0.35 + 0.40 = 0.75$.

Nota. Al graficar una pmf se dibuja una línea vertical (un “palo”) sobre cada valor posible, de altura igual a su probabilidad. Cada punto se llama *masa de probabilidad*.

4. Variables continuas: la función de densidad (pdf)

Para una variable continua *no* tiene sentido preguntar $\mathbb{P}(X = x)$: ese valor es 0 para cualquier x (hay un continuo de valores posibles). Lo que se da es una **densidad** que asigna probabilidad a *intervalos* vía integrales.

Definición 4.1 (Función de densidad (pdf)). X es **continua** si existe una función no negativa f_X , su **función de densidad (probability density function, pdf)**, tal que para todo intervalo $A = [a, b]$,

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx.$$

Propiedades de una pdf

$$(i) \quad f_X(x) \geq 0, \quad (ii) \quad \int_{-\infty}^{\infty} f_X(x) dx = 1.$$

La pmf *suma* a 1; la pdf *integra* a 1. La probabilidad de un intervalo es el **área bajo la curva** en ese intervalo.

$f_X(x)$ *no* es una probabilidad: es una densidad (probabilidad por unidad de x). Puede tomar valores mayores que 1. Lo que nunca pasa de 1 es el *área* total, que vale exactamente 1. También: $\mathbb{P}(X = x) = 0$ para todo x , así que $\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b)$; los extremos no aportan área.

Ejemplo 4.1 (Hallar la constante de normalización c). El número de años de educación X (reescalado al intervalo $[0, 4]$) de cierta población se modela con la densidad

$$f_X(x) = \begin{cases} cx & 0 \leq x \leq 4, \\ 0 & \text{en otro caso.} \end{cases}$$

(a) ¿Cuánto vale c ? Imponemos que integre a 1:

$$\int_0^4 cx dx = c \frac{x^2}{2} \Big|_0^4 = c \frac{16}{2} = 8c \stackrel{!}{=} 1 \implies \boxed{c = \frac{1}{8}}.$$

(b) ¿Probabilidad de tener entre 2 y 3 (en la escala)?

$$\mathbb{P}(2 \leq X \leq 3) = \int_2^3 \frac{1}{8}x dx = \frac{1}{8} \cdot \frac{x^2}{2} \Big|_2^3 = \frac{1}{16}(9 - 4) = \frac{5}{16} \approx 0.3125.$$

4.1. Una densidad muy usada: la exponencial

La densidad **exponencial** modela tiempos de espera (p.ej. tiempo hasta encontrar empleo, duración de un episodio de pobreza). Su pdf es

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad (\lambda > 0).$$

Integra a 1: $\int_0^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_0^{\infty} = 1$. La Figura 1 muestra su forma para $\lambda = 1$.

5. La función de distribución acumulada (cdf)

Definición 5.1 (Función de distribución acumulada (cdf)). La **función de distribución acumulada (cumulative distribution function, cdf)** de *cualquier* variable aleatoria X (discreta o continua) es

$$F_X(x) = \mathbb{P}(X \leq x).$$

La gran ventaja de la cdf es que *tiene la misma definición* sea X discreta o continua. Sus propiedades:

$$0 \leq F_X(x) \leq 1, \quad F_X \text{ es no decreciente}, \quad \lim_{x \rightarrow -\infty} F_X(x) = 0, \quad \lim_{x \rightarrow +\infty} F_X(x) = 1.$$

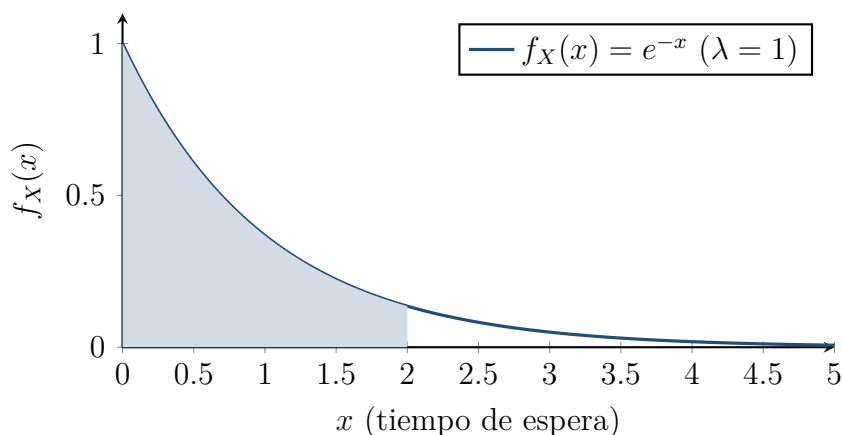


Figura 1: Densidad exponencial con $\lambda = 1$. El área sombreada es $\mathbb{P}(0 \leq X \leq 2) = 1 - e^{-2} \approx 0.865$.

