

Capítulo 1

Conceptos básicos de probabilidad

1.1. Espacio de probabilidad

En lo que sigue daremos una noción intuitiva del concepto de *espacio de probabilidad* y los elementos que intervienen en su definición.

1.1.1. Espacio muestral

Dado un experimento, se llama **espacio muestral** al conjunto de resultados del experimento.

Ejemplo 1.1. En una carrera de 3 caballos a , b y c , se considera el orden de llegada a la meta y se representa el resultado con una terna. Así, la terna (b, c, a) indica que el caballo b llegó primero, c llegó segundo y a llegó tercero. Asumimos que no hay posibilidad de empate.

El espacio muestral de este experimento es el conjunto S formado por todos los resultados posibles:

$$S = \{(a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), (c, b, a)\}.$$

Ejemplo 1.2. Si arrojamus una moneda dos veces, y denotamos por C si sale cara y por X si sale cruz, entonces, el espacio muestral formado por los resultados del experimento puede representarse por el conjunto

$$S = \{(C, C), (C, X), (X, C), (X, X)\}.$$

Ejemplo 1.3. En una urna hay tres bolas verdes y dos rojas. Se sacan dos bolas *simultáneamente*. El espacio muestral está formado por todos los conjuntos posibles de dos bolas. Si denotamos V_1, V_2 y V_3 a las bolas verdes y R_1, R_2 a las rojas, entonces el espacio muestral tiene $\binom{5}{2} = 10$ elementos:

$$S = \{\{V_1, V_2\}, \{V_1, V_3\}, \{V_1, R_1\}, \{V_1, R_2\}, \{V_2, V_3\}, \{V_2, R_1\}, \{V_2, R_2\}, \\ \{V_3, R_1\}, \{V_3, R_2\}, \{R_1, R_2\}\}.$$

Si en cambio el experimento consiste en sacar dos bolas *consecutivamente*, sin reposición, el espacio muestral consiste en 20 elementos, ya que deben considerarse todos los pares ordenados. Esto es, deben distinguirse los resultados (V_1, V_2) de (V_2, V_1) y así sucesivamente.

Cualquier subconjunto del espacio muestral es un **evento**. En el Ejemplo 1.1, el subconjunto

$$U = \{(b, c, a), (c, b, a)\}$$

es un evento, que se puede describir como el conjunto de los resultados en los que el caballo a sale último en la carrera.

Si A y B son eventos de un espacio muestral S , entonces también lo son su unión, su intersección y el complemento, que se denotan $A \cup B$, AB y A^c , respectivamente.

En particular, si A_1, A_2, \dots, A_n son eventos, también lo son:

- La unión de los eventos: $A_1 \cup A_2 \cup \dots \cup A_n$.
- La intersección de los eventos: $A_1 A_2 \dots A_n$.
- El espacio muestral: $S = A \cup A^c$.
- El conjunto vacío: $\emptyset = AA^c$.

Dos eventos A y B se dicen **mutuamente excluyentes** si $AB = \emptyset$. Los eventos A_1, A_2, \dots, A_n se dicen **mutuamente excluyentes dos a dos** si $A_i A_j = \emptyset$ para cada $i \neq j$.

1.1.2. Axiomas de probabilidad

Consideramos un espacio muestral S , y suponemos que existe una función P definida sobre el conjunto de eventos de S que satisface los siguientes axiomas:

Axioma 1: $0 \leq P(A) \leq 1$, para todo evento A .

Axioma 2: $P(S) = 1$.

Axioma 3: Si $A_1, A_2, \dots, A_n, \dots$, son mutuamente excluyentes dos a dos, entonces

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i), \quad n = 1, 2, \dots$$

Una tal función P se la denomina **probabilidad**, y $P(A)$ es la probabilidad del evento A . El Axioma 1 indica que la probabilidad es un número real entre 0 y 1. El Axioma 2 indica que la probabilidad de que ocurra algún resultado de S es 1. El Axioma 3 indica que si dos o más eventos son mutuamente excluyentes, entonces la probabilidad de que ocurra alguno de ellos es la suma de sus probabilidades.

De los axiomas podemos concluir además que:

- $P(A^c) = 1 - P(A)$, para todo evento A .
- $P(\emptyset) = 0$.
- $P(A \cup B) = P(A) + P(B) - P(AB)$, para cualquier par de eventos A y B .

Un **espacio de probabilidad** es un par (S, P) donde S es un espacio muestral y P es una probabilidad sobre S .

