# Temas 3 y 4. Variables aleatorias

Estadística

Grado en Ingeniería Informática

---

## Índice

En este tema se presentan los conceptos fundamentales para el estudio de variables aleatorias, tanto discretas como continuas. Si bien en el programa de la materia se incluyen dos temas diferenciados (uno para variables aleatorias discretas y otro para continuas), los contenidos teóricos de ambos se recogen en este documento para facilitar la comprensión y la construcción de relaciones entre los distintos tipos de variables y, de manera específica, de algunas distribuciones notables.

## 1. Concepto de variable aleatoria

Dado un experimento aleatorio, una variable aleatoria es una aplicación que asocia a cada elemento del espacio muestral de un experimento aleatorio un número. De manera informal, una variable aleatoria nos permite *traducir a números* los resultados de un experimento aleatorio. Así, en el lanzamiento de un dado, nos podría interesar el valor que se obtiene, o simplemente analizar si es par o impar; de manera análoga, en el lanzamiento de dos monedas, podríamos contar el número de caras que se obtienen. Estos serían ejemplos de variables aleatorias: valor al lanzar un dado, indicadora de si es par o impar y número de caras al lanzar dos monedas.

Dependiendo de las posibles asignaciones numéricas a la variable, distinguiremos entre variables aleatorias discretas y continuas. Formalmente, si denotamos por $(\Omega, \mathcal{A}, \mathbb{P})$ el espacio de probabilidad de un experimento aleatorio ($\Omega$ es el espacio muestral, $\mathcal{A}$ es la $\sigma-$álgebra de sucesos y $\mathbb{P}$ la probabilidad asociada), una variable aleatoria $X$ es una aplicación que asocia a cada suceso elemental un número real, de manera que para cualquier real $r$, el conjunto:

$$\{\omega \in \Omega : X(\omega) \leq r\} = X^{-1}((-\infty, r])$$

es también un suceso. El espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$ contiene todos los elementos que necesitamos para caracterizar el experimento aleatorio en estudio. Así, en $\Omega$ se tienen todos los sucesos elementales; en la $\sigma$-álgebra de sucesos están todos los sucesos posibles, incluyendo aquellos que se obtienen a partir de las operaciones usuales con sucesos que se han visto en el capítulo anterior y con $\mathbb{P}$ se conocería la probabilidad de ocurrencia de cada suceso.

**Definición 1.** *Dado un experimento aleatorio con espacio de probabilidad* $(\Omega, \mathcal{A}, \mathbb{P})$*, una variable aleatoria* $X$ *es una función*

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathbb{B}, \mathbb{P}_X),$$

*tal que si* $B \in \mathbb{B}$ *(* $\mathbb{B}$ *es el conjunto de todos los intervalos de* $\mathbb{R}$*), entonces* $X^{-1}(B) \in \mathcal{A}$ *y por tanto* $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$.

Dependiendo del tipo de valores que tomen, distinguiremos entre variables aleatorias discretas y continuas. En el siguiente ejemplo se presenta una variable aleatoria discreta muy sencilla.

**Ejemplo.** Consideremos el experimento aleatorio que consiste en lanzar dos monedas y como variable aleatoria tomemos el numero de caras. Los posibles valores de la variable son Sop$(X) = \{0, 1, 2\}$ y la probabilidad de cada uno de ellos es:

$$\mathbb{P}_X(0) = \mathbb{P}((+,+)) = 1/4, \quad \mathbb{P}_X(1) = \mathbb{P}((c,+) \cup (+,c)) = 1/2, \quad \mathbb{P}_X(2) = \mathbb{P}((c,c)) = 1/4$$

Consideremos distintos intervalos $B \in \mathbb{B}$ y veamos cómo es su imagen inversa a través de la variable aleatoria:

- $B = (-1, 1/2) \Rightarrow X^{-1}(B) = X^{-1}(0) = (+,+)$ porque de los posibles valores que puede tomar $X$ (es decir, los valores de Sop$(X)$), el intervalo solo contiene al cero.
- $B = [1, 2) \Rightarrow X^{-1}(B) = \{(c,+), (+,c)\}$, porque $B$ solo contiene al 1.
- $B = (-3, 35) \Rightarrow X^{-1}(B) = \Omega$.
- $B = (-3, -2) \Rightarrow X^{-1}(B) = \emptyset$.

---

**Ejercicio.** Consideremos un experimento que consiste en lanzar dos dados y registrar el menor valor. Es decir, la variable $X$ será el menor valor en el lanzamiento de dos dados. £Cuál es su soporte? Calcula $X^{-1}(B)$ con los intervalos del ejemplo anterior.

---

En muchas ocasiones, los resultados de un experimento aleatorio no solo se presentan a través de los valores de una variable aleatoria, si no que a estos se les aplica alguna función, o bien se combinan varias variables. En este sentido, algunas propiedades de las variables aleatorias son las siguientes:

1. Si $X$ es una variable aleatoria sobre $(\Omega, \mathcal{A}, \mathbb{P})$ y $c$ es una constante, entonces $cX$ también es una variable aleatoria. Por ejemplo, imaginemos que en el ejemplo del lanzamiento de dos monedas, recibimos 3 euros según el valor que obtenga $X$. Es decir, la variable de interés ahora es $3X$, que tomará valores $\{0, 3, 6\}$, con probabilidades $\{1/4, 1/2, 1/4\}$ respectivamente.
2. Si $X$ e $Y$ son variables aleatorias sobre $(\Omega, \mathcal{A}, \mathbb{P})$, entonces $X + Y$ y $XY$ son también variables aleatorias. Supongamos que lanzamos un dado y registramos el número que sale (variable $X$) y si es par o impar (variable $Y$, con valor 1 si es par y 0 si es impar). La suma de las dos variables es una variable aleatora con valores $\{1, 3, 5, 7\}$, al igual que lo es el producto, que toma valores $\{0, 2, 4, 6\}$.

## 2. Variable aleatoria discreta

Si $X$ es una variable aleatoria (v.a.) sobre $(\Omega, \mathcal{A}, \mathbb{P})$, y solo toma valores en un conjunto finito (o infinito numerable) entonces diremos que $X$ es una variable aleatoria discreta. Las variables aleatorias discretas se caracterizan conociendo el conjunto de sus posibles valores y las probabilidades asociadas. Al conjunto de posibles valores se le denomina soporte, y se denota por $Sop(X)$.

### 2.1. Masa de probabilidad

Si $x_1, \dots, x_K$ son los posibles valores que toma una v.a. discreta, $Sop(X) = \{x_1, \dots, x_K\}$, se le denomina función de masa de probabilidad al conjunto de probabilidades $p_1, \dots, p_K$ tales que:

$$\mathbb{P}(X = x_i) = p_i, i = 1, \dots, K \quad \text{donde, } p_i \geq 0, \sum_{i=1}^{K} p_i = 1.$$

---

**Ejercicio.** Los ejemplos que se han comentado anteriormente (número de caras en el lanzamiento de dos monedas, resultado del lanzamiento de un dado, menor valor al lanzar dos dados,...) son ejemplos de variables aleatorias discretas. Para todas ellas, el soporte (conjunto de posibles valores) es finito. Escribe los valores del soporte en cada caso y la masa de probabilidad asociada.

---

### 2.2. Función de distribución

Como se ha comentado, el comportamiento de una variable aleatoria se puede caracterizar conociendo qué valores puede tomar y con qué probabilidad. Del mismo modo, el comportamiento de una variable aleatoria también se puede describir a través de la función de distribución. La función de distribución de una v.a. $X$ es una función que a cada $x \in \mathbb{R}$ le asocia la probabilidad de que la variable tome valores menores o iguales a dicho número:

$$F(x) = \mathbb{P}_X(X \leq x) = \mathbb{P}(\{A \in \mathcal{A} : X(A) \leq x\}) = \mathbb{P}(X^{-1}((-\infty, x])).$$

Se trata por tanto de una función de probabilidad acumulada. Algunas propiedades de la función de distribución son las siguientes:

1. $0 \leq F(x) \leq 1$.
2. $F$ es no decreciente.
3. $F(+\infty) = \lim_{x \to \infty} F(x) = 1$ y $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$.
4. $F$ es continua por la derecha.

La función de distribución de una v.a. $X$ en un punto $x$ nos da la probabilidad acumulada hasta este valor. Esta función toma valores entre 0 y 1, y es no decreciente.

---

**Ejemplo.** La función de distribución de la variable que cuenta el número de caras en el lanzamiento de dos monedas es una función escalonada (constante a trozos) que toma valores:

$$F(x) = \begin{cases} 0, & \text{si } x < 0, \\ 1/4, & \text{si } x \in [0,1) \\ 3/4, & \text{si } x \in [1,2) \\ 1, & \text{si } x \geq 2 \end{cases}$$

A partir de la función de distribución, es fácil identificar la masa de probabilidad. En concreto, se puede ver que la función tiene saltos en los valores del soporte de la variable (0, 1 y 2 en este caso). Además, la probabilidad asociada a cada valor viene dada por la altura del salto.

