

Capítulo 2

Probabilidad y simulación

2.1. Modelos: variables y vectores aleatorios	1
2.1.1. Variables aleatorias	1
2.1.2. Vectores aleatorios	6
2.2. Transformaciones de variables y vectores aleatorios . . .	12
2.2.1. Transformaciones de variables aleatorias	12
2.2.2. Transformaciones de vectores aleatorios	15
2.3. Algunos modelos aleatorios básicos	17
2.3.1. Variables uniformes	18
2.3.2. Variables exponenciales	19
2.3.3. La normal (unidimensional)	20
2.3.4. Variables Gamma	24
2.4. Simulación de variables aleatorias	28
2.4.1. Simulación de variables discretas	29
2.4.2. Simulación de variables continuas	30

Este capítulo contiene material básico de probabilidad.

Nuestro propósito es que sirva de recordatorio de los principales resultados de la teoría de la probabilidad, sobre variables y vectores aleatorios, que se usarán en capítulos subsiguientes. Al tiempo, nos va a permitir fijar notaciones y definiciones varias de uso sistemático.

2.1. Modelos: variables y vectores aleatorios

2.1.1. Variables aleatorias

Una variable aleatoria X es un modelo matemático (abstracto) que pretende recoger los posibles resultados de un experimento (los valores), junto con la “frecuencia relativa potencial” (las probabilidades) con que ocurre cada uno.

A. Variables aleatorias discretas

Una variable aleatoria discreta X viene dada¹ por una lista de valores² (números reales cualesquiera)

$$x_1, \dots, x_n$$

con sus respectivas probabilidades:

$$p_1, \dots, p_n, \quad \text{donde } p_j > 0 \quad \text{y} \quad \sum_{j=1}^n p_j = 1,$$

de manera que, para cada $i = 1, \dots, n$, $\mathbf{P}(X = x_i) = p_i$.

A la lista de valores (x_1, \dots, x_n) junto con la lista de probabilidades (p_1, \dots, p_n) nos referimos como la **función de masa** de X .

Llamaremos **soprote** de X al conjunto $\{x_1, \dots, x_n\}$ de posibles valores que toma X ; la función de masa de X es, pues, la función

$$x \in \text{sop}(X) \mapsto \mathbf{P}(X = x).$$

Las variables discretas son, conceptualmente, bien sencillas, aunque en los cálculos con ellas sólo en contadas ocasiones se obtienen fórmulas (maneables y compactas).

En un modelo discreto, las probabilidades se calculan *sumando*: para cualquier conjunto $A \subset \mathbb{R}$,

$$\mathbf{P}(X \in A) = \sum_{i: x_i \in A} p_i.$$

Es decir, sumamos (probabilidades) en todos los índices i para los que el correspondiente valor x_i pertenezca a A .

También nos interesan variables aleatorias discretas que toman infinitos valores, es decir, en las que $n = \infty$. En este caso $\text{sop}(X)$ es una sucesión $(x_j)_{j=1}^{\infty}$ y las probabilidades respectivas $(p_j)_{j=1}^{\infty}$ conforman una sucesión de números no negativos, $p_j \geq 0$, para $j \geq 1$, tales que $\sum_{j=1}^{\infty} p_j = 1$ y de manera que

$$\mathbf{P}(X = x_j) = p_j, \quad \text{para cada } j \geq 1.$$

En este caso, $n = \infty$, habrá que manipular series, con las precauciones correspondientes, y no tan sólo sumas, como en el caso $n < +\infty$. A veces, para distinguir, nombramos como *variables finitas* a las variables aleatorias (discretas) que toman un número finito de valores.

Media. La **esperanza o media** $\mathbf{E}(X)$ de la variable aleatoria X (finita) se define como

$$\mathbf{E}(X) = \sum_{j=1}^n x_j p_j$$

(“valores multiplicados por probabilidades”, y sumados).

¹No hacemos mención al espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$ subyacente.

²Ojo con la notación!, ahora se trata de un modelo abstracto: x_1, \dots, x_n no significa muestra, sino que es la lista de los posibles valores de la variable.

En el caso en el que el soporte de X sea infinito, se define como

$$\mathbf{E}(X) = \sum_{j=1}^{\infty} x_j p_j,$$

aunque aquí, ¡atención!, se requiere, como requisito previo, que

$$\mathbf{E}(|X|) = \sum_{j=1}^{\infty} |x_j| p_j < +\infty,$$

para que la serie que define $\mathbf{E}(X)$ sea incondicionalmente convergente.

Medias y funciones de variables. Sea X una variable aleatoria discreta. Consideremos una función $h: \mathbb{R} \rightarrow \mathbb{R}$ y la variable “transformada” $Y = h(X)$. Esta variable toma los valores (transformados) $h(x_1), h(x_2), \dots$ con las probabilidades (originales, aquellas con las que X toma sus valores) p_1, p_2, \dots . Por lo tanto, su esperanza, $\mathbf{E}(Y) = \mathbf{E}(h(X))$ se calcula como sigue:

$$\mathbf{E}(h(X)) = \sum_{j=1}^n h(x_j) p_j.$$

En caso de tener una serie (caso $n = \infty$), se requiere de nuevo que $\mathbf{E}(|h(X)|) < +\infty$.

Nota 2.1.1. Por supuesto, en la lista $h(x_1), h(x_2), \dots$ podría haber valores repetidos, y establecer la función de masa de Y podría ser asunto complicado.

Hacemos notar aquí, lector, que si h es una transformación lineal, $h(x) = ax + b$, entonces

$$\mathbf{E}(h(X)) = \mathbf{E}(aX + b) = a \mathbf{E}(X) + b = h(\mathbf{E}(X)),$$

pero que, en general,

$$\mathbf{E}(h(X)) \neq h(\mathbf{E}(X))$$

(por si le tentaba la simple permutación de los símbolos \mathbf{E} y h).

Si la función h es convexa, las dos cantidades anteriores, la media de la variable transformada, y la transformación de la media, están ordenadas. Recordamos que una función $h(x)$ se dice *convexa* si, para todo $t \in [0, 1]$, y todo $x, y \in \mathbb{R}$,

$$(2.1) \quad h(tx + (1-t)y) \leq t h(x) + (1-t) h(y).$$

Geométricamente, una función es convexa si su gráfica queda siempre *por debajo* de la recta que une dos cualesquiera de sus puntos. La función se dice *cóncava* si la desigualdad (2.1) va en el otro sentido.

Teorema 2.1 (Desigualdad de Jensen) *Si X es una variable aleatoria, y $h(x)$ una función convexa, entonces*

$$(2.2) \quad h(\mathbf{E}(X)) \leq \mathbf{E}(h(X)).$$

Este resultado, la desigualdad de Jensen, es general y válido para cualesquiera variables aleatorias (tanto discretas como las continuas que vienen a continuación).

Se tiene igualdad en (2.2) cuando $h(x)$ es una función lineal, como se vio antes; y si $h(x)$ no es lineal, cuando X sea una constante.

Si la función $h(x)$ es cóncava, entonces el signo de la desigualdad (2.2) de Jensen se invierte. Como ejemplos relevantes, observamos que la función e^x es convexa, lo que nos da que, para toda variable aleatoria X ,

$$(2.3) \quad \mathbf{E}(e^X) \geq e^{\mathbf{E}(X)}$$

mientras que $\ln(x)$ es cóncava, por lo que, para toda variable aleatoria X *positiva*,

$$(2.4) \quad \mathbf{E}(\ln(X)) \leq \ln(\mathbf{E}(X)).$$

Varianzas y otros momentos. La elección de las potencias como funciones h en la discusión anterior da lugar a la siguiente nomenclatura: definimos los **momentos de orden** $k \geq 1$:

$$\begin{array}{c|c} \text{centrados} & \text{no centrados} \\ \hline \mathbf{E}((X - \mathbf{E}(X))^k) & \mathbf{E}(X^k) \end{array}$$

Estas definiciones requieren, en el caso $n = \infty$, que $\mathbf{E}(|X|^k) < +\infty$.

La media $\mathbf{E}(X)$ es el momento (no centrado) de orden 1, y la **varianza** $\mathbf{V}(X)$ es el momento centrado de orden 2:

$$(2.5) \quad \mathbf{V}(X) = \mathbf{E}[(X - \mathbf{E}(X))^2] = \mathbf{E}(X^2) - \mathbf{E}(X)^2.$$

La **desviación típica** se define como $\sigma(X) = \sqrt{\mathbf{V}(X)}$.



Nota 2.1.2. Como por definición $\mathbf{V}(X) \geq 0$, de (2.5) se deduce que $\mathbf{E}(X^2) \geq \mathbf{E}(X)^2$. Esto también se obtiene de la desigualdad de Jensen, teorema 2.1, tomando allí $h(x) = x^2$; o también de la desigualdad de Cauchy-Schwarz, teorema 2.3 más adelante, tomando allí $Y \equiv 1$.

Obsérvese que la varianza $\mathbf{V}(X)$ es cero si y sólo si X es una constante c , puesto que

$$0 = \sum_j (x_j - \mathbf{E}(X))^2 p_j$$

equivale a que $x_j = \mathbf{E}(X)$ para todo j , dado que todos los p_j son positivos.

Compruebe, lector, usando (2.5), que si $a, b \in \mathbb{R}$,

$$\mathbf{V}(aX + b) = a^2 \mathbf{V}(X).$$

Con los momentos centrados de orden 3 y 4 se definen los **coeficientes de asimetría** y la **curtosis** de X :

$$(2.6) \quad \text{asim}(X) = \frac{\mathbf{E}((X - \mu)^3)}{\sigma(X)^3}, \quad \text{curtosis}(X) = \frac{\mathbf{E}((X - \mu)^4)}{\sigma(X)^4}.$$



Nota 2.1.3. Hay quien define la curtosis de una variable aleatoria restando 3 en la definición anterior. La razón de esa definición alternativa es que así las variables normales tienen curtosis 0.

Recogemos a continuación un par de desigualdades, conocidas como de Chebyshev y de Markov, que son válidas tanto para variables discretas como para las variables continuas que definiremos más adelante. La primera acota, en términos de la varianza, la probabilidad de que una variable tome valores alejados de su media. La segunda es válida para variables no negativas (aquellas para las que $\mathbf{P}(X < 0) = 0$).

Teorema 2.2 a) (Desigualdad de Chebyshev). *Si X es una variable aleatoria con media y varianza finitas, entonces, para todo $\lambda > 0$,*

$$\mathbf{E}(|X - \mathbf{E}(X)| > \lambda) \leq \frac{\mathbf{V}(X)}{\lambda^2}.$$

b) (Desigualdad de Markov). *Si X es una variable aleatoria no negativa con media finita, entonces, para $a > 0$,*

$$\mathbf{E}(X > a) \leq \frac{\mathbf{E}(X)}{a}.$$

B. Variables aleatorias continuas

Una variable aleatoria continua X viene definida por una **función de densidad**, a la que nombraremos en general como $f_X(x)$, que es una función integrable en \mathbb{R} que cumple

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

(no negativa y con área total de 1 bajo su gráfica). El soporte de la variable X se define como $\text{sop}(X) = \{x \in \mathbb{R} : f_X(x) > 0\}$.

Las variables continuas son conceptualmente más complicadas que las discretas, pero en muchos casos resultan más fáciles de manipular (pues integrar es frecuentemente más sencillo que sumar).