En teoría de Probabilidad, la familia de subconjuntos en la que está definida la función P puede no ser toda la familia de eventos, sino algunos de ellos, siempre que éstos constituyan una sigma-álgebra de subconjuntos Σ . En esta introducción asumiremos que Σ está conformado por todos los subconjuntos de S .

Ejemplo 1.4. Si S son los posibles resultados de la carrera de tres caballos, y se asigna a cada resultado la misma probabilidad:

$$P(\{(a, b, c)\}) = P(\{(b, a, c)\}) = \dots = \frac{1}{6}$$

entonces (S, P) es un espacio de probabilidad.

En particular, la probabilidad que el caballo a salga último es la probabilidad del evento $U = \{(b, c, a), (c, b, a)\}$, es decir:

$$P(U) = P(\{(b, c, a)\} \cup \{(c, b, a)\}) = P(\{(b, c, a)\}) + P(\{(c, b, a)\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

1.1.3. Probabilidad condicional

Retomando el Ejemplo 1.1, supongamos que el caballo a ha ganado la carrera. Con esta información, ¿cuál es la probabilidad de que el caballo c haya salido último?

El evento que el caballo a haya salido primero está dado por el siguiente subconjunto de resultados:

$$F = \{(a, b, c), (a, c, b)\}.$$

Asumiendo que todos los resultados son igualmente probables, se podría decir que *dado que el caballo a salió primero, la probabilidad que c salga último es $\frac{1}{2}$* .

Dados dos eventos A y F , y una probabilidad P , la **probabilidad condicional** de que ocurra A dado F se define como:

$$P(A | F) = \frac{P(AF)}{P(F)}$$

siempre que $P(F) > 0$.

Ejemplo 1.5. De una urna con 4 bolas rojas y 7 bolas azules se extraen dos bolas de manera consecutiva, sin reposición. ¿Cuál es la probabilidad que la segunda bola sea azul si la primera también lo es?

Para el cálculo de estas probabilidades es útil pensar a la urna como un vector $v = (v_1, \dots, v_{11})$, donde cada v_1, v_2, v_3 y v_4 representan las bolas rojas y las siete restantes una azul. El espacio muestral S consiste en todas las extracciones de dos bolas, por lo cual cada elemento de S puede representarse como un par ordenado (v_i, v_j) , y la probabilidad de cada uno de estos resultados puntuales es:

$$P(\{(v_i, v_j)\}) = \frac{1}{11 \cdot 10}.$$

Debemos calcular una probabilidad condicional $P(A | F)$, donde F es el evento donde la primera bola es azul y A es el evento que la segunda sea azul. Así:

$$P(A | F) = \frac{P(AF)}{P(F)}$$

El evento AF consiste en que ambas bolas sean azules, por lo cual

$$P(AF) = \frac{7 \cdot 6}{11 \cdot 10}.$$

Luego

$$P(A | F) = \frac{7 \cdot 6}{7 \cdot 10} = \frac{6}{10}.$$

Intuitivamente es un resultado razonable, ya que habiendo sacado una bola azul quedan 6 azules sobre un total de 10 bolas.

Dos eventos A y B se dicen **independientes** si $P(A | B) = P(A)$. En este caso se cumple:

$$P(AB) = P(A) \cdot P(B).$$

Notemos que para cualquier par de eventos A y F , se tiene que $A = AF \cup AF^c$. Dado que F y F^c son mutuamente excluyentes, entonces también lo son AF y AF^c . Por lo tanto

$$P(A) = P(AF) + P(AF^c).$$

Esto nos permite calcular la probabilidad de A como una suma ponderada de probabilidades condicionales:

$$\begin{aligned} P(A) &= P(A | F)P(F) + P(A | F^c)P(F^c) \\ &= P(A | F)P(F) + P(A | F^c)(1 - P(F)). \end{aligned} \quad (1.1)$$

En particular, también tenemos que $P(AF) = P(A | F)P(F)$ y $P(AF) = P(F | A)P(A)$. Entonces:

$$P(F | A) = \frac{P(AF)}{P(A)} = \frac{P(A | F)P(F)}{P(A)}. \quad (1.2)$$

La igualdad (1.2) se conoce como **Fórmula de Bayes**, y permite calcular la probabilidad condicional $P(F | A)$ (a posteriori) en términos de $P(F)$ (a priori). Puede generalizarse a un número finito de eventos. Esto es si F_1, F_2, \dots, F_n son eventos mutuamente excluyentes tales que $F_1 \cup F_2 \cup \dots \cup F_n = S$, entonces para cualquier evento A se tiene que:

$$\begin{aligned} P(A) &= P(AF_1) + P(AF_2) + \dots + P(AF_n) \\ &= P(A | F_1)P(F_1) + P(A | F_2)P(F_2) + \dots + P(A | F_n)P(F_n) \\ &= \sum_{i=1}^n P(A | F_i)P(F_i). \end{aligned}$$