---

**Nota.** Para definir la función de distribución no se tiene en cuenta el carácter discreto o continuo de la variable aleatoria de interés. Por tanto, la definición será la misma para el caso de variables continuas.

---

### 2.3. Medidas características

Al igual que en estadística descriptiva se utilizan medidas características para sintentizar la información contenida en una muestra, en el estudio de variables aleatorias también se pueden construir medidas que ayuden a resumir y caracterizar su comportamiento. Tal y como ocurre a nivel muestral al analizar medidas descriptivas, se puede diferenciar entre medidas de localización o posición, dispersión y forma. En el contexto de variables aleatorias se prestará especial atención a la media poblacional o esperanza matemática, como medida de posición de tendencia central y a la varianza, como medida de dispersión.

Sea $X$ es una v.a. discreta, con valores $x_1, \ldots, x_K$ y con masa de probabilidad $p_1, \ldots, p_K$. Como medida de posición de tendencia central, se define la media o esperanza matemática:

$$\mathbb{E}(X) = \mu = \sum_{i=1}^{K} x_i p_i.$$

La esperanza matemática tiene propiedades similares a la media aritmética, como por ejemplo, la linealidad:

1. $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b, \quad a, b \in \mathbb{R}.$
2. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$

Otras medidas características son la mediana y la moda. La mediana es el valor $Me$ que divide la distribución en dos partes iguales: $F(Me) = \frac{1}{2}$. La moda es el valor del soporte donde la función de masa de probabilidad es máximo.

Como medidas de dispersión, se consideran la varianza y la desviación típica:

$$\sigma^2 = \mathbb{V}\text{ar}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] \quad \text{y} \quad \sigma = +\sqrt{\mathbb{E}[(X - \mathbb{E}(X))^2]}.$$

Para variables discretas:

$$\mathbb{V}\text{ar}(X) = \sum_{i=1}^{K} (x_i - \mu)^2 p_i = \sum_{i=1}^{K} (x_i - \sum_{j=1}^{K} x_j p_j)^2 p_i.$$

En cuanto a las propiedades de la varianza, se tienen las siguientes:

1. $\mathbb{V}\text{ar}(aX + b) = a^2 \mathbb{V}\text{ar}(X)$.
2. $\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \mathbb{E}(X^2) - \mu^2$.

Esta segunda propiedad, en el caso de variables discretas, la varianza se escribiría como:

$$\mathbb{V}\text{ar}(X) = \sum_{i=1}^{K} x_i^2 p_i - \left( \sum_{i=1}^{K} x_i p_i \right)^2$$

---

**Nota.** Las propiedades de la esperanza y de la varianza se mantienen para variables aleatorias continuas, aunque la fórmula de cálculo veremos que es diferente. La definición de mediana también es la misma, ya que está dada en términos de la función de distribución.

---

### 2.4. Variables tipificadas

Si tenemos una v.a. $X$ con media $\mu$ y desviación típica $\sigma$, podemos transformarla en una variable que tenga media 0 y varianza 1 (variable estandarizada). Para estandarizarla, debemos restar la media y dividir por la desviación típica:

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

Por las propiedades descritas anteriormente es fácil comprobar que $\mathbb{E}(Y) = 0$ y $\mathbb{V}\text{ar}(Y) = 1$.

---

**Nota.** La tipificación o estandarización de variables aleatorias se puede aplicar también a variables continuas.

---

# 3. Variable aleatoria continua

Una v.a. continua es aquella que toma valores en un intervalo (o varios intervalos) de la recta real, y por tanto, $Sop(X) \subset \mathbb{R}$. Para caracterizar su comportamiento se puede considerar la función de distribución o la función de densidad.

## 3.1. Función de distribución

La función de distribución de una v.a. continua se define de igual manera a la de una v.a. discreta, es decir, la función de distribución de una v.a. $X$ es una función que a cada valor real $x$ le asocia la probabilidad de que la variable tome valores menores o iguales a dicho número: $F(x) = \mathbb{P}(X \leq x)$.

---

**Ejemplo.** Supongamos que $X$ es la altura (en cm) de una persona en una cierta población y supongamos que $F$ es su función de distribución. £Cómo interpretamos $F(160)$ o $F(170)$? Si $F(160) = 1/4$ y $F(170) = 1/2$, £puedes calcular la probabilidad de que una person mida entre 160 y 170 cm?

El valor de $F(160) = \mathbb{P}(X \leq 160)$, por definición. Por tanto, nos daría la probabilidad de que una persona mida como mucho 160 cm. De manera análoga podemos interpretar $F(170)$. Por otro lado, la probabilidad de que una persona mida entre 160 y 170 la podemos escribir como: $\mathbb{P}(160 \leq X \leq 170)$, que en términos de distribución es igual que $\mathbb{P}(X \leq 170) - \mathbb{P}(X \leq 160) = F(170) - F(160) = 1/4$. Nótese que, en el caso de variables continuas, la igualdad en las probabilidades no resulta relevante, ya que la probabilidad *puntual* es nula, como se expondrá más adelante.

---

## 3.2. Función de densidad

Como generalización al caso continuo de la función de masa de probabilidad se tiene la función de densidad. Dada una v.a. continua $X$, la función de densidad se define como:

$$f(x) = \lim_{h \to 0} \frac{\mathbb{P}(x - h < X < x + h)}{2h}.$$

Si $F$ es la función de distribución de $X$, entonces tenemos la siguiente relación entre densidad y distribución:

$$f(x) = F'(x) \quad \text{o} \quad F(x) = \int_{-\infty}^{x} f(t)dt.$$

Para ver de manera intuitiva la dualidad entre función de densidad y función de distribución, consideremos una variable aleatoria $X$ que en un intervalo $(a, b)$ tiene una densidad constante, es decir, cualquier subintervalo dentro de $(a, b)$ de longitud $l$ tiene la misma probabilidad. Más adelante, veremos que si una variable tiene esta función de densidad, se dice que tiene una distribución Uniforme continua en $(a, b)$. Para ser más concretos, fijemos $a = 0$ y $b = 1$. La función de densidad y la función de distribución se pueden ver en la Figura 1:

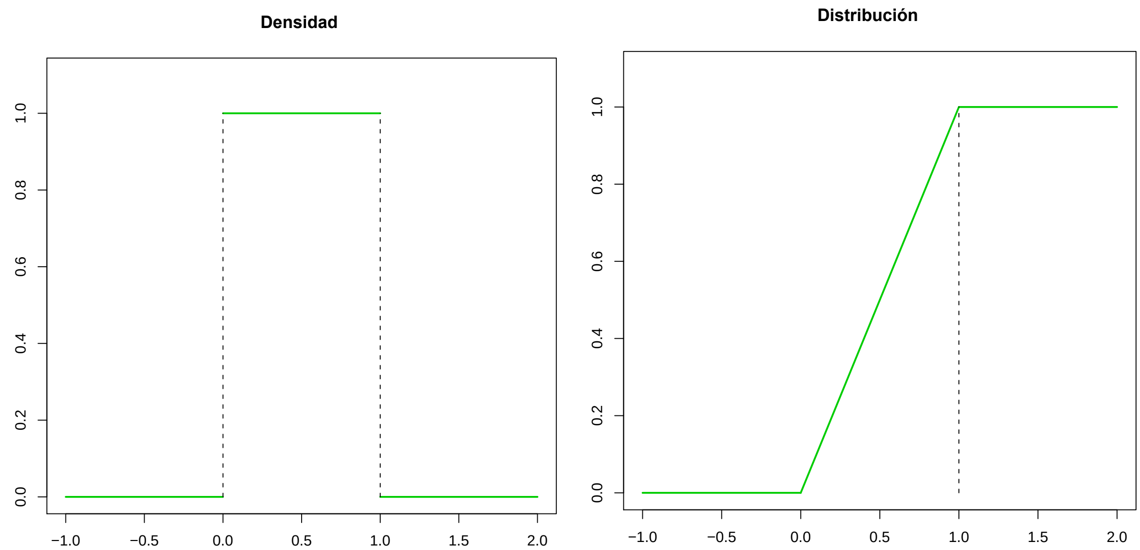


Figura 1. Función de densidad y función de la distribución Uniforme en $(0, 1)$.

La función de distribución en un punto, al darnos la probabilidad acumulada hasta ese valor, se calcula como el área bajo la curva de densidad entre el extremo inferior del soporte de la variable (el valor más pequeño que puede tomar) y ese punto. Entonces, en el ejemplo anterior, el área bajo la densidad hasta $x_0$ se calcula simplemente como el área de un rectángulo.

Podemos ver en la Figura 2 una representación gráfica de las áreas bajo la densidad para $x_0 = 0.25$, $x_0 = 0.5$ y $x_0 = 0.75$. El valor de estas áreas nos da la altura de la función de distribución en esos puntos, como se puede comprobar en la Figura 2.

Para la función de densidad, se pueden deducir las siguientes propiedades:

1. Dado que la función de densidad es la derivada de la función de distribución, y ésta es una función no decreciente: $f(x) \geq 0$, $-\infty < x < \infty$.