En estos modelos continuos, el cálculo de las probabilidades de que X tome valores en ciertos conjuntos requiere *integrar* (calcular áreas encerradas bajo la función de densidad). Por ejemplo, la probabilidad de que X tome valores en un cierto intervalo $[a, b]$ es

$$\mathbf{P}(X \in [a, b]) = \int_a^b f_X(x) dx.$$

La **esperanza o media** $\mathbf{E}(X)$ de la variable aleatoria X se calcula como sigue:

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

(“valores multiplicados por probabilidades”, pero ahora integrados). Se requiere aquí que $\mathbf{E}(|X|) = \int_{-\infty}^{\infty} |x| f_X(x) dx < +\infty$.

Si $Y = h(X)$, donde $h(x)$ es una cierta función (medible Borel), entonces

$$\mathbf{E}(h(X)) = \int_{-\infty}^{\infty} h(x) f_X(x) dx.$$

Se requiere aquí de nuevo $\mathbf{E}(|h(X)|) < +\infty$.

Como en el caso discreto, los momentos de X son las cantidades $\mathbf{E}(X^k)$ para $k \geq 1$, o quizás $\mathbf{E}((X - \mathbf{E}(X))^k)$ para $k \geq 1$, cuyo cálculo requiere, claro, integrar.



Nota 2.1.4. Determinar la función de densidad de $Y = h(X)$ suele ser asunto más delicado. Véase la discusión del apartado 2.2.1.



Nota 2.1.5. Función de distribución. Recordamos seguidamente una descripción alternativa de variable aleatoria común a las variables discretas y a las continuas. Una variable aleatoria X viene dada por su función de distribución F_X , que se define como

$$F_X(x) = \mathbf{P}(X \leq x) \quad \text{para cada } x \in \mathbb{R}.$$

La función de distribución F_X está definida en todo \mathbb{R} y toma valores en $[0, 1]$.

- En el caso de una variable discreta X que toma los valores x_1, x_2, \dots con probabilidades respectivas p_1, p_2, \dots , la función de distribución es una función con saltos:

$$F_X(x) = \sum_{j: x_j \leq x} p_j.$$

- Para una variable continua X con función de densidad $f_X(x)$,

$$F_X(x) = \int_{-\infty}^x f_X(y) dy.$$

En este caso, $F'_X(x) = f_X(x)$.

2.1.2. Vectores aleatorios

La siguiente noción considera *conjuntamente* una colección (finita) de variables aleatorias; de ahí la expresión “vector aleatorio”.

A. Caso discreto

Definimos la **función de masa conjunta** del vector (X_1, \dots, X_n) como la colección de números (probabilidades conjuntas)

$$\mathbf{P}(X_1 = a_1, \dots, X_n = a_n) \geq 0,$$

donde cada $a_i \in \text{sop}(X_i)$, para $i = 1, \dots, n$, tales que

$$\sum_{a_1 \in \text{sop}(X_1)} \cdots \sum_{a_n \in \text{sop}(X_n)} \mathbf{P}(X_1 = a_1, \dots, X_n = a_n) = 1.$$

La probabilidad de que (X_1, \dots, X_n) tome valores en un cierto subconjunto $A \subset \mathbb{R}^n$ viene dada por

$$\sum_{(a_1, \dots, a_n) \in A} \mathbf{P}(X_1 = a_1, \dots, X_n = a_n).$$

La función de masa de una coordenada X_i se obtiene fijando el valor de X_i y sumando en los valores de todas las demás variables. Por ejemplo,

$$\mathbf{P}(X_1 = \alpha) = \sum_{a_2 \in \text{sop}(X_2)} \cdots \sum_{a_n \in \text{sop}(X_n)} \mathbf{P}(X_1 = \alpha, X_2 = a_2, \dots, X_n = a_n)$$

para cada $\alpha \in \text{sop}(X_1)$. A estas funciones de masa de las variables coordenadas X_i se les conoce como **funciones de masa marginales de \mathbb{X}** .

Dada una función $h : \mathbb{R}^n \rightarrow \mathbb{R}$ (medible Borel), consideramos la variable aleatoria $Z = h(X_1, \dots, X_n)$. Su media se calcula como

$$\mathbf{E}(Z) = \sum_{a_1 \in \text{sop}(X_1)} \cdots \sum_{a_n \in \text{sop}(X_n)} h(a_1, \dots, a_n) \mathbf{P}(X_1 = a_1, \dots, X_n = a_n);$$

en palabras, «suma de valores por probabilidades». (Aquí, como es habitual, se requiere que $\mathbf{E}(|Z|) < +\infty$).

B. Caso continuo

El vector aleatorio (X_1, \dots, X_n) se define a través de una función $f_{(X_1, \dots, X_n)}$ de densidad conjunta definida en \mathbb{R}^n tal que

$$\begin{aligned} & \text{(no negativa)} \quad f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \geq 0; \\ & \text{(integral 1)} \quad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) dx_1 \cdots dx_n = 1. \end{aligned}$$

Es decir, la función $f_{(X_1, \dots, X_n)}$ es no negativa y el volumen total bajo su gráfica es 1.

La probabilidad de que (X_1, \dots, X_n) tome valores en un determinado subconjunto $A \subset \mathbb{R}^n$ viene dada por

$$\int_A f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

Las funciones de densidad de cada X_i , es decir, las **funciones de densidad marginales** del vector (X_1, \dots, X_n) , se calculan

$$f_{X_i}(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{(X_1, \dots, X_n)}(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

Nótese que se integra en todas las variables excepto la i -ésima.

La media de una variable aleatoria $Z = h(X_1, \dots, X_n)$, donde h es una función $h : \mathbb{R}^n \rightarrow \mathbb{R}$, se calcula en este caso como

$$\mathbf{E}(Z) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \dots, x_n) f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) dx_1 \cdots dx_n$$

(con el requisito de que $\mathbf{E}(|Z|) < +\infty$).

C. Independencia

Las variables (X_1, \dots, X_n) se dicen **independientes** si

$$\mathbf{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbf{P}(X_1 \in A_1) \cdot \mathbf{P}(X_2 \in A_2) \cdots \mathbf{P}(X_n \in A_n),$$

para cualesquiera conjuntos (de Borel) $A_1, \dots, A_n \subset \mathbb{R}$.



Nota 2.1.6. Usaremos en ocasiones la siguiente observación: si las variables (X_1, \dots, X_n) son independientes, entonces las variables $(Y_1, \dots, Y_n) = (T_1(X_1), \dots, T_n(X_n))$, donde T_1, \dots, T_n son funciones medibles de \mathbb{R} en \mathbb{R} , también lo son.

La independencia en el caso discreto equivale a cerciorarse de que

$$\mathbf{P}(X_1 = a_1, \dots, X_n = a_n) = \mathbf{P}(X_1 = a_1) \cdot \mathbf{P}(X_2 = a_2) \cdots \mathbf{P}(X_n = a_n),$$

para cada $(a_1, \dots, a_n) \in \mathbb{R}^n$, con $a_j \in \text{sop}(X_j)$ para $1 \leq j \leq n$.

En el caso continuo, las variables coordenadas X_j son independientes si y sólo si la función de densidad conjunta $f_{(X_1, \dots, X_n)}$ se factoriza como producto de las funciones de densidad de las coordenadas X_j en la forma siguiente:

$$(2.7) \quad f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n), \quad \text{para } (x_1, \dots, x_n) \in \mathbb{R}^n.$$



Nota 2.1.7. Usaremos en ocasiones la observación siguiente: si $f_{(X_1, \dots, X_n)}(x_1, \dots, x_n)$ se factoriza como producto de funciones $g_j(x_j)$ no negativas e integrables (cada g_j es función de una sola variable, x_j), es decir, si

$$(b) \quad f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n g_i(x_i), \quad \text{para } (x_1, \dots, x_n) \in \mathbb{R}^n,$$

entonces las X_j son independientes. Veamos. Denotemos las integrales $\int_{\mathbb{R}} g_j(x_j) dx_j = \lambda_j$. De (b) se deduce que $\prod_{j=1}^n \lambda_j = 1$. Asimismo de (b) se deduce que

$$f_{X_j}(x_j) = g_j(x_j) \prod_{i \neq j} \lambda_i = \frac{1}{\lambda_j} g_j(x_j), \quad \text{para cada } x_j \in \mathbb{R}.$$

Así que, para $x_1, \dots, x_n \in \mathbb{R}$,

$$\prod_{j=1}^n f_{X_j}(x_j) = \prod_{j=1}^n \frac{1}{\lambda_j} g_j(x_j) = \prod_{j=1}^n \frac{1}{\lambda_j} \cdot \prod_{j=1}^n g_j(x_j) = \prod_{j=1}^n g_j(x_j) = f_{(X_1, \dots, X_n)}(x_1, \dots, x_n).$$

En ese caso, $f_{X_j} \equiv g_j$, para $1 \leq j \leq n$.

La frase, que nos encontraremos a menudo, que dice que X e Y son variables aleatorias independientes con funciones de densidad respectivas f_X y f_Y especifica de manera completa la distribución de probabilidad del vector aleatorio (X, Y) , pues ha de ser

$$f_{(X, Y)}(x, y) = f_X(x) f_Y(y), \quad \text{para } (x, y) \in \mathbb{R}^2.$$

D. Covarianzas y coeficientes de correlación

Sea (X_1, \dots, X_n) un vector aleatorio. Para cada $i \neq j$, definimos la **covarianza** entre las variables X_i y X_j como sigue:

$$(2.8) \quad \text{cov}(X_i, X_j) = \mathbf{E}[(X_i - \mathbf{E}(X_i)) \cdot (X_j - \mathbf{E}(X_j))] = \mathbf{E}(X_i X_j) - \mathbf{E}(X_i) \mathbf{E}(X_j).$$

Para $1 \leq i, j \leq n$ se tiene que

$$\mathbf{V}(X_i + X_j) = \mathbf{V}(X_i) + \mathbf{V}(X_j) + 2 \text{cov}(X_i, X_j).$$

En general, si $a_1, \dots, a_n \in \mathbb{R}$, entonces la combinación lineal $\sum_{j=1}^n a_j X_j$ tiene varianza

$$(2.9) \quad \begin{aligned} \mathbf{V}\left(\sum_{j=1}^n a_j X_j\right) &= \sum_{1 \leq i, j \leq n} a_i a_j \text{cov}(X_i, X_j) \\ &= \sum_{j=1}^n a_j^2 \mathbf{V}(X_j) + \sum_{1 \leq i \neq j \leq n} a_i a_j \text{cov}(X_i, X_j). \end{aligned}$$

Por su parte, el **coeficiente de correlación** entre X_i y X_j es

$$(2.10) \quad \rho(X_i, X_j) = \frac{\text{cov}(X_i, X_j)}{\sqrt{\mathbf{V}(X_i)} \sqrt{\mathbf{V}(X_j)}}.$$

El coeficiente de correlación $\rho(X_i, X_j)$ está definido sólo si X_i y X_j son variables no constantes, es decir, si $\mathbf{V}(X_i) \neq 0$ y $\mathbf{V}(X_j) \neq 0$.

Como vemos a continuación, el coeficiente de correlación entre dos variables cualesquiera es siempre un número entre -1 y 1 .

Partimos de la desigualdad general:

Teorema 2.3 (Desigualdad de Cauchy–Schwarz) *Si (X, Y) es un vector aleatorio, entonces*

$$|\mathbf{E}(X \cdot Y)|^2 \leq \mathbf{E}(X^2) \cdot \mathbf{E}(Y^2).$$

Si X e Y son variables aleatorias no nulas, entonces se tiene la igualdad si y sólo si $X = \alpha Y$ con probabilidad 1.