Relación cdf $\leftrightarrow$ pmf/pdf

- **Discreta:** $F_X(x) = \sum_{x_i \leq x} f_X(x_i)$. La cdf es una *escalera*: salta en cada valor posible, y el tamaño del salto es la pmf.
- **Continua:** $F_X(x) = \int_{-\infty}^x f_X(t) dt$, y recíprocamente $f_X(x) = F'_X(x)$. La cdf es continua y suave.

Ejemplo 5.1 (De pdf a cdf). Para la densidad de educación $f_X(x) = \frac{1}{8}x$ en $[0, 4]$, la cdf es, para $0 \leq x \leq 4$,

$$F_X(x) = \int_0^x \frac{1}{8}t dt = \frac{1}{8} \cdot \frac{x^2}{2} = \frac{x^2}{16}.$$

Así $F_X(0) = 0$, $F_X(4) = 16/16 = 1$ (correcto), y por ejemplo $\mathbb{P}(X \leq 3) = F_X(3) = 9/16$. Además $\mathbb{P}(2 \leq X \leq 3) = F_X(3) - F_X(2) = \frac{9}{16} - \frac{4}{16} = \frac{5}{16}$, igual que antes.

6. Distribuciones conjuntas

A menudo nos interesa la relación entre *dos* variables: ingreso y educación, lluvia y rendimiento agrícola, gasto en salud y edad. Para eso necesitamos una distribución de las dos a la vez.

Definición 6.1 (Densidad conjunta). Si X e Y son continuas sobre el mismo espacio muestral, su **densidad conjunta (joint pdf)** es la superficie $f_{XY}(x, y) \geq 0$ tal que, para cualquier región A del plano xy ,

$$\mathbb{P}((X, Y) \in A) = \iint_A f_{XY}(x, y) dx dy.$$

Para variables discretas el análogo es la **pmf conjunta** $f_{XY}(x, y) = \mathbb{P}(X = x, Y = y)$.

Propiedades de la densidad conjunta

$$f_{XY}(x, y) \geq 0, \quad \iint_{\mathbb{R}^2} f_{XY}(x, y) dx dy = 1.$$

Cualquier punto aislado o curva unidimensional tiene probabilidad cero.

6.1. La región de soporte R_{XY}

Definición 6.2 (Región de soporte). La **región de soporte (support region)** R_{XY} es el conjunto de puntos (x, y) donde $f_{XY}(x, y) > 0$. Fuera de R_{XY} la densidad vale 0, así que todas las integrales se hacen *solo* sobre R_{XY} .

En la práctica, el error más común no está en integrar, sino en **poner los límites de integración**. Siempre dibuja R_{XY} antes de integrar. La Figura 2 muestra un soporte triangular típico.

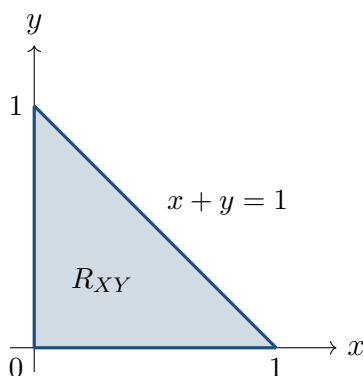


Figura 2: Región de soporte triangular $R_{XY} = \{(x, y) : x \geq 0, y \geq 0, x + y \leq 1\}$. Toda integral doble se calcula *solo* sobre esta región.

Ejemplo 6.1 (Constante de normalización en una densidad conjunta). Sobre el triángulo $R_{XY} = \{x \geq 0, y \geq 0, x + y \leq 1\}$ proponemos la densidad *constante* $f_{XY}(x, y) = c$. ¿Cuánto vale c ?

Como la densidad es constante, $\iint c dx dy = c \cdot (\text{área de } R_{XY})$. El área del triángulo es $\frac{1}{2}$, así que

$$c \cdot \frac{1}{2} = 1 \implies \boxed{c = 2}.$$

Verifiquemos con la integral doble, fijando x y dejando y ir de 0 a $1 - x$:

$$\int_0^1 \int_0^{1-x} 2 dy dx = \int_0^1 2(1-x) dx = 2 \left(x - \frac{x^2}{2} \right) \Big|_0^1 = 2 \left(1 - \frac{1}{2} \right) = 1. \quad \checkmark$$

7. Distribuciones marginales

A partir de la conjunta podemos recuperar la distribución de cada variable por separado: su **distribución marginal**. La idea es “borrar” (integrar o sumar) la otra variable.