2. El área bajo la curva de la función de densidad integra 1, esto es, $\lim_{x\to\infty} \int_{-\infty}^{x} f(t)dt = 1$.

3. La probabilidad de que $X$ tome un valor concreto es nula, esto es, $\mathbb{P}(X = x_0) = \int_{x_0}^{x_0} f(t)dt = 0$.

4. La probabilidad de un intervalo $(a, b)$ es el área bajo la curva $f(x)$ entre las rectas $x = a$ y $x = b$, es decir, $\mathbb{P}(a < X < b) = \int_a^b f(x)dx = F(b) - F(a)$. Por tanto, la probabilidad de $(a, b)$ es la misma que la de $[a, b]$, $[a, b)$ o $(a, b]$.

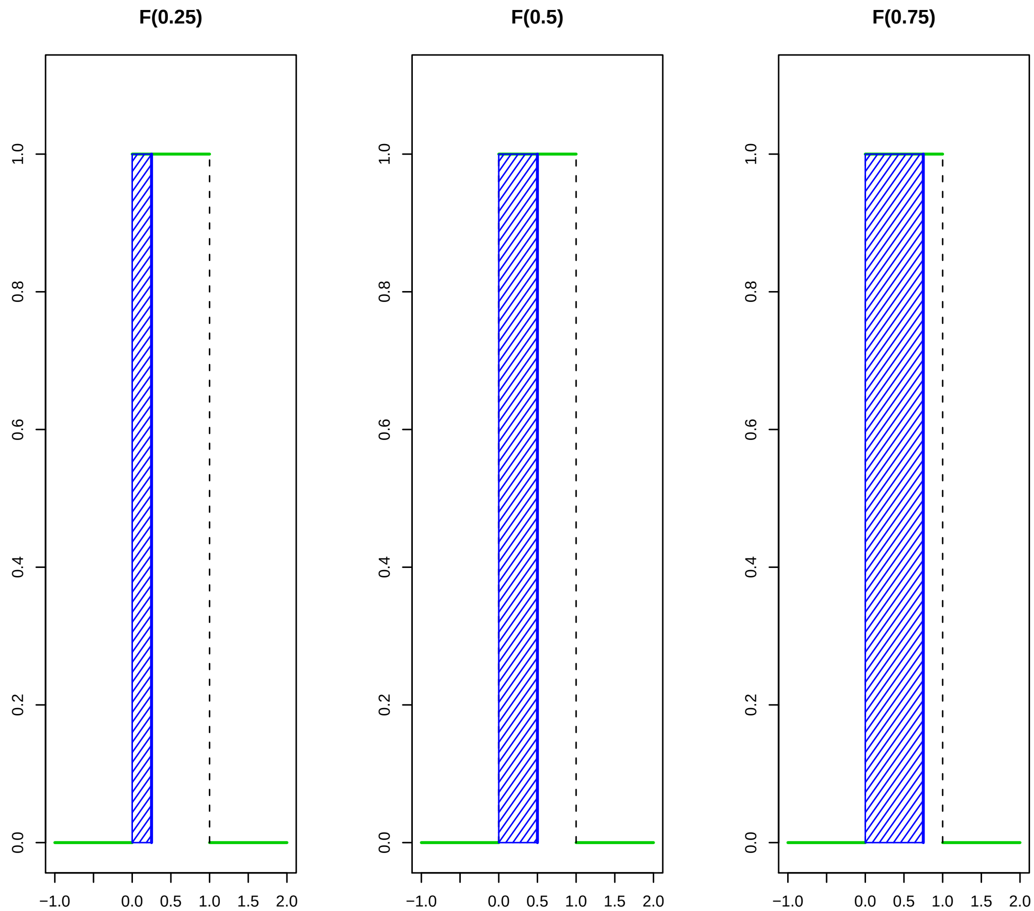


Figura 2. Función de distribución para $x_0 = 0.25$, $x_0 = 0.5$ y $x_0 = 0.75$ representada como área bajo la función de densidad.

La función de densidad extiende el concepto de masa de probabilidad al caso continuo. Si recordamos la definición de masa de probabilidad, las probabilidades de cada valor debían ser no negativas y sumar 1. En este caso, la función de densidad toma valores positivos (propiedad 1) y su integral es 1 (propiedad 2). La principal diferencia radica en las probabilidades puntuales. En el caso de una v.a. continua, la probabilidad de que tome exactamente un valor $x_0$ es nula (propiedad 3), por lo que la probabilidad de tomar valores entre $a$ y $b$ no varía, se incluya o no alguno de los extremos del intervalo.

### 3.3. Medidas características

Si $X$ es una v.a. continua, con función de densidad $f$ se pueden definir la esperanza matemática, la varianza y la desviación típica de modo análogo al caso discreto. En las definiciones que se darán a continuación, se puede observar que la masa de probabilidad es *reemplazada* por la densidad y los sumatorios por integrales.

La media o esperanza matemática viene dada por:

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Otras medidas de posición de tendencia central son la mediana y la moda. La mediana viene dada en términos de la función de distribución, por lo que su definición es la misma que en el caso discreto: $F(Me) = 1/2$. La moda es el punto o puntos donde se maximiza la densidad (se alcanza un máximo relativo). Es decir: $Mo = x_0$ si se verifica que $f'(x_0) = 0$, $f''(x_0) < 0$.

Como medidas de dispersión, la varianza y la desviación típica se definen como:

$$\sigma^2 = \mathbb{V}\text{ar}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx,$$

$$\sigma = +\sqrt{\int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx}.$$

La varianza admite un cálculo más sencillo de la siguiente forma:

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

Las propiedades de la media y la varianza son las mismas que en el caso de variables discretas. La tipificación de variables continuas, que será de gran utilidad en el caso de las variables con distribución Normal como veremos más adelante, se realiza del mismo modo que para variables discretas (restando la media y dividiendo por la desviación típica).

## 4. Independencia de variables aleatorias

**Definición 2.** *Dos v.a. $X$ e $Y$ son independientes si y solo si se verifica:*

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y),$$

*para todo $x, y \in \mathbb{R}$.*

En el caso de v.a. discretas, equivale a obtener la masa de probabilidad como producto de las marginales, mientras que en el caso de v.a. continuas, la condición es la misma pero para las funciones de densidad. $X$ e $Y$ son independientes si, y solo si, en el caso discreto:

$$p_{ij} = \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) = p_i \cdot p_j, \quad i = 1, \ldots, K,\ j = 1, \ldots, L,$$

y en el caso continuo:

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y).$$

En el caso de que las v.a. $X$ e $Y$ sean independientes, entonces se verifican las siguientes propiedades para la media y la varianza:

1. $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.
2. $\mathbb{V}\text{ar}(X + Y) = \mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y)$.

# 5. Modelos de distribuciones discretas

En las siguientes secciones veremos casos particulares de variables aleatorias discretas cuya masa de probabilidad y función de distribución resultan sencillas de obtener y son de utilidad en numerosos problemas aplicados. El primer grupo de variables discretas que estudiaremos serán aquellas relacionadas con el Experimento de Bernoulli. También veremos el modelo de Poisson y la distribución Hipergeométrica. Finalizaremos con la distribución Uniforme discreta, que extenderemos en la siguiente sección al caso continuo.

## 5.1. Experimento de Bernoulli

Un experimento de Bernoulli es aquel que solo presenta dos posibles resultados (por ejemplo, éxito o fracaso, válido o defectuoso, 0 o 1, sano o enfermo, etc). En general, llamaremos éxito (E) a la ocurrencia del suceso que nos interesa estudiar, y fracaso (F) a la no ocurrencia. Además, la probabilidad de obtener un éxito se mantiene constante, y la denotaremos por $p$. La v.a. $X$ que registra el resultado de una prueba de este tipo se dice que tienen una distribución de Bernoulli de parámetro $p$, $X \in Ber(p)$:

$$X = \begin{cases} 1 & \text{si ocurre E,} \\ 0 & \text{si ocurre F,} \end{cases} \quad \text{o también} \quad X = \begin{cases} 1 & \text{con probabilidad } p, \\ 0 & \text{con probabilidad } q = 1 - p. \end{cases}$$

La función de masa de probabilidad de $X \in \text{Ber}(p)$ es:

| $x_i$ | 1 | 0 |
|---|---|---|
| $p_i$ | $p$ | $q = 1 - p$ |

.

En general, podemos escribir la función de masa de probabilidad como:

$$\mathbb{P}(X = x) = p^x(1-p)^{1-x} = p^x q^{1-x}, \quad x \in Sop(X) = \{0, 1\}.$$

La esperanza y la varianza son:

$$\mathbb{E}(X) = p, \quad \mathbb{V}\text{ar}(X) = pq.$$

### 5.2. Distribución Binomial

Si repetimos el experimento de Bernoulli $n$ veces y consideramos la variable que cuenta el número de éxitos esta seguirá una distribución Binomial de parámetros $n$ y $p$, $X \in Bi(n, p)$:

$$X = \text{nž de éxitos en } n \text{ pruebas de Bernoulli} \Rightarrow X \in Bi(n, p)$$

Esta variable, puede tomar valores $Sop(X) = \{0, 1, 2, \ldots, n-1, n\}$. La función de masa de probabilidad es:

$$\mathbb{P}(X = x) = \binom{n}{x} p^x q^{n-x}, \quad \text{donde } \binom{n}{x} = \frac{n!}{x!(n-x)!}, \; x \in Sop(X)$$

y el factorial de un número $n$, denotado por $n!$.