Que nos dice, tomando $\tilde{X} = X - \mathbf{E}(X)$ e $\tilde{Y} = Y - \mathbf{E}(Y)$, lo siguiente:

Corolario 2.4 *Si (X, Y) es un vector aleatorio, entonces*

$$|\text{cov}(X, Y)| \leq \sqrt{\mathbf{V}(X)} \sqrt{\mathbf{V}(Y)}.$$

Es decir, $|\rho(X, Y)| \leq 1$.

Si X e Y son variables aleatorias no constantes, entonces se tiene la igualdad si y sólo si, para ciertos $\alpha, \beta \in \mathbb{R}$, $X = \alpha Y + \beta$ con probabilidad 1.

Si $\rho(X, Y) = 0$ (o equivalentemente, si $\text{cov}(X, Y) = 0$), se dice que las variables X e Y son **incorreladas**. Si las variables X e Y son independientes, entonces son incorreladas (basta observar que la independencia nos da que $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$); el recíproco es falso en general.

E. Notación matricial

En lo que sigue usaremos a menudo notaciones matriciales en los cálculos. Denotaremos por \mathbb{X} , \mathbf{x} y $\mathbf{E}(\mathbb{X})$ a vectores (columna) cuyas n componentes son, respectivamente, variables aleatorias, números reales, y medias de las variables aleatorias anteriores:

$$\underbrace{\mathbb{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}}_{\text{vector aleatorio}}, \quad \underbrace{\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}_{\text{vector de } \mathbb{R}^n}, \quad \underbrace{\mathbf{E}(\mathbb{X}) = \begin{pmatrix} \mathbf{E}(X_1) \\ \vdots \\ \mathbf{E}(X_n) \end{pmatrix}}_{\text{vector de medias}}.$$

Si por razones tipográficas interesara escribirlos como filas, lo indicaríamos siempre usando traspuestas. Como, por ejemplo, al escribir $\mathbb{X} = (X_1, \dots, X_n)^\top$.

Con más generalidad, si $\mathbb{M} = (X_{i,j})_{i,j}$ es una *matriz* de dimensiones $n \times m$ cuyas componentes son variables aleatorias, escribiremos $\mathbf{E}(\mathbb{M})$ para referirnos la matriz $(\mathbf{E}(X_{i,j}))_{i,j}$ de medias de esas variables.

Obsérvese que si \mathbb{X} es un vector aleatorio de dimensión n , si A es una matriz $n \times n$ y si \mathbf{b} es un vector de \mathbb{R}^n , entonces

$$\mathbf{E}(A\mathbb{X} + \mathbf{b}) = A \mathbf{E}(\mathbb{X}) + \mathbf{b}.$$

Usamos aquí, simplemente, la linealidad de la media.

Dado el vector aleatorio $\mathbb{X} = (X_1, \dots, X_n)^\top$, definimos su **matriz de covarianzas** (o de varianzas/covarianzas) como la matriz $n \times n$ que registra las covarianzas entre cada pareja de las variables X_j . La denotamos con el símbolo $\mathbf{Cov}(\mathbb{X})$:

$$\mathbf{Cov}(\mathbb{X}) = (\text{cov}(X_i, X_j))_{1 \leq i, j \leq n}.$$

Usamos aquí la notación unificadora $\mathbf{V}(X_i) = \text{cov}(X_i, X_i)$ para las varianzas.

Obsérvese que $(\mathbb{X} - \mathbf{E}(\mathbb{X})) \cdot (\mathbb{X} - \mathbf{E}(\mathbb{X}))^\top$ es una matriz (aleatoria) $n \times n$, y que

$$(2.11) \quad \mathbf{E}((\mathbb{X} - \mathbf{E}(\mathbb{X})) \cdot (\mathbb{X} - \mathbf{E}(\mathbb{X}))^\top) = \mathbf{Cov}(\mathbb{X}).$$



Nota 2.1.8. En contraste, $(\mathbb{X} - \mathbf{E}(\mathbb{X}))^\top \cdot (\mathbb{X} - \mathbf{E}(\mathbb{X}))$ es una *variable aleatoria* con media

$$\mathbf{E}((\mathbb{X} - \mathbf{E}(\mathbb{X}))^\top \cdot (\mathbb{X} - \mathbf{E}(\mathbb{X}))) = \mathbf{E}(\|\mathbb{X} - \mathbf{E}(\mathbb{X})\|^2) = \sum_{i=1}^n \mathbf{V}(X_i).$$

Si \mathbb{X} es un vector aleatorio de dimensión n , si A es una matriz $n \times n$ y si \mathbf{b} es un vector de \mathbb{R}^n , entonces

$$\mathbf{Cov}(A\mathbb{X} + \mathbf{b}) = A \cdot \mathbf{Cov}(\mathbb{X}) \cdot A^\top.$$

La **matriz de correlaciones** $\Sigma(\mathbb{X})$ del vector $\mathbb{X} = (X_1, \dots, X_n)^\top$ es la matriz cuadrada $n \times n$ que registra los coeficientes de correlación entre cada pareja de las variables X_j :

$$\Sigma(\mathbb{X}) = (\rho(X_i, X_j))_{1 \leq i, j \leq n},$$

Obsérvese que la matriz de correlaciones tiene unos en la diagonal. Y que, claro, $\Sigma(\mathbb{X})$ está definida sólo cuando $V(X_j) \neq 0$, para $1 \leq j \leq n$.



Nota 2.1.9. La matriz $\Sigma(\mathbb{X})$ de correlaciones de \mathbb{X} es la matriz de covarianzas del vector $\hat{\mathbb{X}}$, cuyas componentes son

$$\hat{X}_j = \frac{X_j}{\sqrt{V(X_j)}}.$$

Si las variables del vector $\mathbb{X} = (X_1, \dots, X_n)^\top$ son independientes (dos a dos), entonces son incorreladas, y por consiguiente tanto $\mathbf{Cov}(\mathbb{X})$ como $\Sigma(\mathbb{X})$ son matrices diagonales.

Lema 2.5 Sea \mathbb{X} un vector aleatorio.

La matriz de covarianzas $\mathbf{Cov}(\mathbb{X})$ y la matriz de correlaciones $\Sigma(\mathbb{X})$ de \mathbb{X} son matrices simétricas y semidefinidas positivas.

Véase el apartado 3.1 (o el capítulo 9) para estas nociones de álgebra lineal.

DEMOSTRACIÓN. Hay que comprobar que, para todo vector $\mathbf{a}^\top = (a_1, \dots, a_n) \in \mathbb{R}^n$, se tiene que

$$\mathbf{a}^\top \mathbf{Cov}(\mathbb{X}) \mathbf{a} \geq 0.$$

Pero por la identidad (2.9),

$$(\star) \quad \mathbf{a}^\top \mathbf{Cov}(\mathbb{X}) \mathbf{a} = \sum_{j=1}^n a_j^2 V(X_j) + \sum_{i \neq j} \text{cov}(X_i, X_j) a_i a_j = V(a_1 X_1 + \dots + a_n X_n),$$

lo que concluye la comprobación, pues $V(Z) \geq 0$ para cualquier variable aleatoria Z .

Si $\mathbf{Cov}(\mathbb{X})$ no es definida positiva, entonces es que existe un $\mathbf{a} = (a_1, \dots, a_n) \neq \mathbf{0}$ tal que $\mathbf{a}^\top \mathbf{Cov}(\mathbb{X}) \mathbf{a} = 0$. Pero entonces, por (\star) , se tendría que $V(a_1 X_1 + \dots + a_n X_n) = 0$, es decir, que la variable aleatoria $a_1 X_1 + \dots + a_n X_n$ sería constante.

Veamos ahora el caso de la matriz de correlaciones $\Sigma(\mathbb{X})$. Para que $\Sigma(\mathbb{X})$ esté definida, suponemos de partida que ninguna X_j es constante, es decir, que $V(X_j) > 0$. Sea $\mathbf{a}^\top = (a_1, \dots, a_n)$ un vector de \mathbb{R}^n y consideremos el vector $\mathbf{b}^\top = (b_1, \dots, b_n)$, auxiliar, dado por

$$b_j = \frac{a_j}{\sqrt{V(X_j)}}, \quad \text{para } 1 \leq j \leq n.$$

Entonces,

$$\mathbf{a}^\top \Sigma(\mathbb{X}) \mathbf{a} = \mathbf{b}^\top \mathbf{Cov}(\mathbb{X}) \mathbf{b} \geq 0,$$

pues $\mathbf{Cov}(\mathbb{X})$, como se ha probado antes, es semidefinida positiva. ■

2.2. Transformaciones de variables y vectores aleatorios

Analizamos a continuación cómo determinar las funciones de densidad de variables (y vectores) aleatorios que son función de otras variables aleatorias.

2.2.1. Transformaciones de variables aleatorias

Supongamos que X es una variable aleatoria *continua* cuyo soporte $\text{sop}(X)$ está incluido en un intervalo (a, b) . Este intervalo (a, b) puede ser finito, o quizás $(-\infty, \infty)$, o $(0, +\infty)$ o...

Sea H un difeomorfismo³ del intervalo (a, b) sobre un intervalo (c, d) y consideremos la variable aleatoria $Y = H \circ X = H(X)$, que toma valores en (c, d) .

La relación entre las funciones de densidad, f_X y f_Y , de las variables X e Y es la siguiente:

Lema 2.6 *Sea X una variable aleatoria continua con $\text{sop}(X) = (a, b)$ y sea Y la variable aleatoria dada por $Y = H(X)$, donde H es un difeomorfismo $H: (a, b) \rightarrow (c, d)$. Entonces,*

$$(2.12) \quad f_Y(H(t)) |H'(t)| = f_X(t), \quad \text{para todo } t \in (a, b).$$

La fórmula (2.12) se puede reescribir como

$$(2.13) \quad f_Y(s) = f_X(H^{-1}(s)) |(H^{-1})'(s)|, \quad \text{para todo } s \in (c, d),$$

expresión que suele resultar más práctica, pues da directamente la densidad de la variable transformada (aunque para ello haya que calcular la inversa de H).



Nota 2.2.1. Para demostrar el lema 2.6, analizamos por separado las dos posibilidades que caben: que H sea monótona creciente, o que sea monótona decreciente.

Caso 1. Pongamos, primero, que H es creciente ($H' > 0$). Para $t \in (a, b)$, tenemos que

$$\mathbf{P}(Y \leq H(t)) = \mathbf{P}(H(X) \leq H(t)) = \mathbf{P}(X \leq t)$$

(en la última igualdad hemos usado que H es creciente). Es decir,

$$F_Y(H(t)) = F_X(t), \quad \text{para todo } t \in (a, b).$$

Derivando respecto de t obtenemos que

$$f_Y(H(t))H'(t) = f_X(t), \quad \text{para todo } t \in (a, b).$$

Caso 2. Si H fuera decreciente ($H' < 0$), entonces para $t \in (a, b)$ tendríamos que

$$\mathbf{P}(Y \leq H(t)) = \mathbf{P}(H(X) \leq H(t)) = \mathbf{P}(X \geq t);$$

nótese cómo, al ser H decreciente, se invierte el signo de la desigualdad. Como X es una variable continua, y por tanto $\mathbf{P}(X = t) = 0$, deducimos que

$$F_Y(H(t)) = 1 - F_X(t), \quad \text{para todo } t \in (a, b).$$

Derivando respecto de t obtenemos que

$$f_Y(H(t))H'(t) = -f_X(t), \quad \text{para todo } t \in (a, b).$$

³Biyección f tal que tanto f como su inversa f^{-1} son diferenciables.