Marginales

$$\text{Continuas: } f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx.$$

$$\text{Discretas: } f_X(x) = \sum_y f_{XY}(x, y), \quad f_Y(y) = \sum_x f_{XY}(x, y).$$

La marginal de X “no le importa” qué valor tomó Y : por eso sumamos/integramos sobre *todos* los valores de Y . Es la versión continua de la regla de probabilidad total.

Ejemplo 7.1 (Marginales del triángulo). Con $f_{XY}(x, y) = 2$ sobre $R_{XY} = \{x \geq 0, y \geq 0, x + y \leq 1\}$, la marginal de X es, para $0 \leq x \leq 1$ (porque y va de 0 a $1 - x$):

$$f_X(x) = \int_0^{1-x} 2 dy = 2(1 - x).$$

Verificamos que es una pdf válida: $\int_0^1 2(1 - x) dx = 1$ (✓). Por simetría del problema, $f_Y(y) = 2(1 - y)$ para $0 \leq y \leq 1$.

8. Distribuciones condicionales

Así como en la Clase 2 vimos la probabilidad condicional $\mathbb{P}(B | A)$, ahora queremos la distribución de una variable *dado que ya conocemos* la otra.

Definición 8.1 (Densidad condicional). La **densidad condicional de Y dado $X = x$** es

$$f_{Y|X}(y | x) = \frac{f_{XY}(x, y)}{f_X(x)}, \quad \text{siempre que } f_X(x) > 0.$$

Análogamente, $f_{X|Y}(x | y) = f_{XY}(x, y) / f_Y(y)$.

Vista como función de y (con x fijo), una condicional se comporta *exactamente* como una pdf: es ≥ 0 e integra a 1.

Ejemplo 8.1 (Condicional del triángulo). Con $f_{XY} = 2$ y $f_X(x) = 2(1 - x)$, la condicional de Y dado $X = x$ (para $0 \leq x < 1$) es

$$f_{Y|X}(y | x) = \frac{2}{2(1 - x)} = \frac{1}{1 - x}, \quad 0 \leq y \leq 1 - x.$$

¡Es una densidad **uniforme** en $[0, 1 - x]$! Tiene sentido: como la conjunta es plana, fijar $X = x$ deja a Y repartido uniformemente en lo que “queda” del triángulo. Comprobamos que integra a 1: $\int_0^{1-x} \frac{1}{1-x} dy = \frac{1-x}{1-x} = 1$ (✓).

Independencia de variables aleatorias

X e Y son **independientes** si y solo si la conjunta *se factoriza* como el producto de las marginales:

$$f_{XY}(x, y) = f_X(x) f_Y(y) \quad \text{para todo } (x, y).$$

Equivalentemente, $f_{Y|X}(y|x) = f_Y(y)$: conocer X no cambia la distribución de Y .

Ejemplo 8.2 (*¿Independientes o no?*). (a) Sea $f_{XY}(x, y) = 4xy$ en $[0, 1] \times [0, 1]$. Las marginales son $f_X(x) = \int_0^1 4xy \, dy = 2x$ y, por simetría, $f_Y(y) = 2y$. Como $f_X(x)f_Y(y) = (2x)(2y) = 4xy = f_{XY}(x, y)$, **sí son independientes**.

(b) En cambio, para el triángulo $f_{XY} = 2$ teníamos $f_X(x) = 2(1-x)$, $f_Y(y) = 2(1-y)$. El producto $f_X(x)f_Y(y) = 4(1-x)(1-y) \neq 2$, así que **no son independientes**. La pista esta en el soporte: como $x + y \leq 1$ liga las dos variables, saber X *sí* informa sobre Y .

Si la región de soporte R_{XY} no es un “rectángulo” (producto de un intervalo en x por uno en y), las variables **no pueden** ser independientes, por más bonita que sea la fórmula de la densidad.

9. Funciones de variables aleatorias

Muchas cantidades de interés son *funciones* de variables aleatorias: el ingreso total $Z = X + Y$ de una pareja, el logaritmo del ingreso $\log X$, el cuadrado de una desviación. Si X es aleatoria, también lo es $Z = g(X)$, y queremos su distribución.