La masa de probabilidad y la función de distribución de una variable aleatoria $Bi(8, 0.4)$ se representan en la Figura 3.

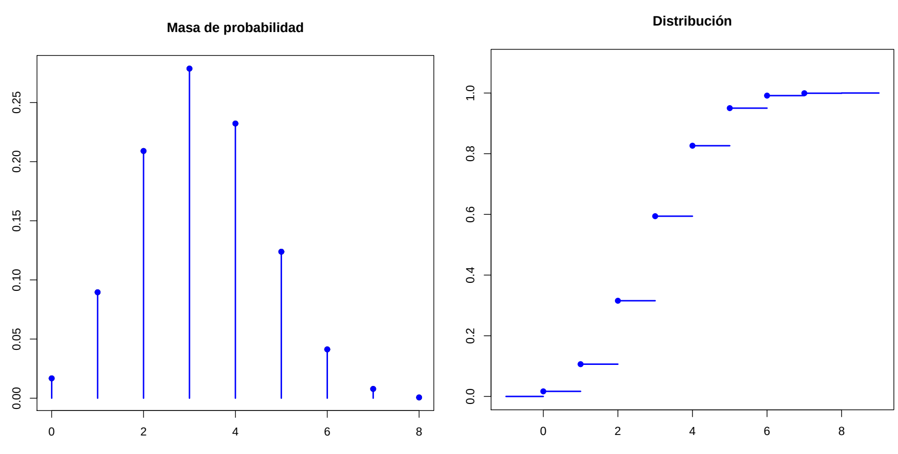


Figura 3. Masa de probabilidad y función de distribución de una variable aleatoria $Bi(8, 0.4)$.

La esperanza y la varianza son:

$$\mathbb{E}(X) = np, \quad \mathbb{V}\text{ar}(X) = npq = np(1-p).$$

Algunas de las propiedades de la distribución Binomial son las siguientes:

1. La distribución de Bernoulli es una Binomial con $n = 1$, esto es, $Ber(p) = Bi(1, p)$.
2. La masa de probabilidad está tabulada, al igual que la función de distribución.
3. Si $X \in Bi(n_1, p)$, $Y \in Bi(n_2, p)$ independientes, entonces $X + Y \in Bi(n_1 + n_2, p)$.
4. Si $X \in Bi(n, p)$ entonces $X = \sum_{i=1}^{n} X_i$, donde $X_i \in Ber(p)$ independientes.

---

**Ejemplo.** En R podemos calcular valores de la masa de probabilidad o de la función de distribución de una variable Binomial (o de una Bernoulli, simplemente considerándola como caso particular con $n = 1$). Sea $X \sim Bi(n, p)$ con $n = 5$ y $p = 1/2$. El soporte de la variable es Sop$(X) = \{0, 1, 2, 3, 4, 5\}$, y los valores de la masa de probabilidad se pueden obtener como:

```
> dbinom(0:5,size=5,prob=0.5)
[1] 0.03125 0.15625 0.31250 0.31250 0.15625 0.03125
```

Del mismo modo, los valores de la función de distribución en cada punto del soporte serán las probabilidades acumuladas hasta dicho punto, es decir:

```
> pbinom(0:5,size=5,prob=0.5)
[1] 0.03125 0.18750 0.50000 0.81250 0.96875 1.00000
```

También podemos calcular, por ejemplo, cuál es la probabilidad de que la variable valga 6 (que será 0) o el valor de la función de distribución en 3.5.

```
> dbinom(6,size=5,prob=0.5);pbinom(3.5,size=5,prob=0.5)
[1] 0
[1] 0.8125
```

Además de las funciones para calcular la masa de probabilidad y la función de distribución, R también nos permite generar de manera aleatoria muestras de una distribución. Por ejemplo, generemos dos muestras, de tamaño 6 y 4, de la variable $X$:

```
> rbinom(6,size=5,prob=0.5)
[1] 3 2 1 2 3 4


> rbinom(4,size=5,prob=0.5)
[1] 2 3 2 3
```

---

### 5.3. Distribución Geométrica

Supongamos que tenemos un experimento de Bernoulli y consideremos la v.a. $X$ que cuenta el número de fracasos hasta que se obtiene el primer éxito:

$$X = \text{nž fracasos hasta el primer éxito} \Rightarrow X \in G(p)$$

Esta variable toma valores $Sop(X) = \{0, 1, 2, \ldots\}$. La función de masa de probabilidad es:

$$\mathbb{P}(X = x) = q^x p = (1-p)^x p, \quad x \in Sop(X).$$

La esperanza y la varianza son:

$$\mathbb{E}(X) = \frac{q}{p}, \quad \mathbb{V}\text{ar}(X) = \frac{q}{p^2}.$$

La masa de probabilidad y función de distribución de una $G(0.3)$ se representan en la Figura 4 en la que se puede ver cómo la masa de probabilidad va decayendo a medida que aumentan los valores en el eje horizontal.

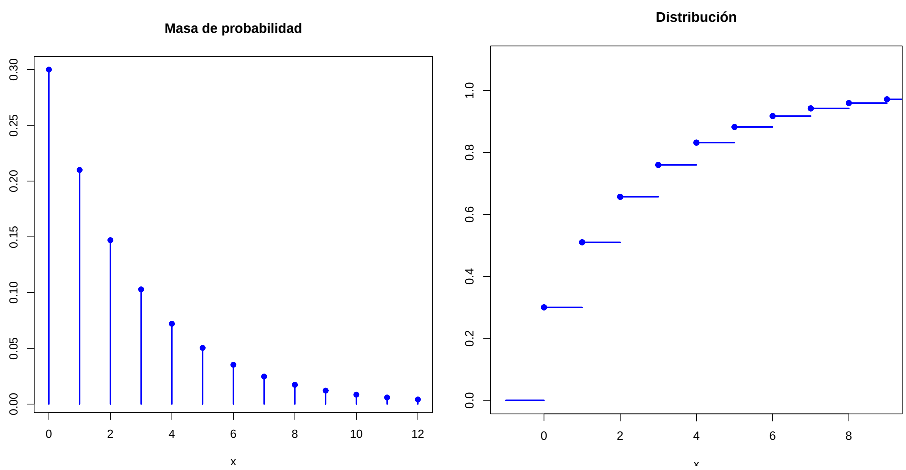


Figura 4. Masa de probabilidad y función de distribución de una $G(0.3)$.

---

**Ejemplo.** Al igual que para variables binomiales, R también dispone de comandos para el cálculo de masa de probabilidad, distribución y generación de muestras de variables geométricas: `dgeom(x, prob)`, `pgeom(q,prob)` y `rgeom(n,prob)`, respectivamente.

---

### 5.4. Distribución Binomial Negativa

La distribución Binomial Negativa generaliza a la distribución Geométrica. Consideremos un experimento de Bernoulli y definamos la variable $X$ que cuenta el número de fracasos hasta obtener el éxito

$n$-ésimo:

$$X = \text{nž de fracasos hasta obtener el éxito } n \Rightarrow X \in BN(n, p).$$

La función de masa de probabilidad es:

$$\mathbb{P}(X = x) = \binom{n + x - 1}{x} q^x p^n, \quad x \in Sop(X) = \{0, 1, 2, \ldots\}$$

La media y la varianza:

$$\mathbb{E}(X) = \frac{nq}{p}, \quad \mathbb{V}\text{ar}(X) = \frac{nq}{p^2}.$$

La masa de probabilidad y la función de distribución de una $BN(5, 0.6)$ pueden verse en la Figura 5. La distribución $G(p)$ es un caso particular de la Binomial Negativa. Si $X \in G(p)$ esto es equivalente a $X \in BN(1, p)$.

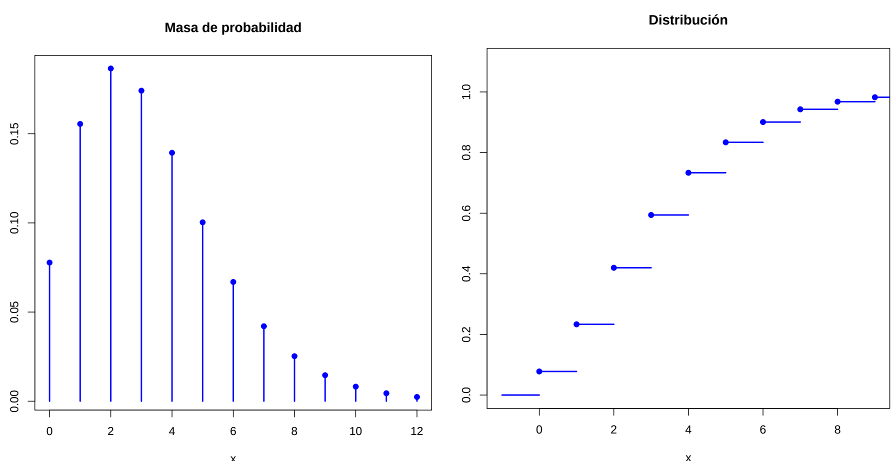


Figura 5. Masa de probabilidad y función de distribución de una $BN(5, 0.6)$.