EJEMPLO 2.2.1. *Cambios de escala: de X a λX , con $\lambda > 0$.*

Sea $\lambda > 0$ y definamos $H(x) = \lambda x$, para $x \in \mathbb{R}$. Esta función H es un difeomorfismo de \mathbb{R} sobre \mathbb{R} y $H'(x) = \lambda$.

Sea $Y = \lambda X$. Entonces, usando (2.12) tenemos que

$$f_Y(\lambda t) \cdot \lambda = f_X(t),$$

o mejor, llamando $s = \lambda t$,

$$f_Y(s) = \frac{1}{\lambda} f_X\left(\frac{s}{\lambda}\right), \quad \text{para todo } s \in \mathbb{R}$$

(esto es lo que se habría obtenido usando directamente (2.13)).

Al pasar de X a λX los valores de X quedan multiplicados por λ . Esto es un cambio de escala, o si se prefiere, un cambio en las unidades con las que se miden los valores de X .

Como ilustración, para una variable $X \sim \text{EXP}(\lambda)$ con $\lambda > 0$, tenemos que $f_X(t) = \lambda e^{-\lambda t}$, para $t > 0$. Para la variable $Y = \lambda X$, se tiene que

$$f_Y(s) = \frac{1}{\lambda} f_X\left(\frac{s}{\lambda}\right) = e^{-s}, \quad \text{para todo } t \in (0, +\infty).$$

Es decir, $Y \sim \text{EXP}(1)$; una exponencial de parámetro 1. ♣

EJEMPLO 2.2.2. *Consideremos la función $H(t) = 1/t$.*

En este caso, si X toma valores en $(0, +\infty)$, entonces para $Y = 1/X$ se tiene que

$$f_Y\left(\frac{1}{t}\right) \frac{1}{t^2} = f_X(t), \quad \text{para todo } t \in (0, +\infty),$$

o alternativamente,

$$f_Y(s) = \frac{1}{s^2} f_X\left(\frac{1}{s}\right), \quad \text{para todo } s \in (0, +\infty).$$

Como ilustración, para una variable $X \sim \text{EXP}(\lambda)$ con $\lambda > 0$, tenemos que $f_X(t) = \lambda e^{-\lambda t}$, para $t > 0$. Para la variable $Y = 1/X$, se tiene que

$$f_Y(s) = \frac{\lambda}{s^2} e^{-\lambda/s}, \quad \text{para todo } s \in (0, +\infty). \quad \clubsuit$$

EJEMPLO 2.2.3. *Sea X una variable aleatoria continua con función de distribución F_X estrictamente creciente. Comprobemos que la variable $Y = F_X(X)$ es uniforme en el intervalo $[0, 1]$.*

En este caso, tenemos que $(a, b) = (-\infty, \infty)$ (puesto que $\lim_{t \rightarrow +\infty} F_X(t) = 1$ y $\lim_{t \rightarrow -\infty} F_X(t) = 0$), y $(c, d) = (0, 1)$. La transformación es $H(t) = F_X(t)$.

Si $Y = F_X(X)$, entonces la función de densidad de Y es 0 fuera de $(0, 1)$ y para $t \in \mathbb{R}$ se tiene $F_X(t) \in (0, 1)$ y se cumple que

$$f_Y(F_X(t)) F_X'(t) = f_X(t) \quad \implies \quad f_Y(F_X(t)) f_X(t) = f_X(t).$$

Usando que $f_X(t) > 0$ para todo $t \in \mathbb{R}$, podemos cancelar factores para obtener que $f_Y(F_X(t)) = 1$. Como esto sucede para todo $t \in \mathbb{R}$, se deduce que

$$f_Y(s) = \begin{cases} 1, & \text{si } s \in (0, 1), \\ 0, & \text{en otro caso.} \end{cases}$$

En otros términos,

$$Y = F_X(X)$$

es una variable $\sim \text{UNIF}[0, 1]$. ♣



Nota 2.2.2. En particular, si F_X es un difeomorfismo (entre el soporte de X y el intervalo $(0, 1)$), entonces X y $F_X^{-1}(U)$, donde U es una variable uniforme en $[0, 1]$, tienen la misma distribución. Es decir, todas las variables continuas son *transformaciones* de la uniforme. (En realidad esto también es válido para variables discretas, sin más que interpretar adecuadamente el significado de F_X^{-1}). Usaremos esta observación para la simulación de variables aleatorias cualesquiera, apartado 2.4.

Si la función H que lleva (a, b) en (c, d) no fuera monótona, como en el enunciado del lema 2.6, entonces la relación hay entre f_Y y f_X es más delicada, y requiere un análisis más detallado, como se ilustra en el siguiente ejemplo 2.2.4.

EJEMPLO 2.2.4. Sea X una variable continua con $\text{sop}(X) = (-\infty, \infty)$, y sea Y la variable $Y = X^2$, que toma valores en $(0, +\infty)$.

En este caso, para $t > 0$ se tiene que

$$F_Y(t) = F_X(\sqrt{t}) - F_X(-\sqrt{t})$$

de donde

$$f_Y(t) = \frac{1}{2\sqrt{t}}(f_X(\sqrt{t}) + f_X(-\sqrt{t}))$$

Como caso particular, si X es una variable normal estándar, entonces de $Y = X^2$ se dice que es una $\chi^2(1)$ (que se lee: una chi-cuadrado con 1 grado de libertad). Véase la sección 3.3.1. La función de densidad de $Y = X^2$ es, en este caso,

$$f_Y(t) = \begin{cases} \frac{1}{\sqrt{2\pi}} t^{-1/2} e^{-t/2}, & \text{si } t > 0; \\ 0, & \text{si } t \leq 0. \end{cases}$$



2.2.2. Transformaciones de vectores aleatorios

Consideremos una transformación $H : \mathbb{R}^n \rightarrow \mathbb{R}^n$ dada por

$$H(x_1, \dots, x_n) = (H_1(x_1, \dots, x_n), \dots, H_n(x_1, \dots, x_n))$$

que sea un difeomorfismo (de \mathbb{R}^n en sí mismo).

Denotamos por $J_H(\mathbf{x})$ al jacobiano de H en \mathbf{x} , es decir, al determinante de la matriz jacobiana de H en \mathbf{x} :

$$J_H(\mathbf{x}) = \det \begin{pmatrix} \frac{\partial H_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial H_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial H_n}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial H_n}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

Consideramos dos vectores aleatorios \mathbb{X} e \mathbb{Y} *continuos* de dimensión n ligados mediante H :

$$\mathbb{Y} \equiv H(\mathbb{X}).$$

La relación entre $f_{\mathbb{Y}}$ y $f_{\mathbb{X}}$ se recoge en el siguiente resultado.

Lema 2.7 *Sea un difeomorfismo $H : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Si $\mathbb{Y} = H(\mathbb{X})$, entonces*

$$(2.14) \quad f_{\mathbb{X}}(\mathbf{x}) = f_{\mathbb{Y}}(H(\mathbf{x})) |J_H(\mathbf{x})|, \quad \text{para todo } \mathbf{x} \in \mathbb{R}^n.$$

Nota 2.2.3. Recordamos al lector la idea de la demostración del lema 2.7. Sea A un conjunto (de Borel) cualquiera en \mathbb{R}^n y sea $B = H(A)$. Entonces,

$$\mathbf{P}(\mathbb{Y} \in B) = \mathbf{P}(H(\mathbb{X}) \in H(A)) = \mathbf{P}(\mathbb{X} \in A) = \int_A f_{\mathbb{X}}(\mathbf{x}) d\mathbf{x}.$$

Además, con el cambio de variables $\mathbf{y} = H(\mathbf{x})$, se tiene

$$\mathbf{P}(\mathbb{Y} \in B) = P(\mathbb{Y} \in H(A)) = \int_{H(A)} f_{\mathbb{Y}}(\mathbf{y}) d\mathbf{y} = \int_A f_{\mathbb{Y}}(H(\mathbf{x})) |J_H(\mathbf{x})| d\mathbf{x}.$$

Por tanto, *para todo* conjunto de Borel $A \subset \mathbb{R}^n$ se tiene que

$$\int_A f_{\mathbb{X}}(\mathbf{x}) d\mathbf{x} = \int_A f_{\mathbb{Y}}(H(\mathbf{x})) |J_H(\mathbf{x})| d\mathbf{x},$$

de donde se concluye el enunciado del lema.

EJEMPLO 2.2.5. *El caso de las transformaciones lineales.*

Si H es una transformación lineal, es decir, si $\mathbb{Y} = \mathbf{h} + M\mathbb{X}$, para una cierta matriz M invertible y un cierto vector \mathbf{h} , entonces

$$(2.15) \quad f_{\mathbb{X}}(\mathbf{x}) = f_{\mathbb{Y}}(\mathbf{h} + M\mathbf{x}) |\det(M)|, \quad \text{para todo } \mathbf{x} \in \mathbb{R}^n,$$

y también

$$(2.16) \quad f_{\mathbb{Y}}(\mathbf{y}) = \frac{1}{|\det(M)|} f_{\mathbb{X}}(M^{-1}(\mathbf{y} - \mathbf{h})), \quad \text{para todo } \mathbf{y} \in \mathbb{R}^n.$$



Como aplicación de esta fórmula del cambio de variables, se obtiene una expresión para la función de densidad de la *suma de variables aleatorias*.

Supongamos que $\mathbb{X} = (X, Y)^\top$ es un vector aleatorio con valores en \mathbb{R}^2 y con función de densidad conjunta $f_{(X,Y)}(x, y)$. Interesa hallar la función de densidad de la variable suma, $V = X + Y$.

Para ello consideramos el vector $\mathbb{U} = (U, V)^\top$ dado por

$$U = X \quad \text{y} \quad V = X + Y.$$

Obsérvese que $\mathbb{U} = M \cdot \mathbb{X}$, donde M es la matriz

$$M = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

que tiene $\det(M) = 1$. Así que, siguiendo (2.15),

$$f_{(X,Y)}(x, y) = f_{(U,V)}(x, x + y),$$

de manera que

$$(2.17) \quad f_{(U,V)}(u, v) = f_{(X,Y)}(u, v - u), \quad \text{para todo } (u, v) \in \mathbb{R}^2.$$

La densidad marginal $f_V(v)$ se obtiene integrando $f_{(U,V)}(u, v)$ en la variable u ,

$$f_V(v) = \int_{-\infty}^{\infty} f_{U,V}(u, v) du,$$

y como $f_V \equiv f_{X+Y}$, de (2.17) se deduce que, para todo $v \in \mathbb{R}$,

$$(2.18) \quad f_{X+Y}(v) = \int_{-\infty}^{\infty} f_{(X,Y)}(u, v - u) du.$$

En el caso particular en el que las coordenadas X e Y son variables *independientes*, de la fórmula (2.18) se deduce el lema siguiente, que será de uso harto frecuente en lo que sigue.

Lema 2.8 *Si X e Y son variables aleatorias continuas independientes, entonces la variable aleatoria $X + Y$ tiene función de densidad*

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx.$$

Las transformaciones H consideradas en el lema 2.7 son difeomorfismos de \mathbb{R}^n sobre \mathbb{R}^n y, por tanto, permiten transformar cualquier vector aleatorio \mathbb{X} . Si consideramos difeomorfismos de una parte de $A \subset \mathbb{R}^n$ sobre otra, entonces estos difeomorfismos sólo permiten transformar vectores aleatorios que toman valores en A .

En estos casos, aunque la relación entre funciones de densidad es la misma que en el lema 2.7, cierta precaución sobre los rangos de aplicación es necesaria.