Método de la cdf

Para hallar la distribución de $Z = g(X)$:

1. Escribe la cdf de Z como una probabilidad sobre X : $F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(g(X) \leq z)$.
2. Resuelve la desigualdad para expresarla como un evento sobre X y evalúalo con la cdf/pdf de X .
3. Deriva para obtener la pdf: $f_Z(z) = F'_Z(z)$.

Ejemplo 9.1 (Distribución de $Z = X^2$). Sea $X \sim \text{Unif}[0, 1]$, es decir $f_X(x) = 1$ en $[0, 1]$ y $F_X(x) = x$. Definimos $Z = X^2$ (que vive en $[0, 1]$). Para $0 \leq z \leq 1$:

$$F_Z(z) = \mathbb{P}(X^2 \leq z) = \mathbb{P}(X \leq \sqrt{z}) = F_X(\sqrt{z}) = \sqrt{z}.$$

Derivando, $f_Z(z) = \frac{d}{dz} \sqrt{z} = \frac{1}{2\sqrt{z}}$ para $0 < z \leq 1$. Verificación: $\int_0^1 \frac{1}{2\sqrt{z}} \, dz = [\sqrt{z}]_0^1 = 1$ (✓).

10. En R: muestrear y visualizar una densidad

Densidad exponencial en R

```
set.seed(1)
x <- rexp(1e5, rate = 1/5)           # X ~ Exponencial (media 5)
1 - pexp(2, rate = 1/5)              # P(X >= 2) = e^{-2/5} ~ 0.670
hist(x, breaks = 60, freq = FALSE)  # histograma de la muestra
curve(dexp(x, 1/5), add = TRUE, lwd = 2) # pdf teorica encima
```

El histograma de la muestra aproxima la densidad $f(x) = \frac{1}{5}e^{-x/5}$.

Resumen de la clase

Lo esencial de la Clase 3

- **Variable aleatoria:** función $X : S \rightarrow \mathbb{R}$. *Discreta* (valores contables) o *continua* (intervalo).
- **pmf** (discreta): $f_X(x) = \mathbb{P}(X = x)$, con $\sum f_X = 1$. **pdf** (continua): $f_X \geq 0$, $\int f_X = 1$, y $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X$.
- **Constante de normalización:** impón que la (densidad) integre/sume a 1 y despeja c .
- **cdf:** $F_X(x) = \mathbb{P}(X \leq x)$; misma definición para ambos tipos. Continua: $F'_X = f_X$.
- **Conjunta** f_{XY} sobre la región de soporte R_{XY} (¡dibújala!). Integra a 1 en el plano.
- **Marginal:** integra/suma la otra variable. **Condiciona:** $f_{Y|X} = f_{XY}/f_X$. **Independencia:** $f_{XY} = f_X f_Y$.
- **Función de una v.a.:** $Z = g(X)$ por el *método de la cdf*: $F_Z(z) = \mathbb{P}(g(X) \leq z)$, luego derivar.

Fuentes y atribución

Material de repaso **basado en MIT 14.310x – Data Analysis for Social Scientists** (Esther Duflo & Sara Fisher Ellison), MIT OpenCourseWare, licencia **CC BY-NC-SA 4.0** (uso no comercial, con atribución, compartir bajo la misma licencia), y en las fuentes de la bibliografía. Los enunciados se basan en el material del curso y en diversas fuentes abiertas; cuando un problema se adapta de una fuente específica, se cita en el enunciado.

Bibliografía.

- Duflo, E. & Ellison, S. F. *14.310x: Data Analysis for Social Scientists*. MIT OpenCourseWare (slides, lecture notes y problem sets). ocw.mit.edu.
- Larsen, R. J. & Marx, M. L. *An Introduction to Mathematical Statistics and Its Applications*, 6.^a ed., Pearson, 2018.
- DeGroot, M. H. & Schervish, M. J. *Probability and Statistics*, 4.^a ed., Pearson, 2012.
- Blitzstein, J. & Hwang, J. *Introduction to Probability* (Harvard Stat 110); W. Chen, *Probability Cheatsheet*.
- Diez, D., Çetinkaya-Rundel, M. & Barr, C. *OpenIntro Statistics*. openintro.org.
- Materiales abiertos de la comunidad del curso en GitHub: [gevorg-hunanyan/probability](https://github.com/gevorg-hunanyan/probability); [hanmochen/Probability-Theory-2020](https://github.com/hanmochen/Probability-Theory-2020); [mynameisjanus/14310xDataAnalysis](https://github.com/mynameisjanus/14310xDataAnalysis).

creativecommons.org/licenses/by-nc-sa/4.0