---

**Ejemplo.** Las funciones de masa de probabilidad, distribución y generación de muestras aleatorias para una variable binominal negativa en R son `dnbinom(x,size,prob)`, `pnbinom(x,size,prob)` y `rnbinom(n,size,prob)` donde `x` sería el valor donde evaluamos la masa de probabilidad o la distribución y `n` sería el tamaño de la muestra a generar. En este caso, `size` es el número de éxitos que se desea obtener y `prob` la probabilidad de éxito. Además, como la distribución geométrica es un caso particular de la distribución binomial negativa, se puede obtener los mismos resultados que en el apartado anterior utilizando estas funciones con `size=1`.

---

### 5.5. Distribución de Poisson

Un proceso de Poisson es un experimento aleatorio que consiste en observar la aparición de sucesos en un soporte continuo, por ejemplo, en el tiempo. Este proceso ha de ser estable: el número medio de sucesos en un intervalo de tiempo fijado, que se denota por $\lambda$, se mantiene constante. Además, los sucesos han de ocurrir de manera independiente.

En este contexto, si consideramos la variable:

$$X = \text{nž de sucesos en un intervalo} \Rightarrow X \in Pois(\lambda).$$

Esta variable toma valores en $Sop(X) = \{0, 1, 2, \dots\}$. La masa de probabilidad es:

$$\mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x \in Sop(X).$$

La media y la varianza son:

$$\mathbb{E}(X) = \mathbb{V}\text{ar}(X) = \lambda.$$

Sean $X_i \in Pois(\lambda_i)$, con $i = 1, \dots, n$ variables independientes, entonces $X = \sum_{i=1}^n X_i \in Pois(\sum_{i=1}^n \lambda_i)$, si los $\lambda_i$ están referidos a la misma unidad temporal.

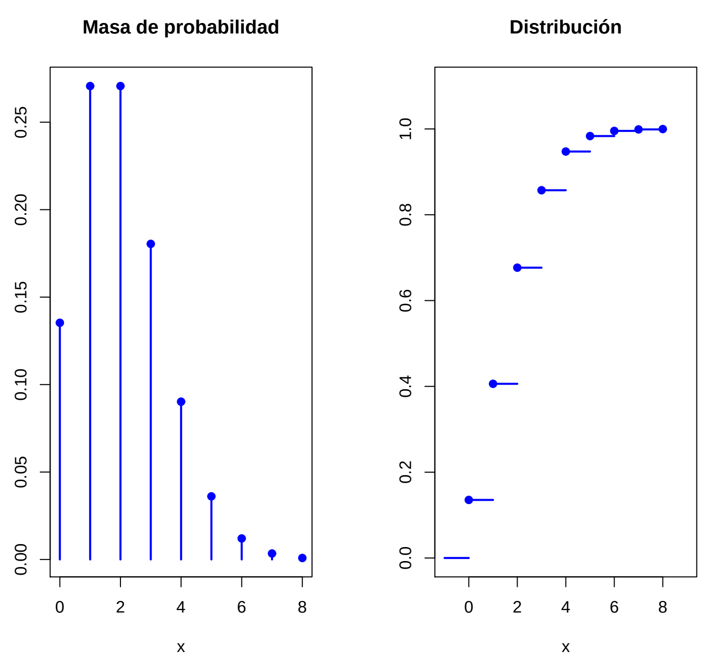


Figura 6. Masa de probabilidad y función de distribución de una variable aleatoria $Pois(2)$.

**Ejemplo.** Las funciones en R para calcular probabilidades sobre variables de Poisson, con número medio de sucesos `lambda`, son `ppois(x,lambda)`, `ppois(x,lambda)` y `rpois(n,lambda)`.

## 5.6. Distribución Hipergeométrica

Supongamos que hay una población finita con $N$ elementos, de los cuales $k$ son de la clase $D$ y $N-k$ son de la clase $\overline{D}$. Tomando una muestra aleatoria sin reemplazamiento de tamaño $n$, la variable:

$$X = \text{nž de elementos de la clase } D \text{ en la muestra de tamaño } n \Rightarrow X \in H(N,n,k).$$

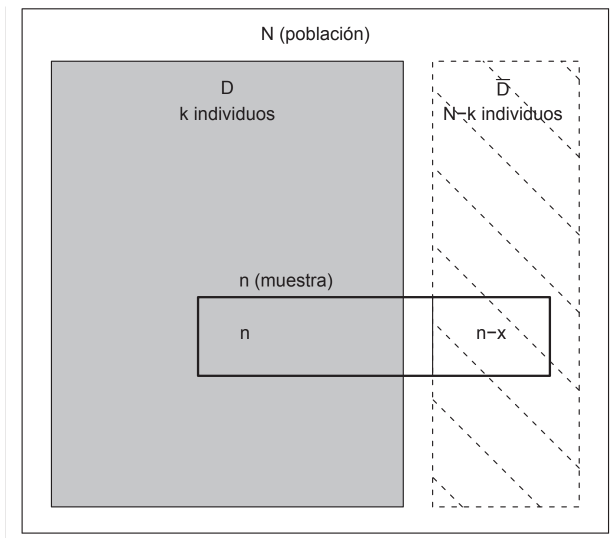


Figura 7. Contexto de la distribución Hipergeométrica.

La masa de probabilidad es:

$$\mathbb{P}(X=x) = \frac{\begin{pmatrix} k \\ x \end{pmatrix}\begin{pmatrix} N-k \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}, \quad x \in Sop(X) = \{\text{máx}(0, n+k-N), \text{mín}(k,N)\}.$$

La media y la varianza son:

$$\mathbb{E}(X) = \frac{nk}{N} = np, \quad \mathbb{V}\text{ar}(X) = npq\frac{N-n}{N-1}.$$

A efectos prácticos, la Hipergeométrica puede aproximarse por una distribución Binomial, cuando el tamaño de la población $N$ es suficientemente grande. Si denotamos por $p = k/N$ (la probabilidad de pertenecer a la clase $D$), podemos observar que el valor esperado de una variable $H(N, n, k)$ coincide con el de una $Bi(n, p)$. La varianza no es la misma, pero se diferencian tan solo en un factor $(N - n)/(N - 1)$ que cuando $N$ es grande y $n$ relativamente pequeño con respecto a $N$, se aproxima a 1.

### 5.7. Distribución Uniforme discreta

Si $X$ toma valores $\{x_1, \ldots, x_K\}$ y todos ellos tienen la misma probabilidad, entonces se dice que $X$ tiene una distribución Uniforme:

$$X \in U\{x_1, \ldots, x_K\}.$$

La masa de probabilidad asigna a todos los valores la misma probabilidad:

$$\mathbb{P}(X = x) = \frac{1}{K}.$$

La media y la varianza son:

$$\mathbb{E}(X) = \frac{1}{K}\sum_{i=1}^{K} x_i, \quad \mathbb{V}\text{ar}(X) = \frac{1}{K}\sum_{i=1}^{K}(x_i - \mu)^2.$$

## 6. Modelos de distribuciones continuas

En esta sección comentaremos los principales modelos de distribuciones continuas. Comenzaremos con la distribución Uniforme continua, como extensión de la Uniforme discreta que hemos visto en el apartado anterior. Veremos también de manera intuitiva cómo se pueden interpretar las funciones de densidad asociadas. Para ello, antes de introducir la distribución Normal (que será la distribución de referencia en esta asignatura), presentaremos algunas gráficas de densidades y discutiremos su significado.

En la Figura 8 se muestran tres ejemplos de densidades continuas. La de la izquierda se corresponde con una densidad $N(0, 1)$, que veremos en el siguiente apartado. A partir de la gráfica podemos interpretar que la zona de valores más probables está en torno al cero (que es la moda) y que se distribuyen simétricamente alrededor de este valor. La gráfica del centro nos muestra una densidad bimodal: es decir, hay dos grupos de valores *más probables*. La gráfica de la derecha nos muestra una densidad asimétrica. Además, la altura de la densidad para valores por debajo de $-3$ es cero, así que el soporte de la variable será $(-3, \infty)$ (o un subconjunto de este).