EJEMPLO 2.2.6. A partir de un vector aleatorio (X, Y) con valores en $(0, \infty) \times (0, \infty)$, construimos el vector aleatorio (U, V) dado por $U = X + Y$ y $V = X/Y$.

La transformación $H(x, y)$ para este caso viene dada por $H(x, y) = (x + y, x/y)$, de manera que si $(u, v) = H(x, y)$ entonces

$$\begin{cases} u = x + y, \\ v = \frac{x}{y}, \end{cases} \quad y \quad \begin{cases} x = \frac{uv}{1+v}, \\ y = \frac{u}{1+v}. \end{cases}$$

La transformación H es un difeomorfismo de $(0, \infty) \times (0, \infty)$ sobre sí mismo.

El jacobiano de H viene dado por

$$|J_H(x, y)| = \frac{x + y}{y^2}.$$

La ecuación $(\star\star)$ dice entonces que

$$f_{(U,V)}\left(x + y, \frac{x}{y}\right) \cdot \frac{x + y}{y^2} = f_{(X,Y)}(x, y), \quad \text{para } x, y > 0.$$

Traduciendo en variables u, v se obtiene finalmente que

$$(\#) \quad f_{(U,V)}(u, v) = \frac{u}{(1+v)^2} \cdot f_{(X,Y)}\left(\frac{uv}{1+v}, \frac{u}{1+v}\right), \quad \text{para } x, y > 0.$$

Como ilustración consideremos el caso en que (X, Y) es un par de variables exponenciales independientes con el mismo parámetro λ , es decir, que la distribución conjunta es

$$f_{(X,Y)}(x, y) = \lambda^2 e^{-\lambda(x+y)}.$$

Para el par (U, V) con $U = X + Y$ y $V = X/Y$ se tiene por $(\#)$ que

$$f_{(U,V)}(u, v) = \lambda^2 \frac{u}{(1+v)^2} e^{-\lambda u} = (\lambda^2 u e^{-\lambda u}) \left(\frac{1}{(1+v)^2} \right),$$

de donde U y V son independientes.

La $U = X + Y$ tiene densidad $f_U(u) = \lambda^2 u e^{-\lambda u}$, para $u > 0$, que es una distribución GAMMA($\lambda, 2$) (como veremos en el apartado 2.3.4).

La $V = X/Y$ tiene densidad $f_V(v) = 1/(1+v)^2$, para $v > 0$. ♣

2.3. Algunos modelos aleatorios básicos

Describimos a continuación las principales características de cuatro familias de distribuciones (continuas) de probabilidad: uniformes, exponenciales, normales y gammas, atendiendo en especial a tres cuestiones:

- funciones de densidad y de distribución y momentos;
- cambios de escala;
- suma de variables independientes.

2.3.1. Variables uniformes

El modelo continuo más sencillo, y en cierto sentido (véase el ejemplo 2.2.3) el más básico y fundamental, es aquel en el que la variable toma todos los valores de un cierto intervalo con “igual probabilidad”.

A. Función de densidad, de distribución y momentos

Decimos que una variable X sigue una distribución uniforme en el intervalo $[0, a]$, donde $a > 0$, y escribimos $X \sim \text{UNIF}(0, a)$, si su función de densidad viene dada por

$$f_X(x) = \frac{1}{a} \mathbf{1}_{[0,a]}(x) = \begin{cases} 1/a & \text{para } x \in [0, a], \\ 0 & \text{en caso contrario.} \end{cases}$$

La función de distribución viene dada por

$$F_X(x) = \begin{cases} 0 & \text{para } x < 0, \\ x/a & \text{para } x \in [0, a], \\ 1 & \text{para } x > a. \end{cases}$$

La media y la varianza de $X \sim \text{UNIF}(0, a)$ se calculan como sigue:

$$\begin{aligned} \mathbf{E}(X) &= \frac{1}{a} \int_0^a x \, dx = \frac{1}{a} \frac{a^2}{2} = \frac{a}{2}, \\ \mathbf{V}(X) &= \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \frac{1}{a} \int_0^a x^2 \, dx - \left(\frac{a}{2}\right)^2 = \frac{a^2}{3} - \frac{a^2}{4} = \frac{a^2}{12}. \end{aligned}$$

B. Cambio de escala

Si $X \sim \text{UNIF}(0, a)$, entonces $X/a \sim \text{UNIF}(0, 1)$, como se comprueba con la fórmula del cambio de variables.

De manera que todas las distribuciones $\text{UNIF}(0, a)$ se obtienen de la $\text{UNIF}(0, 1)$ cambiando la escala de los valores de la variable (multiplicando por a).

C. Suma de uniformes independientes

A diferencia de los modelos que siguen, en particular las variables normales y las variables de la familia Gamma, la suma de uniformes independientes *no* es uniforme.

Por ejemplo, la suma $Z = X + Y$ de dos variables uniformes en $[0, a]$ independientes tiene función de densidad

$$\begin{aligned} f_Z(z) &= \frac{1}{a^2} \int_{-\infty}^{\infty} \mathbf{1}_{[0,a]}(x) \mathbf{1}_{[0,a]}(z-x) \, dx = \frac{1}{a^2} |[0, a] \cap [z-a, z]| \\ &= \begin{cases} 0 & \text{si } z < 0, \\ z/a^2 & \text{si } 0 \leq z \leq a, \\ (2a-z)/a^2 & \text{si } a \leq z \leq 2a, \\ 0 & \text{si } z > 2a; \end{cases} \end{aligned}$$

que corresponde a una variable conocida como *triangular* en el intervalo $[0, 2a]$.

2.3.2. Variables exponenciales

A. Función de densidad, de distribución y momentos

Si una variable X sigue una distribución exponencial de parámetro $\lambda > 0$, que escribimos $X \sim \text{EXP}(\lambda)$, entonces su función de densidad viene dada por

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

Compactamente, escribimos la función de densidad de $X \sim \text{EXP}(\lambda)$ en la forma

$$f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}}, \quad \text{para todo } x \in \mathbb{R}.$$

La función de distribución de $X \sim \text{EXP}(\lambda)$ viene dada por

$$F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0,$$

de manera que

$$\mathbf{P}(X > x) = 1 - F_X(x) = e^{-\lambda x}, \quad x \geq 0.$$

La media y la varianza de $X \sim \text{EXP}(\lambda)$ vienen dadas por

$$\mathbf{E}(X) = \frac{1}{\lambda} \quad \text{y} \quad \mathbf{V}(X) = \frac{1}{\lambda^2},$$

como se puede comprobar integrando por partes (para un λ general, o quizás únicamente para $\lambda = 1$, y utilizando la observación sobre el cambio de escala que sigue).

B. Cambio de escala en variables exponenciales

Si $X \sim \text{EXP}(\lambda)$, entonces $\lambda X \sim \text{EXP}(1)$, como ya vimos en el ejemplo 2.2.1.

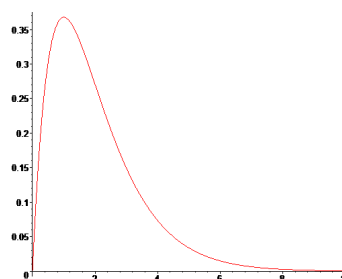
De manera que todas las distribuciones $\text{EXP}(\lambda)$ se obtienen de la $\text{EXP}(1)$ cambiando la escala de los valores de la variable, o si se prefiere, de las unidades con las que se mide X .

C. Suma de exponenciales independientes

Supongamos que X e Y son variables exponenciales independientes, ambas de parámetro $\lambda > 0$. Entonces, aplicando el lema 2.8, para $z > 0$,

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}} \lambda e^{-\lambda(z-x)} \mathbf{1}_{\{z-x \geq 0\}} dx \\ &= \lambda^2 \int_0^z e^{-\lambda z} dx = \lambda^2 z e^{-\lambda z} = \lambda (\lambda z) e^{-\lambda z}, \end{aligned}$$

como ya habíamos obtenido en el ejemplo 2.2.6. El aspecto típico de la función de densidad de una suma $X + Y$ de exponenciales con el mismo parámetro se muestra a la derecha.



La función de densidad obtenida para $X + Y$ no se corresponde con la de una variable exponencial, sino con la de una variable de la familia de las Gammas, en concreto una $\text{GAMMA}(\lambda, 2)$ (véase la sección 2.3.4).



Nota 2.3.1. Si las variables X e Y son variables exponenciales con distintos parámetros e independientes, digamos $X \sim \text{EXP}(\lambda)$ e $Y \sim \text{EXP}(\mu)$, con $\lambda \neq \mu$, entonces la función de densidad de su suma $X + Y$ es una combinación lineal de las respectivas densidades:

$$f_{X+Y}(z) = \frac{\mu}{\mu - \lambda} f_X(z) - \frac{\lambda}{\mu - \lambda} f_Y(z).$$

2.3.3. La normal (unidimensional)

El adjetivo “unidimensional” aparece para distinguir este modelo de modelo homónimo multidimensional que trataremos en el capítulo 3.

A. Función de densidad, de distribución y momentos

Una *variable normal de parámetros* $\mu \in \mathbb{R}$ y $\sigma^2 > 0$ es una variable continua X con valores en todo \mathbb{R} y con función de densidad dada por

$$(2.19) \quad f(x) = \frac{1}{\sqrt{2\pi} \sigma^2} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}.$$

Denotamos $X \sim \mathcal{N}(\mu, \sigma^2)$.



Nota 2.3.2. ¡Atención!, escribimos $\mathcal{N}(\mu, \sigma^2)$ (y no $\mathcal{N}(\mu, \sigma)$) significando como segundo parámetro la varianza, y no la desviación típica.

Reservamos la notación ϕ para la función de densidad de la **normal estándar**, esto es, para $\mathcal{N}(0, 1)$:

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

La comprobación de que $\phi(x)$ encierra área 1, es decir, que

$$\int_{-\infty}^{\infty} \phi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = 1,$$

queda como ejercicio 2.5 para el lector. De esta observación se sigue con facilidad (cambiando variables) que la función definida en (2.19) es también una función de densidad para todo $\mu \in \mathbb{R}$ y $\sigma^2 > 0$.


Si X es una variable normal estándar, entonces X tiene media $\mathbf{E}(X) = 0$ (por simetría alrededor de 0 de la función de densidad) y varianza $\mathbf{V}(X) = 1$, puesto que, integrando por partes,

$$\begin{aligned} \mathbf{E}(X^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x d(-e^{-x^2/2}) \\ &= \frac{1}{\sqrt{2\pi}} (-xe^{-x^2/2}) \Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1. \end{aligned}$$

Este cálculo nos dice también, como comprobaremos en el apartado B, que si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces su media es μ y su varianza, σ^2 . Para otros momentos de orden superior, véase la nota 2.3.4.

La función de distribución de $\mathcal{N}(0, 1)$ se denota por Φ :

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}z^2} dz.$$

 **Nota 2.3.3.** La función de distribución Φ de la normal estándar no tiene fórmula cerrada, pero tenemos las siguientes estimaciones, válidas en general, pero de interés sobre todo cuando $x \rightarrow \infty$: para $x > 0$,

$$\frac{x}{1+x^2} \phi(x) \leq 1 - \Phi(x) \leq \frac{1}{x} \phi(x).$$

Para la cota superior, basta escribir

$$\begin{aligned} 1 - \Phi(x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{1}{z} z e^{-z^2/2} dz \leq \frac{1}{x} \frac{1}{\sqrt{2\pi}} \int_x^\infty z e^{-z^2/2} dz \\ &= \frac{1}{x} \frac{1}{\sqrt{2\pi}} [-e^{-z^2/2}]_x^\infty = \frac{\phi(x)}{x}. \end{aligned}$$

Como consecuencia, para $t > 0$,

$$\mathbf{P}(|X| > t) \leq \frac{2}{t} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

Para la cota inferior, aprovechamos que ϕ satisface la ecuación diferencial

$$(\dagger) \quad \phi'(x) + x\phi(x) = 0, \quad \text{para todo } x \in \mathbb{R}.$$

Supongamos que u es una función positiva, diferenciable en $(0, \infty)$, y tal que

$$(\star) \quad (-u\phi)'(t) \leq \phi(t), \quad \text{para todo } t > 0 \quad \text{y} \quad (\star\star) \quad \lim_{t \rightarrow \infty} u(t)\phi(t) = 0.$$

Entonces, integrando desde x a ∞ , se obtiene que

$$u(x)\phi(x) \leq \int_x^\infty \phi(t) dt = 1 - \Phi(x), \quad \text{para todo } x > 0.$$

Cualquier función u con las propiedades anteriores da una cota. Busquemos una apropiada. Obsérvese primero que

$$(-u\phi)'(t) = -u'(t)\phi(t) - u(t)\phi'(t) = (-u'(t) + tu(t))\phi(t), \quad \text{para todo } t > 0,$$

donde en la segunda igualdad hemos usado (\dagger) . De manera que la condición (\star) se puede escribir como

$$(\star') \quad -u'(t) + tu(t) \leq 1, \quad \text{para todo } t > 0.$$

Por ejemplo, la función $u(t) = t/(1+t^2)$ cumple todas las propiedades requeridas (y nos da la cota del enunciado), pues es positiva y diferenciable en $t > 0$, cumple $(\star\star)$, y en cuanto a (\star') , obsérvese que, para todo $t > 0$,

$$-u'(t) + tu(t) = -\frac{1-t^2}{(1+t^2)^2} + \frac{t^2}{1+t^2} = \frac{-1+2t^2+t^4}{1+2t^2+t^4} \leq 1.$$

B. Tipificación de la normal

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces

$$Y := \frac{X - \mu}{\sigma}$$

es una $\mathcal{N}(0, 1)$.

Para verificar que $Y \sim \mathcal{N}(0, 1)$, usamos la fórmula (2.13) (con $H^{-1}(s) = \sigma s + \mu$) y la expresión de la función de densidad (2.19) de la $\mathcal{N}(\mu, \sigma^2)$ para comprobar que

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} = \phi(y)$$

Y viceversa: si $Y \sim \mathcal{N}(0, 1)$, entonces la variable $X = \mu + \sigma Y$ es $\mathcal{N}(\mu, \sigma^2)$, aplicando el cambio de variables en el otro sentido (fórmula (2.12))

De manera que no hay más que una variable normal: la estándar. Todas las distribuciones $\mathcal{N}(\mu, \sigma^2)$ se obtienen de la normal estándar cambiando la escala de la variable (y trasladándola). Es decir, si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces

$$X = \mu + \sigma Y, \quad \text{donde } Y \sim \mathcal{N}(0, 1).$$

Esto nos dice, en particular, que si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces su media es μ , pues

$$\mathbf{E}(X) = \mathbf{E}(\mu + \sigma Y) = \mu + \sigma \mathbf{E}(Y) = \mu,$$

y su varianza es σ^2 :

$$\mathbf{V}(X) = \mathbf{V}(\mu + \sigma Y) = \mathbf{V}(\sigma Y) = \sigma^2 \mathbf{V}(Y) = \sigma^2.$$

Con más generalidad,

Lema 2.9 Si $X \sim \mathcal{N}(\mu, \sigma^2)$ y definimos

$$Y = aX + b,$$

entonces Y es una variable normal y, de hecho, $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.



Nota 2.3.4. Función generatriz de momentos de normal estándar. Para una normal estándar X , y para todo $\lambda \in \mathbb{R}$,

$$M_X(\lambda) = \mathbf{E}(e^{\lambda X}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\lambda x} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} e^{\lambda^2/2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\lambda)^2} dx = e^{\lambda^2/2},$$

Como

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mathbf{E}(X^k) = M_X(\lambda) = e^{\lambda^2/2} = \sum_{k=0}^{\infty} \lambda^{2k} \frac{1}{2^k k!},$$

donde hemos usado que $e^z = \sum_{k=0}^{\infty} z^k/k!$, se deduce que los momentos de orden impar de X son nulos (como es obvio por simetría), mientras que los momentos de orden par son:

$$\mathbf{E}(X^{2k}) = \frac{(2k)!}{2^k k!}, \quad \text{para } k \geq 0.$$

(Para el cálculo de los momentos pares de una normal estándar, véase también el ejercicio 3.5.)

Esto nos da que, por ejemplo, el coeficiente de asimetría es 0 y la kurtosis vale 3 si $X \sim \mathcal{N}(0, 1)$ (véanse en (2.6) las respectivas definiciones).

Los momentos de una variable $X \sim \mathcal{N}(\mu, \sigma^2)$ genérica se pueden calcular a partir de los momentos de una estándar usando el correspondiente cambio lineal. Por ejemplo, escribiendo Y para una normal estándar,

$$\mathbf{E}(X^2) = \mathbf{E}((\mu + \sigma Y)^2) = \mu^2 + 2\mu\sigma \mathbf{E}(Y) + \sigma^2 \mathbf{E}(Y^2) = \mu^2 + \sigma^2.$$

Registramos, para ulterior referencia, los valores de los momentos de orden 2, 3 y 4 para $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$\begin{aligned} \mathbf{E}(X^2) &= \mu^2 + \sigma^2, & \mathbf{V}(X) &= \sigma^2, \\ \mathbf{E}(X^3) &= \mu^3 + 3\mu\sigma^2, & \mathbf{E}((X - \mu)^3) &= 0, \\ \mathbf{E}(X^4) &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4, & \mathbf{E}((X - \mu)^4) &= 3\sigma^4. \end{aligned}$$

Por sus definiciones, ni el coeficiente de asimetría ni la kurtosis son sensibles a cambios lineales, y por tanto valen 0 y 3, respectivamente, para cualquier variable normal.

C. Suma de normales independientes

Como vamos a ver a continuación, la suma (o la combinación lineal) de variables normales *independientes* es también una normal.

Proposición 2.10 *Si $X \sim \mathcal{N}(\mu, \sigma^2)$ e $Y \sim \mathcal{N}(\nu, \eta^2)$ son variables normales independientes, entonces $X + Y$ es una variable normal. De hecho,*

$$X + Y \sim \mathcal{N}(\mu + \nu, \sigma^2 + \eta^2).$$

En general, si $a, b \in \mathbb{R}$ no ambos nulos,

$$aX + bY \sim \mathcal{N}(a\mu + b\nu, a^2\sigma^2 + b^2\eta^2),$$

DEMOSTRACIÓN. Por la observación del apartado B anterior, podemos escribir $X = \mu + \sigma \tilde{X}$ e $Y = \nu + \eta \tilde{Y}$, donde \tilde{X} e \tilde{Y} son normales estándar independientes. De manera que $X + Y = (\mu + \nu) + (\sigma \tilde{X} + \eta \tilde{Y})$. Vamos a comprobar que $Z = \sigma \tilde{X} + \eta \tilde{Y}$ es una $\mathcal{N}(0, \sigma^2 + \eta^2)$.

Aplicando el lema 2.8 se tiene que

$$f_Z(z) = \int_{\mathbb{R}} f_{\tilde{X}}(x) f_{\tilde{Y}}(z - x) dx = \frac{1}{2\pi\sigma\eta} \int_{\mathbb{R}} e^{-\frac{1}{2}x^2/\sigma^2} e^{-\frac{1}{2}(z-x)^2/\eta^2} dx.$$

Llamando $\omega = \sqrt{\sigma^2 + \eta^2}$, usando la identidad (“completar cuadrados”)

$$\frac{x^2}{\sigma^2} + \frac{(z - x)^2}{\eta^2} = \frac{z^2}{\omega^2} + \left(\frac{\omega}{\eta\sigma} x - \frac{\sigma}{\eta\omega} z \right)^2,$$

y el cambio de variables

$$y = \frac{\omega}{\eta\sigma} x - \frac{\sigma}{\eta\omega} z,$$

se obtiene que

$$\begin{aligned} f_Z(z) &= \frac{1}{2\pi\sigma\eta} e^{-\frac{1}{2}z^2/\omega^2} \left(\frac{\eta\sigma}{\omega}\right) \int_{\mathbb{R}} e^{-\frac{1}{2}\left(\frac{\omega}{\sigma\eta}x - \frac{\sigma}{\eta\omega}z\right)^2} dx \\ &= \frac{1}{2\pi\sigma\eta} e^{-\frac{1}{2}z^2/\omega^2} \left(\frac{\eta\sigma}{\omega}\right) \int_{\mathbb{R}} e^{-\frac{1}{2}y^2} dy = \frac{1}{\omega 2\pi} e^{-\frac{1}{2}z^2/\omega^2} \sqrt{2\pi} = \frac{1}{\omega\sqrt{2\pi}} e^{-\frac{1}{2}z^2/\omega^2}. \end{aligned}$$

Es decir, Z es una variable normal de media 0 y varianza ω^2 , como queríamos.

La segunda parte sigue de la primera sin más que observar que $aX \sim \mathcal{N}(a\mu, a^2\sigma^2)$ y que $bY \sim \mathcal{N}(b\nu, b^2\eta^2)$. ■

Combinando el lema 2.9 y el uso reiterado de la proposición 2.10, obtenemos que cualquier combinación lineal de normales independientes es una variable normal:

Proposición 2.11 *Si X_1, \dots, X_n son variables aleatorias independientes, de manera que cada $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, entonces la variable aleatoria*

$$Z = \sum_{j=1}^n a_j X_j$$

es una normal (unidimensional) tal que

$$\mathbf{E}(Z) = \sum_{j=1}^n a_j \mu_j \quad y \quad \mathbf{V}(Z) = \sum_{j=1}^n a_j^2 \sigma_j^2.$$

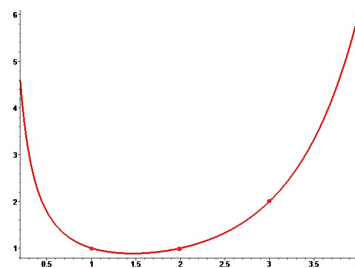
2.3.4. Variables Gamma

Las variables aleatorias de la familia de las Gammas tienen funciones de densidad que dependen de la función Γ de Euler, que describimos primero: su definición y algunas propiedades útiles.

La función Γ es la función definida en $(0, +\infty)$ con valores positivos dada por

$$(2.20) \quad \Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx = \int_0^\infty x^t e^{-x} \frac{dx}{x}.$$

La integral es finita para todo $t > 0$ y $\Gamma(t) > 0$. Obsérvese que $\Gamma(1) = 1$. A la derecha mostramos el aspecto de la función Γ de Euler.



Integrando por partes en la definición (2.20) de la función $\Gamma(t)$ se obtiene que, para todo $t > 1$,

$$(2.21) \quad \Gamma(t) = (t-1)\Gamma(t-1).$$

En particular, para entero $n \geq 1$ se tiene que

$$\Gamma(n) = (n-1)!,$$

que es consistente con la convención por la que $0! = 1$. La función Γ es una función continua (convexa) que interpola los factoriales de los números enteros.

Los valores de la función $\Gamma(t)$ en los semienteros son también relevantes.

Proposición 2.12 $\Gamma(1/2) = \sqrt{\pi}$.