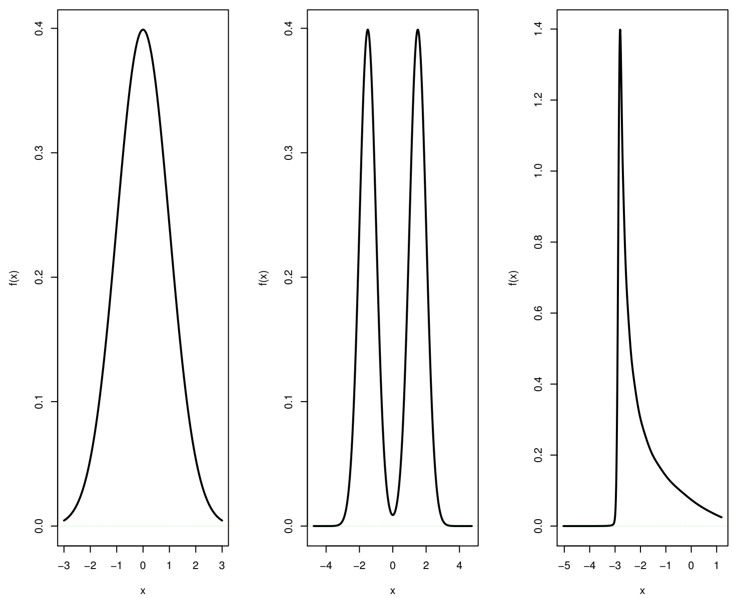

Figura 8. Tres ejemplos de densidades continuas: Normal, bimodal y asimétrica.

### 6.1. Distribución Uniforme continua

Si $X$ toma valores en un intervalo $(a, b)$, diremos que $X$ tiene una **distribución uniforme continua**, $X \in U(a, b)$, si su función de densidad es:

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b), \\ 0 & \text{en otro caso.} \end{cases}$$

La función de distribución asociada:

$$F(x) = \begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & a \leq x < b, \\ 1 & x \geq b. \end{cases}$$

La esperanza y la varianza son:

$$\mathbb{E}(X) = \mu = \frac{a+b}{2} \quad \text{y} \quad \mathbb{V}\text{ar}(X) = \sigma^2 = \frac{(b-a)^2}{12}.$$

El caso de la distribución Uniforme continua es muy simple, ya que su función de densidad es un segmento plano sobre el soporte de la variable y con altura el inverso de la longitud del soporte (para garantizar que el área bajo la densidad integra $1$). En general, la altura que alcanza una función de densidad nos da información sobre los rangos de valores *más probables* (o menos probables).

### 6.2. Distribución Normal

Una variable aleatoria $X$ tiene **distribución Normal** con media $\mu$ y varianza $\sigma^2$, $X \in N(\mu, \sigma^2)$, si su función de densidad es:

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

Si tenemos $\mu = 0$ y $\sigma^2 = 1$, entonces $X \in N(0,1)$ y su densidad será:

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

Cambiar los valores de los parámetros de localización (media) y escala (desviación típica), tiene distinto efecto sobre la forma de la función de densidad y por tanto, de la función de distribución, como podemos ver en la Figura 9. Tomando como referencia la distribución Normal de media $\mu = 0$ y varianza $\sigma^2 = 1$, si cambiamos la media, lo que hacemos es trasladar la gráfica, hacia la derecha si la media es positiva, y hacia la izquierda si es negativa. Cuando modificamos la varianza, si la aumentamos lo que estamos haciendo es incrementar la dispersión, con lo que la curva se *achata*, incrementando la probabilidad de valores más altos y más bajos con respecto al modelo estándar. Si reducimos la varianza, lo que veríamos es que se concentra más alrededor de la media. Al incremetar la varianza, la densidad se vuelve mesocúrtica (curtosis negativa, más achatada), mientras que al disminuir la varianza, lo que se obtiene es una curva leptocúrtica (curtosis positiva, más apuntada). La densidad de una $N(0,1)$ tiene curtosis nula (platocúrtica).

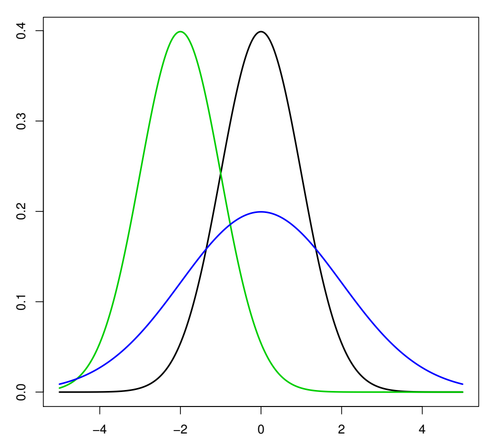


Figura 9. Funciones de densidad y funciones de distribución para diferentes variables que siguen una distribución Normal. $N(0,1)$ (negro), $N(-2,1)$ (verde) y $N(0,4)$ (azul).

---

**Ejemplo.** Para dibujar las densidades de la Figura 9 podemos utilizar los siguientes comandos de R:

```
> curve(dnorm,xlim=c(-5,5),lwd=2)
> curve(dnorm(x,mean=-2),add=TRUE,col=3,lwd=2)
> curve(dnorm(x,sd=2),add=TRUE,col=4,lwd=2)
```

La función `dnorm` permite calcular la función de densidad de una Normal. Los parámetros (de media y desviación típica) por defecto son los de la Normal estándar. Para modificarlos, se deben incluir como argumentos `mean` y `sd` (desviación típica, no varianza aunque esta sea la notación habitual en esta asignatura).

---

La función de distribución de la Normal estándar se denota por $\Phi(z)$, aunque no existe una expresión explícita (fórmula) para ella. En la distribución $N(0,1)$ será de utilidad identificar en qué intervalos se encuentran el $90\,\%$, $95\,\%$ y $99\,\%$ de los valores. En la Figura 10 se muestran los tres intervalos más usuales para los posibles valores de una $N(0,1)$. Estos intervalos serán de utilidad tanto en la estimación por intervalos de confianza como para los contrastes de hipótesis.

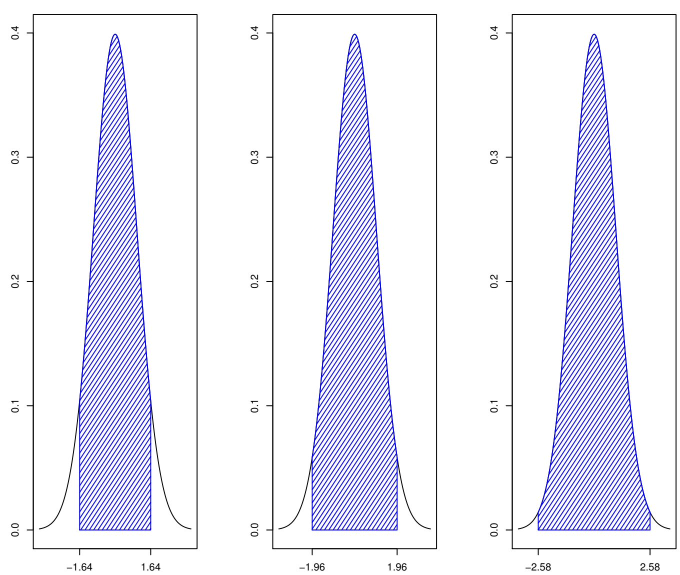


Figura 10. Intervalos de probabilidad en una $N(0,1)$. El $90\,\%$ de los valores están en $(-1.64, 1.64)$. El $95\,\%$ de los valores están en $(-1.96, 1.96)$. El $99\,\%$ de los valores están en $(-2.58, 2.58)$.

Todas las variables Normales se pueden transformar, mediante tipificación o estandarización, a una Normal estándar. Esto es:

$$X \in N(\mu, \sigma^2) \Leftrightarrow Z = \frac{X-\mu}{\sigma} \in N(0,1).$$

En algunos casos lo que nos interesa es, dada una probabilidad, ver cuál es el punto de corte que deja esa proporción de valores de la variable por encima o por debajo. Para ello, se define la función de distribución inversa o la **función cuantil** del siguiente modo. Dada una probabilidad $p_0$, la función cuantil $q$ nos devuelve el punto $x_0$ tal que:

$$q(p_0) = x_0, \quad \text{si} \quad \mathbb{P}(X \leq x_0) = F(x_0) = p_0.$$

En la Figura 11 representamos la idea de la función cuantil para una distribución Normal.

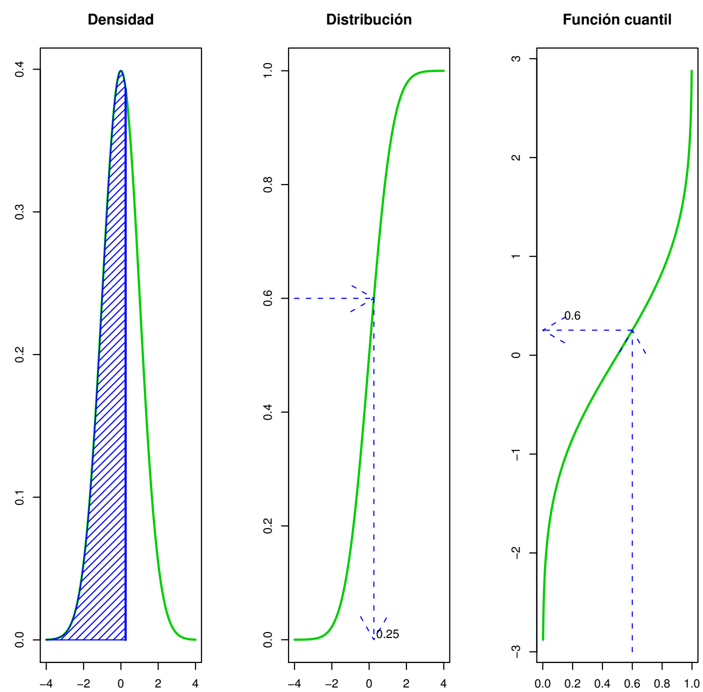


Figura 11. Función de densidad, función de distribución y función cuantil (distribución inversa) de una $N(0,1)$. Ejemplo para el sexto decil (cuantil $0.6$). $\mathbb{P}(Z \leq z_0) = 0.6 \Rightarrow z_0 = 0.2533$.

Por otra parte, la distribución Normal también satisface la siguiente propiedad de aditividad; si $X \in N(\mu_1, \sigma_1^2)$ e $Y \in N(\mu_2, \sigma_2^2)$ son independientes, entonces:

$$(X+Y) \in N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

En general, la suma de variables Normales también sigue distribución Normal, donde la media es la suma de las medias, pero en la varianza se debe tener en cuenta la posible dependencia entre las variables (covarianza), que vale cero cuando son independientes.

### 6.3. Distribución Exponencial

En la sección anterior definíamos el proceso de Poisson como aquel en que los sucesos aparecen de manera independiente y con una frecuencia de aparición constante, que llamábamos $\lambda$. Habitualmente, estudiaremos los procesos de Poisson en el tiempo; una variable que contabiliza el número de sucesos en un intervalo de tiempo fijado tiene distribución de Poisson. En este tipo de procesos, además de la aparición de los sucesos, nos puede interesar el tiempo entre ellos.

Si tenemos un proceso de Poisson de parámetro $\lambda$ y consideramos la variable $X$ que mide el *tiempo entre dos sucesos consecutivos*, entonces diremos que la variable $X$ sigue una distribución Exponencial de parámetro $\lambda$ y lo denotaremos por $X \in Exp(\lambda)$. Como es una variable que mide el tiempo entre sucesos, su soporte es $Sop(X) = (0, +\infty)$. Su función de densidad es:

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

La función de distribución se obtiene mediante integración por partes de la densidad:

$$F(x) = 1 - e^{-\lambda x}.$$

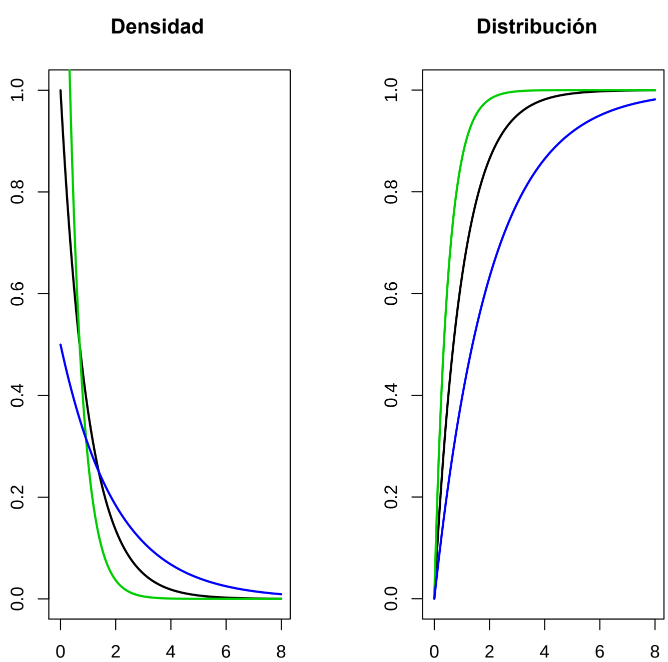


Figura 12. Funciones de densidad y funciones de distribución para variables que siguen una distribución exponencial. Negro: $\lambda = 1$. Verde: $\lambda = 2$. Azul: $\lambda = 0.5$.

En la Figura 12 tenemos las funciones de densidad y funciones de distribución de variables Exponenciales con distintos parámetros. Las curvas en negro se corresponden con la Exponencial estándar, es decir, $\lambda = 1$ (frecuencia media de aparición de un suceso por unidad de tiempo). La curva verde ($\lambda = 2$) tiene una frecuencia de aparición mayor. Así, el tiempo de espera entre dos sucesos consecutivos se reduce, ya que los sucesos se producen con mayor frecuencia. La curva verde decae más rápidamente que la negra. Por otro lado, la curva azul tiene una frecuencia de aparición inferior ($\lambda = 0.5$), lo que retrasa la aparición de sucesos, haciendo más probables tiempos de espera mayores.

En las funciones de distibución (Figura 12, gráfico derecho) observamos que cuando la frecuencia media de aparición aumenta, la curva de distribución crece más rápidamente, mientras que si la frecuencia disminuye, el crecimiento es menor.

---

**Ejemplo.** La Figura 12 con las densidades y distribuciones exponenciales se puede reproducir con el siguiente código de R:

```
> par(mfrow=c(1,2))
> curve(dexp,lwd=2,ylim=c(0,1),from=0,to=6)
> curve(dexp(x,rate=2),col=3,lwd=2,add=TRUE)
> curve(dexp(x,rate=0.5),col=4,lwd=2,add=TRUE)
> curve(pexp,lwd=2,ylim=c(0,1),from=0,to=6)
> curve(pexp(x,rate=2),col=3,lwd=2,add=TRUE)
> curve(pexp(x,rate=0.5),col=4,lwd=2,add=TRUE)
```

---

La esperanza de esta variable, el tiempo de espera medio entre sucesos, es la inversa de la frecuencia media de aparición. La esperanza y la varianza, que también se calculan utilizando integrales por partes, son:

$$\mathbb{E}(X) = \frac{1}{\lambda} \quad \text{y} \quad \mathbb{V}\text{ar}(X) = \frac{1}{\lambda^2}.$$

### 6.4. Distribución Gamma

La generalización de la distribución Exponencial (tiempo entre dos sucesos consecutivos) se obtiene cuando consideramos el tiempo hasta la aparición del $n$-ésimo suceso. Si en un proceso de Poisson con frecuencia media de aparición $\lambda$ consideramos la variable $X$ que recoja el *tiempo hasta la aparición del $n$-ésimo suceso*, $Sop(X) = (0, +\infty)$, entonces $X$ tiene una **distribución Gamma** de parámetros $n$ y $\lambda$, $X \in \Gamma(n, \lambda)$. La función de densidad de $X \in \Gamma(n, \lambda)$ es:

$$f(x) = \begin{cases} \frac{\lambda^n}{\Gamma(n)} e^{-\lambda x} x^{n-1} & x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

La distribución Exponencial es un caso particular de la Gamma, ya que $Exp(\lambda)$ es una $\Gamma(\lambda, 1)$ y la esperanza y la varianza son una generalización de las características de la Exponencial:

$$\mathbb{E}(X) = \frac{n}{\lambda} \quad \text{y} \quad \mathbb{V}\text{ar}(X) = \frac{n}{\lambda^2}.$$

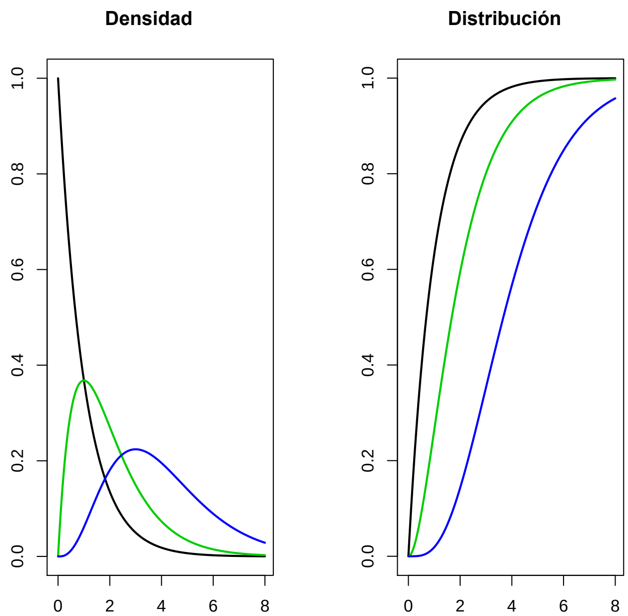


Figura 13. Funciones de densidad y funciones para diferentes variables que siguen una distribución Gamma. $\Gamma(1,1) = Exp(1)$ (negro), $\Gamma(2,1)$ (verde) y $\Gamma(4,1)$ (azul).

La distribución Gamma también verifica una propiedad de aditividad: si tenemos dos variables $X \in \Gamma(\lambda, n_1)$, $Y \in \Gamma(\lambda, n_2)$ independientes, entonces $X + Y \in \Gamma(\lambda, n_1 + n_2)$. En la Figura 13 tenemos las funciones de densidad y las funciones de distribución para tres variables Gamma, consideradas sobre un proceso de Poisson con frecuencia media de aparición $\lambda = 1$, pero con distinto número de sucesos.