DEMOSTRACIÓN. El cambio de variables $x = y^2/2$ da que

$$\Gamma(1/2) = \int_0^\infty \frac{e^{-x}}{\sqrt{x}} dx = \sqrt{2} \int_0^\infty e^{-y^2/2} dy = \frac{1}{\sqrt{2}} \int_{-\infty}^\infty e^{-y^2/2} dy = \frac{1}{\sqrt{2}} \sqrt{2\pi} = \sqrt{\pi},$$

donde se ha usado que $e^{-y^2/2}$ es una función par y que $\int_{\mathbb{R}} e^{-y^2/2} dy = \sqrt{2\pi}$. ■

Proposición 2.13 Para $k \geq 0$,

$$\Gamma(k + \frac{1}{2}) = \frac{(2k)!}{4^k k!} \sqrt{\pi} \quad \text{para cada entero } k \geq 0.$$

DEMOSTRACIÓN. Aplicamos reiteradamente la regla de recursión:

$$\begin{aligned} \Gamma(k + \frac{1}{2}) &= (k - \frac{1}{2}) \Gamma(k - \frac{1}{2}) = (k - \frac{1}{2})(k - \frac{3}{2}) \Gamma(k - \frac{3}{2}) \\ &= \cdots = (k - \frac{1}{2})(k - \frac{3}{2}) \cdots (\frac{1}{2}) \Gamma(\frac{1}{2}) = (k - \frac{1}{2})(k - \frac{3}{2}) \cdots (\frac{1}{2}) \sqrt{\pi} \\ &\stackrel{(*)}{=} \frac{1}{2^k} (2k-1)(2k-3) \cdots 3 \cdot 1 \cdot \sqrt{\pi} = \frac{1}{2^k} \frac{(2k)!}{2k(2k-2) \cdots 4 \cdot 2} \sqrt{\pi} = \frac{(2k)!}{4^k k!} \sqrt{\pi}. \end{aligned}$$

Hemos usado el valor de $\Gamma(1/2)$ (proposición 2.12) y, en el paso marcado con un asterisco, que hay exactamente k factores. El resto es pura cosmética factorial. ■

Registramos, por último, una variación de la integral que define la función $\Gamma(t)$ que resultará útil más adelante.

Proposición 2.14 Para $b > 0$ y $t > 0$ se tiene que

$$\int_0^\infty x^t e^{-bx} \frac{dx}{x} = \frac{\Gamma(t)}{b^t}.$$

DEMOSTRACIÓN. Cambio de variables $bx = y$. ■

A. La familia Gamma de distribuciones de probabilidad

Una variable $\text{GAMMA}(\lambda, t)$ (con parámetros $\lambda, t > 0$) es una variable aleatoria continua con función de densidad dada para $x > 0$ por

$$(2.22) \quad f_{\lambda,t}(x) = \frac{1}{\Gamma(t)} \lambda^t x^{t-1} e^{-\lambda x} = \frac{1}{\Gamma(t)} (\lambda x)^t e^{-\lambda x} \frac{1}{x},$$

y $f_{\lambda,t}(x) = 0$ para $x \leq 0$.

(La proposición 2.14 nos confirma que $\int_0^\infty f_{\lambda,t}(x) dx = 1$ para cualesquiera valores de los parámetros $\lambda, t > 0$.)



Nota 2.3.5. Conviven en la literatura científica diversas formas de nombrar a los parámetros de una distribución Gamma. Hay quien habla de $\text{GAMMA}(\alpha, \beta)$, que se corresponde a tomar $\alpha = \lambda$ y $\beta = t$ en nuestra notación; y hay quien utiliza $\text{GAMMA}(k, \theta)$, donde $k = \lambda$ y, ¡atención!, $\theta = 1/t$.

B. Cambios de escala en variables Gamma

Si $X \sim \text{GAMMA}(\lambda, t)$ entonces, por el ejemplo 2.2.1, para cualquier $c > 0$, la variable reescalada cX sigue una distribución $\text{GAMMA}(\lambda/c, t)$.

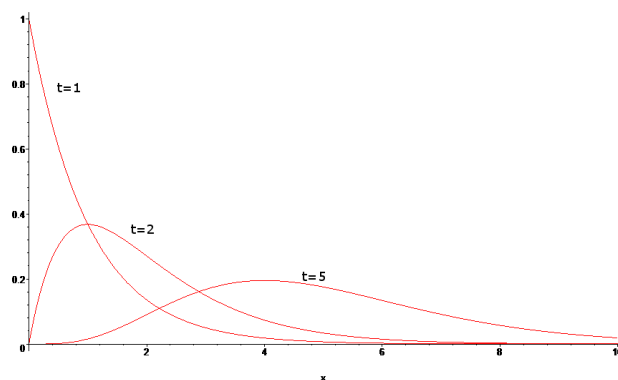
En particular,

$$\text{si } X \sim \text{GAMMA}(\lambda, t), \text{ entonces } \lambda X \sim \text{GAMMA}(1, t).$$

Los parámetros λ y t de una variable Gamma se conocen como:

- λ es el parámetro de “escala”. Según la observación anterior, todas las distribuciones $\text{GAMMA}(\lambda, t)$ se obtienen de la $\text{GAMMA}(1, t)$ cambiando la escala (las unidades) de la variable.
- Mientras que a t se le llama el parámetro de “forma”.

En la siguiente figura aparecen tres casos de una $\text{GAMMA}(1, t)$:



C. Momentos de variables Gamma

El cálculo de los momentos de una variable $X \sim \text{GAMMA}(\lambda, t)$ es directo: para cada entero $k \geq 1$,

$$\begin{aligned} \mathbf{E}(X^k) &= \int_0^\infty x^k \frac{1}{\Gamma(t)} \lambda^t x^{t-1} e^{-\lambda x} \frac{dx}{x} = \frac{\lambda^t}{\Gamma(t)} \int_0^\infty x^{k+t-1} e^{-\lambda x} \frac{dx}{x} \\ (2.23) \quad &\stackrel{(*)}{=} \frac{\lambda^t}{\Gamma(t)} \frac{\Gamma(t+k)}{\lambda^{t+k}} = \frac{\Gamma(t+k)}{\Gamma(t)} \frac{1}{\lambda^k} \stackrel{(**)}{=} \frac{(t+k-1) \cdots t}{\lambda^k}, \end{aligned}$$

donde en $(*)$ hemos usado la proposición 2.14, y en $(**)$ hemos aplicado reiteradamente la recursión (2.21) de la función Γ de Euler.

En particular,

$$(2.24) \quad \mathbf{E}(X) = \frac{t}{\lambda} \quad \text{y} \quad \mathbf{V}(X) = \frac{t}{\lambda^2},$$

pues $\mathbf{E}(X^2) = (t+1)t/\lambda^2$.

Nos interesará también el siguiente cálculo, para $t > 1$:

$$(2.25) \quad \mathbf{E}(1/X) = \frac{\lambda^t}{\Gamma(t)} \int_0^\infty x^{t-1} e^{-\lambda x} \frac{dx}{x} = \frac{\lambda^t}{\Gamma(t)} \frac{\Gamma(t-1)}{\lambda^{t-1}} = \frac{\lambda}{t-1},$$

en el que hemos usado de nuevo la proposición 2.14 y la recursión para $\Gamma(t)$.

D. Suma de variables Gamma independientes

Proposición 2.15 *Si X e Y son, respectivamente, variables aleatorias $\text{GAMMA}(\lambda, t)$ y $\text{GAMMA}(\lambda, s)$ (con el mismo parámetro λ) e independientes, entonces $X + Y$ es $\text{GAMMA}(\lambda, t + s)$.*

DEMOSTRACIÓN. Podemos suponer que $\lambda = 1$. Para $x > 0$ se tiene que

$$\begin{aligned} f_{X+Y}(x) &= \frac{1}{\Gamma(t)} \frac{1}{\Gamma(s)} \int_0^x y^{t-1} e^{-y} e^{-(x-y)} (x-y)^{s-1} dy \\ &= \frac{e^{-x}}{\Gamma(t)\Gamma(s)} \int_0^x y^{t-1} (x-y)^{s-1} dy \\ &\stackrel{[\text{cambio } y = xz]}{=} \frac{x^{t+s-1} e^{-x}}{\Gamma(t)\Gamma(s)} \int_0^1 z^{t-1} (1-z)^{s-1} dz \\ &= \frac{\Gamma(t+s) \int_0^1 z^{t-1} (1-z)^{s-1} dz}{\Gamma(t)\Gamma(s)} \left\{ \frac{x^{t+s-1} e^{-x}}{\Gamma(t+s)} \right\} \\ &= \frac{\Gamma(t+s) \int_0^1 z^{t-1} (1-z)^{s-1} dz}{\Gamma(t)\Gamma(s)} \cdot f_{1,t+s}(x) \end{aligned}$$

Como f_{X+Y} y $f_{1,t+s}$ son funciones de densidad, sus integrales valen 1, de lo que deducimos primero que

$$\frac{\Gamma(t+s) \int_0^1 z^{t-1} (1-z)^{s-1} dz}{\Gamma(t)\Gamma(s)} = 1,$$

y luego que

$$f_{X+Y}(x) = f_{1,t+s}(x),$$

para todo $x > 0$, como queríamos ver. ■



Nota 2.3.6. Como bonus de la demostración se obtiene la relación entre la función gamma y la llamada *función beta*:

$$\text{Beta}(t, s) := \int_0^1 z^{t-1} (1-z)^{s-1} dz = \frac{\Gamma(t)\Gamma(s)}{\Gamma(t+s)}.$$

E. Casos particulares

- **El caso $t = 1$.** Una variable $\text{GAMMA}(\lambda, 1)$ es una $\text{EXP}(\lambda)$.

Lo que nos da, usando las fórmulas de (2.24), que si $Z \sim \text{EXP}(\lambda)$, con $\lambda > 0$,

$$\mathbf{E}(Z) = \frac{1}{\lambda} \quad \text{y} \quad \mathbf{V}(Z) = \frac{1}{\lambda^2},$$

como ya se mencionó en el apartado 2.3.2.

Además, por la proposición 2.15, si X e Y son variables $\text{EXP}(\lambda)$ independientes, entonces $X + Y \sim \text{GAMMA}(\lambda, 2)$. Repase, lector, el ejemplo 2.2.6.

- **El caso $\lambda = 1/2$ y $t = n/2$, con $n \geq 1$ entero.** Una variable $\text{GAMMA}(1/2, n/2)$ es una χ_n^2 (una chi-cuadrado con n grados de libertad, véase la sección 3.3.1).

Las fórmulas de (2.24) dan que, si $Z \sim \chi_n^2$, para entero $n \geq 1$,

$$\mathbf{E}(Z) = n \quad \text{y} \quad \mathbf{V}(Z) = 2n$$

(véase un argumento alternativo en la sección 3.3.1). Y, además, por (2.25),

$$\mathbf{E}\left(\frac{1}{Z}\right) = \frac{1}{n-2},$$

para $n \geq 3$ (para $n = 1, 2$ se tiene que $\mathbf{E}(1/Z) = +\infty$).

Finamente, aplicando la proposición 2.15, si X e Y son variables χ_n^2 y χ_m^2 independientes, entonces $X + Y \sim \chi_{n+m}^2$, pues $\chi_n^2 \sim \text{GAMMA}(1/2, n/2)$ y $\chi_m^2 \sim \text{GAMMA}(1/2, m/2)$. Esta observación será inmediata con la definición de variables chi-cuadrado de la sección 3.3.1.

2.4. Simulación de variables aleatorias

Simular una muestra (independiente) de una variable aleatoria X es el proceso de generar números que

- se ajusten aproximadamente a la ley de X ,
- y que “parezcan” muestras independientes:

“Ajustarse aproximadamente a la ley de X ” significa:

- en el caso de variables discretas X , que cuando la muestra sea *grande*, cada valor $x_j \in \text{sop}(X)$ aparezca aproximadamente en la proporción p_j correspondiente;
- en el caso de variables continuas X , que cuando la muestra sea *grande*, ocurra que en cada intervalo $(a, b]$ aparezca aproximadamente la proporción correspondiente: $F_X(b) - F_X(a)$, es decir,

$$\frac{\#\{x_i \in (a, b]; 1 \leq i \leq n\}}{n} \approx F(b) - F(a).$$

Los datos de la muestra han de ser además (aproximadamente) independientes, en el sentido de que para pares se tenga

$$\frac{\#\{(x_{2i-1}, x_{2i}) \in (a, b] \times (c, d]; 1 \leq i \leq n\}}{n} \approx (F(b) - F(a))(F(d) - F(c)).$$

Y análogamente para triples, etc.

Vamos a ver a continuación que existe un mecanismo “universal” de simulación, que se aplica a cualquier variable aleatoria.

Hipótesis básica: disponemos de una máquina que genera muestras que siguen la ley de una uniforme $\text{UNIF}(0, 1)$. Es decir, que cuando el tamaño n de la muestra (u_1, \dots, u_n) es muy grande, para cada intervalo (a, b) , con $0 \leq a < b \leq 1$, la proporción de datos que caen en ese intervalo es aproximadamente $b - a$:

$$\frac{\#\{u_i \in (a, b]; 1 \leq i \leq n\}}{n} \approx b - a.$$

Además, para la apariencia de independencia,

$$\frac{\#\{(u_{2i-1}, u_{2i}) \in (a, b] \times (c, d]; 1 \leq i \leq n\}}{n} \approx (b - a)(d - c).$$

Y análogamente para triples, etc.

Al procedimiento⁴ que genera estas muestras (que siguen la ley de una $\text{UNIF}[0, 1]$ con apariencia de independencia) nos referiremos como la MÁQUINA U. Cada lenguaje de programación/software tiene su propia MÁQUINA U. En Excel, la MÁQUINA U viene dada por la instrucción `=aleatorio()`.

Vamos a ver cómo con una MÁQUINA U podemos simular cualesquiera variables.

2.4.1. Simulación de variables discretas

A. El caso de la Bernoulli de parámetro p

Supongamos que X toma los valores 1 y 0 con probabilidades respectivas p y $1 - p$, es decir, $X \sim \text{BER}(p)$.

Código de simulación de una muestra (x_1, \dots, x_n) que sigue la ley de X :

- Obtenemos una muestra (u_1, \dots, u_n) con la MÁQUINA U.
- Para cada $i = 1, \dots, n$: si $u_i \leq p$, declaramos $x_i = 1$; y declaramos $x_i = 0$ en caso contrario.

⁴Estos generadores utilizan algoritmos *deterministas*. Por eso se conocen como generadores *pseudoaleatorios*.

B. El caso de una variable discreta (finita) general

La variable X toma valores a_1, \dots, a_k con probabilidades p_1, \dots, p_k .

Código de simulación de una muestra (x_1, \dots, x_n) que sigue la ley de X :

- Dividimos el intervalo $[0, 1]$ en k subintervalos:

$$I_1 = [0, p_1), I_2 = [p_1, p_1 + p_2), I_3 = [p_1 + p_2, p_1 + p_2 + p_3), \dots, I_k = [p_1 + \dots + p_{k-1}, 1].$$

- Obtenemos una muestra (u_1, \dots, u_n) con la MÁQUINA U.
- Para cada $i = 1, \dots, n$: si $u_i \in I_j$, declaramos $x_i = a_j$.

C. El caso de la binomial de parámetros n y p

Los valores de una $X \sim \text{BIN}(k, p)$ son $\{0, 1, \dots, k\}$, con probabilidades respectivas

$$p_j = \mathbf{P}(X = j) = \binom{k}{j} p^j (1 - p)^{k-j} \quad \text{para } j = 0, 1, \dots, k.$$

Podemos simular muestras como en el apartado anterior, pues se trata de una variable discreta.

Alternativamente, recordando que una binomial es una suma de Bernoullis independientes, podemos sortear (independientemente) k variables $\text{BER}(p)$ y sumar los resultados para obtener *un* número que sigue la ley de la binomial.

2.4.2. Simulación de variables continuas

A. El resultado básico

Digamos que X es una variable aleatoria continua, con función de densidad f_X y función de distribución F_X . Pongámonos, para empezar, en el caso en el que F_X es un homeomorfismo de \mathbb{R} en $(0, 1)$ (una biyección de \mathbb{R} sobre $(0, 1)$ tal que tanto la función como su inversa son continuas); ése es el caso, por ejemplo, de una normal. La variable X toma valores en todo \mathbb{R} .

Consideramos la variable $Y = F_X^{-1}(U)$, donde U es una uniforme $U(0, 1)$. La variable Y también toma valores en todo \mathbb{R} . Ahora, para cualquier $y \in \mathbb{R}$,

$$F_Y(y) = \mathbf{P}(Y \leq y) = \mathbf{P}(F_X^{-1}(U) \leq y) \stackrel{\text{homeomorfismo}}{=} \mathbf{P}(U \leq F_X(y)) \stackrel{\text{uniforme}}{=} F_X(y),$$

de manera que Y y X son variables aleatorias con la misma función de distribución (véase el ejemplo 2.2.3). Escribimos $X \stackrel{d}{=} Y$.

Esto permite definir el siguiente método “universal” de simulación:

- Obtenemos una muestra (u_1, \dots, u_n) con la MÁQUINA U.
- Para cada $i = 1, \dots, n$, declaramos $x_i = F_X^{-1}(u_i)$. La muestra (x_1, \dots, x_n) así obtenida sigue la ley de X .

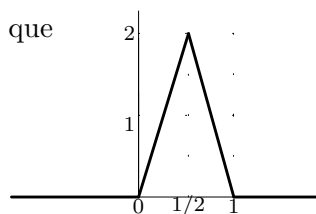
Comentarios:

- Si la variable X toma sólo valores en $[0, \infty)$, es decir, si $F_X(x) = 0$ para $x < 0$ y F_X es homeomorfismo de $(0, \infty)$ sobre $(0, 1)$, entonces $Y = F_X^{-1}(U)$ toma sólo valores en $[0, \infty)$, y $X \stackrel{d}{=} Y$.
- Si X toma valores en un intervalo $[a, b]$ y F_X es homeomorfismo de (a, b) sobre $(0, 1)$, entonces $X \stackrel{d}{=} F_X^{-1}(U)$.
- Para que el mecanismo de simulación anterior sea implementable, se necesita una “fórmula” para la inversa F_X^{-1} de la función de distribución, o al menos un procedimiento eficiente para calcular $F_X^{-1}(u)$ para cada $u \in (0, 1)$.

EJEMPLO 2.4.1. *Muestras de una variable triangular.*

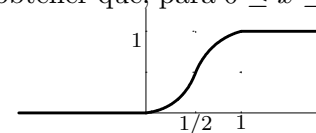
Digamos que X toma valores en el intervalo $[0, 1]$, y que su función de densidad es

$$f_X(x) = \begin{cases} 4x & \text{si } 0 \leq x \leq 1/2, \\ 4(1-x) & \text{si } 1/2 < x \leq 1, \\ 0 & \text{resto.} \end{cases}$$



Para calcular la función de distribución, integramos la de densidad (teniendo cierto cuidado, por la definición a trozos de $f(x)$), para obtener que, para $0 \leq x \leq 1$,

$$F_X(x) = \int_{-\infty}^x f_X(y) dy = \begin{cases} 2x^2 & \text{si } 0 \leq x \leq 1/2, \\ 1 - 2(1-x)^2 & \text{si } 1/2 < x \leq 1, \end{cases}$$



Obsérvese que $F(x)$ se especifica asimismo a trozos. Finalmente, calculamos una expresión para F^{-1} invirtiendo las expresiones anteriores:

$$F^{-1}(u) = \begin{cases} \sqrt{u/2} & \text{si } u < 1/2; \\ 1 - \sqrt{(1-u)/2} & \text{si } u > 1/2. \end{cases}$$

Esta fórmula permite transformar muestras (u_1, \dots, u_n) de la MÁQUINA U en muestras (x_1, \dots, x_n) de X , tomando $x_i = F^{-1}(u_i)$ para cada $i = 1, \dots, n$. ♣

B. Uniformes

Para simular muestras la uniforme en $[0, a]$, basta producir muestras (u_1, \dots, u_n) de la MÁQUINA U, y multiplicar cada número por a . Usamos aquí el cambio de escala discutido en el apartado 2.3.1.

Más generalmente, podemos producir muestras de una variable uniforme en el intervalo $[a, b]$, con $a < b$, y $a, b \in \mathbb{R}$ (esta variable tiene como función de densidad $f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$) transformando muestras (u_1, \dots, u_n) de la MÁQUINA U en muestras (x_1, \dots, x_n) con el cambio $x_i = (b-a)u_i + a$, para cada $i = 1, \dots, n$.

C. Exponenciales

Si $X \sim \text{EXP}(\lambda)$, con $\lambda > 0$, la función de distribución de X viene dada por

$$F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0,$$

y $F_X(x) = 0$, si $x \leq 0$. La inversa F_X^{-1} se calcula explícitamente: para $u \in [0, 1)$,

$$F_X^{-1}(u) = -\frac{\ln(1-u)}{\lambda}.$$

D. Normales

Para la normal estándar, $\mathcal{N}(0, 1)$, la función de distribución viene dada por

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy.$$

Los valores de Φ se calculan numéricamente, y están incorporados en cualquier software científico. En Excel, $\Phi(x)$ se obtiene como `distr.norm.estand(x)`.

Los valores de la inversa de Φ también se calculan numéricamente. En Excel, $\Phi^{-1}(u)$ se obtiene con `distr.norm.estand.inv(u)` o con `inv.norm(u)`.

Si (u_1, \dots, u_n) es una muestra de la MÁQUINA U, entonces

- $(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$ es una muestra de $X \sim \mathcal{N}(0, 1)$;
- y $(\mu + \sigma\Phi^{-1}(u_1), \dots, \mu + \sigma\Phi^{-1}(u_n))$ es una muestra de $X \sim \mathcal{N}(\mu, \sigma^2)$.

E. Gammas

Si $X \sim \text{GAMMA}(\lambda, t)$, su función de densidad es, para $x > 0$,

$$f_{\lambda,t}(x) = \frac{1}{\Gamma(t)} \lambda^t x^{t-1} e^{-\lambda x},$$

y por tanto la función de distribución viene dada por

$$F_{\lambda,t}(x) = \frac{1}{\Gamma(t)} \lambda^t \int_0^x z^{t-1} e^{-\lambda z} dz,$$

cuyos valores, como en el caso de la normal, se calculan numéricamente. En Excel, el valor $F_{\lambda,t}(x)$ se calcula con la instrucción `=distr.gamma(x;λ;1/t;verdadero)`. Nótese el cambio en el parámetro t .

La inversa $F_{\lambda,t}^{-1}(u)$ también se calcula numéricamente, usando bien la instrucción `=inv.gamma(u;λ;1/t)` (en las nuevas versiones de Excel), o bien la instrucción `=distr.gamma.inv(u;λ;1/t)` en las antiguas.