---

**Ejemplo.** Las funciones en R para el cálculo de densidad y distribución son `dgamma` y `pgamma`.

---

## 7. Teorema Central del Límite

Uno de los resultados fundamentales de la Estadística es el Teorema Central del Límite. Este teorema establece que el promedio de variables independientes e idénticamente distribuídas (sea cual sea esa distribución), con varianza finita, sigue una distribución Normal.

**Teorema.** *Sea $\{X_i\}_{i\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con $\mathbb{E}(X_i) = \mu$, $\mathbb{V}ar(X_i) = \sigma^2$, para todo $i \in \mathbb{N}$. Entonces, para $n$ suficientemente grande:*

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2) \Longleftrightarrow \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \sim N(0,1),$$

*donde $\sim$ indica una distribución aproximada.*

Por ejemplo, si $X \in Bi(n,p)$ se puede escribir como:

$$X = \sum_{i=1}^{n} X_i, \quad \text{donde cada variable } X_i \in Ber(p).$$

Por tanto, $\mathbb{E}(X_i) = p$, $\mathbb{V}\text{ar}(X_i) = p(1-p) = pq$. Así, aplicando el Teorema Central del Límite, para $n$ suficientemente grande y $0.1 < p < 0.9$ se tiene:

$$X = \sum_{i=1}^{n} X_i \sim\ N(np, npq).$$

Otro ejemplo lo tenemos con la distribución Gamma que se puede expresar como suma de Exponenciales. Así, $X \in \Gamma(n,\lambda)$ se tiene que:

$$X = \sum_{i=1}^{n} X_i \quad \text{donde cada } X_i \in Exp(\lambda).$$

Entonces, $\mathbb{E}(X_i) = 1/\lambda$ y $\mathbb{V}\text{ar}(X_i) = 1/\lambda^2$, y aplicando el Teorema Central del Límite, para $n$ suficientemente grande, se tiene que $X \sim N(n/\lambda, n/\lambda^2)$, donde recordemos $n/\lambda$ y $n/\lambda^2$ son los valores de media y varianza de una $\Gamma(n,\lambda)$.

## 7.1. Aproximaciones entre distribuciones

Cuando el número de pruebas del experimento de Bernoulli $n$ es grande, podemos hacer aproximaciones de la distribución Binomial que facilitarán el cálculo de probabilidades, bien a la distribución de Poisson o bien a la Normal.

En la Figura 14 podemos ver gráficamente la validez de la aproximación de la Binomial a la Normal. En la parte izquierda tenemos la masa de probabilidad de una $Bi(50, 0.3)$. Como el número de pruebas $n$ es suficientemente grande (se suele fijar el criterio $n > 30$), podríamos aproximar la Binomial por una Normal $N(\mu, \sigma^2)$ con $\mu = np = 50 \cdot 0.3 = 15$ y $\sigma^2 = npq = 50 \cdot 0.3 \cdot 0.7 = 10.5$.

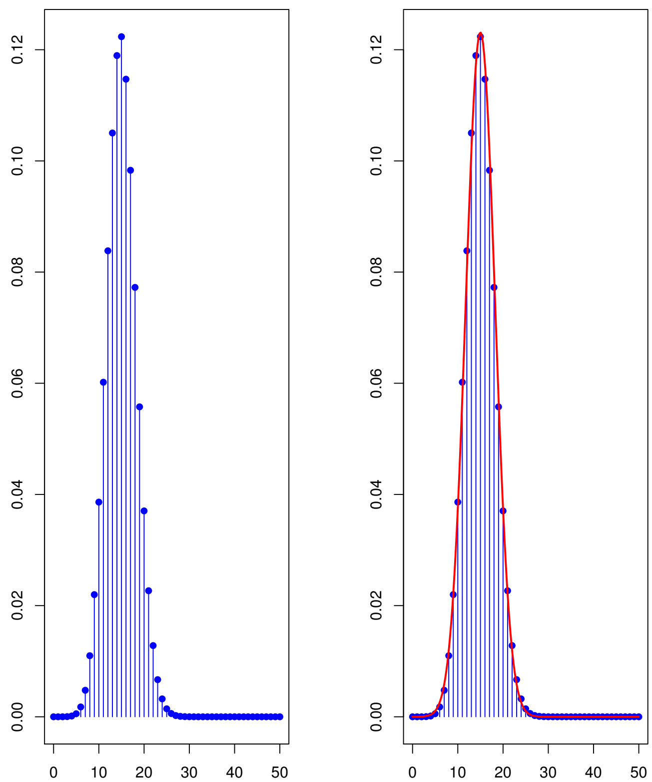


Figura 14. Masa de probabilidad de una $Bi(50, 0'3)$ y función de densidad de una $N(15, 10.5)$.

Cuando en una distribución Binomial con $n$ grande, la probabilidad de éxito es extrema ($p < 0.1$ o $p > 0.9$), se aproxima mejor por una Poisson, con la misma media, es decir, $Pois(np)$.

En el caso de la distribución de Poisson, cuando el parámetro $\lambda$ es grande (normalmente se toma $\lambda > 10$), se puede aproximar por una Normal que conserva la media y la varianza. Es decir, una variable $Pois(\lambda)$ se aproximaría por una $N(\lambda, \lambda)$, para $\lambda$ suficientemente grande.

Al aproximar la distribución Binomial o la Poisson por la Normal estamos conservando la media y la varianza. Sin embargo, al calcular probabilidades utilizando la aproximación a la Normal debemos tener en cuenta que tanto la Binomial como la Poisson son discretas, mientras que la Normal es continua. Para una variable $X \in Bi(50, 0.3)$ podemos calcular la probabilidad de que $X$ sea $20$, $\mathbb{P}(X = 20)$ y podemos ver en la Figura 14 que es positiva. Sin embargo, utilizando la aproximación $X \in N(15, 75)$, al ser continua, la probabilidad $\mathbb{P}(X = 20)$ será nula. Para solucionar este problema, utilizaremos la **corrección de Yates**. La probabilidad de que $X$ sea igual a $20$ es la misma que:

$$\mathbb{P}(X = 20) = \mathbb{P}(19.5 < X < 20.5),$$

y sobre esta segunda expresión podemos utilizar la aproximación Normal, obteniendo un resultado no nulo. Esta corrección también se debe emplear al aproximar la Poisson por la Normal, y en general cualquier aproximación que se haga de una distribución discreta a través de una continua.

| **Distribución** | **Caso** | **Aproximación** |
|---|---|---|
| $Bi(n, p)$ | $n \geq 30$, $p <$0.1* | $Pois(np)$ |
| $Bi(n, p)$ | $n \geq 30$, 0.1$<$ $p$ $<$0.9 | $N(np, npq)$ |
| $Pois(\lambda)$ | $\lambda \geq 10$ | $N(\lambda, \lambda)$ |

Tabla 1. Aproximaciones de la Binomial, Poisson y Normal.

* Si $n \geq 30$, $p > 0.9$, entonces se usa que

$$\mathbb{P}(Bi(n, p) = k) = \mathbb{P}(Bin(n, 1-p) = n-k) \simeq \mathbb{P}(Pois(n(1-p)) = n-k).$$

| **Estadística Descriptiva** | **Variable aleatoria discreta** | **Variable aleatoria continua** |
|---|---|---|
| Muestra<br>$(x_1, \ldots, x_n)$ | $X$ Variable aleatoria discreta<br>$Sop(X) = \{x_1, \ldots, x_K\}$ | $X$ Variable aleatoria continua<br>$Sop(X) = (a, b) \subseteq \mathbb{R}$ |
| $f_i = \frac{n_i}{n}$ frecuencia relativa<br>$f_i \geq 0, \sum_{i=1}^{n} f_i = 1$ | $p_1, \ldots, p_K$ masa de probabilidad<br>$p_i \geq 0, \sum_{i=1}^{K} p_i = 1$ | $f(x)$ función de densidad<br>$f(x) \geq 0, \int_{-\infty}^{\infty} f(x)dx = 1$ |
| $F_i = \frac{N_i}{n}$ frec. relativa acumulada<br>$F_K = 1$ ($K$ clases) | $F(x) = \mathbb{P}(X \leq x)$ distribución<br>$F(-\infty) = 0, \quad F(+\infty) = 1$ | $F(x) = \mathbb{P}(X \leq x)$ distribución<br>$F(-\infty) = 0, \quad F(+\infty) = 1$ |
| $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ | $\mu = \mathbb{E}(X) = \sum_{i=1}^{K} x_i p_i$ | $\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)dx$ |
| $s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$<br>$s^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2$ | $\sigma^2 = \mathbb{V}\text{ar}(X) = \sum_{i=1}^{K} (x_i - \mu)^2 p_i$<br>$\sigma^2 = \sum_{i=1}^{K} x_i^2 p_i - \mu^2$ | $\sigma^2 = \mathbb{V}\text{ar}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$<br>$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$ |

Tabla 2. Relaciones ente Estadística Descriptiva, variable aleatoria discreta y variable aleatoria continua.