Luego, la probabilidad condicional de que ocurra el evento F_j dado A es igual a:

$$P(F_j | A) = \frac{P(A | F_j)P(F_j)}{P(A)} = \frac{P(A | F_j)P(F_j)}{\sum_{i=1}^n P(A | F_i)P(F_i)}. \quad (1.3)$$

1.2. Variables aleatorias

Dado un espacio muestral (S, P) , una **variable aleatoria** es una función $X : S \mapsto \mathbb{R}$ tal que el conjunto $\{s \in S | X(s) \leq x\}$ es un evento sobre el que está definido P , para todo $x \in \mathbb{R}$.

Ejemplo 1.6. Si el experimento consiste en arrojar dos dados, la suma de los valores obtenidos es una variable aleatoria que toma los valores enteros entre 2 y 12.

Ejemplo 1.7. Si el experimento consiste en arrojar una moneda sucesivamente hasta que caiga cara, el número de veces que salió cruz es una variable aleatoria. Específicamente, el espacio muestral determinado por el experimento puede describirse como:

$$S = \{C, XC, XXC, XXXC, \dots\}$$

y la variable aleatoria toma valores en los enteros no negativos: $X : S \mapsto \mathbb{N} \cup 0$.

Ejemplo 1.8. También son ejemplos de variables aleatorias las que resultan de:

- contar el número de personas que ingresan a un local determinado cada día,
- medir el tiempo de servicio en un cajero automático,
- observar las tasas de interés en el mercado financiero.

Si X es una variable aleatoria sobre un espacio de probabilidad (S, P) , y x es un número real, entonces denotamos con $\{X \leq x\}$ al subconjunto de S :

$$\{X \leq x\} := \{s \in S \mid X(s) \leq x\}.$$

Una notación análoga denotará a los eventos $\{X = x\}$, $\{X > x\}$, $\{a < X < b\}$, y así siguiendo.

Dada una variable aleatoria X en un espacio de probabilidad S , se define la **función de distribución acumulada** como:

$$F(x) = P(X \leq x) := P(\{s \in S \mid X(s) \leq x\}). \quad (1.4)$$

La función F tiene dominio en los números reales y toma valores en el intervalo $[0, 1]$. En particular cumple las siguientes propiedades:

- F es no decreciente,
- para todo $x \in \mathbb{R}$ se cumple que $0 \leq F(x) \leq 1$,
- F es continua a derecha.

Diremos que una variable aleatoria es **discreta** si toma sólo un número finito o numerable de valores. En este caso, se define la **función de probabilidad de masa** por

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

Si la variable X toma valores en un conjunto finito $\{x_1, x_2, \dots, x_N\}$, entonces se cumple que

$$\sum_{i=1}^N p(x_i) = 1,$$

y si toma una cantidad infinita numerable de valores x_1, x_2, \dots , entonces

$$\sum_{i=1}^{\infty} p(x_i) = 1.$$

Ejemplo 1.9. Si una variable aleatoria toma dos valores, 1 y 2, y $p(1) = 0.3$, entonces $p(2) = 0.7$.

Ejemplo 1.10. Si la probabilidad de que una moneda salga cara es $\frac{1}{3}$, y las tiradas de moneda son eventos independientes, entonces la probabilidad de realizar n tiradas hasta obtener la primera cara es:

$$p(n) = P(\{n-1 \text{ tiradas cruz y una cara}\}) = \left(\frac{2}{3}\right)^{n-1} \frac{1}{3}, \quad n = 1, 2, \dots$$

y se tiene que

$$\sum_{n=1}^{\infty} p(n) = \frac{1}{3} \left(\frac{1}{1 - \frac{2}{3}} \right) = 1.$$

Una variable aleatoria se dice que es una variable **(absolutamente) continua** si existe una función f tal que para todo subconjunto $C \subseteq \mathbb{R}$ se cumple:

$$P(X \in C) = \int_C f(x) dx.$$

La función f se denomina **función de densidad de probabilidad** de la variable aleatoria X .

Por ejemplo, si $C = \{X < 2\}$ y f es la función de densidad de X , entonces

$$P(X \leq 2) = \int_{-\infty}^2 f(x) dx,$$

o si $C = \{-2 < X < 5\}$, entonces

$$P(-2 < X \leq 5) = \int_{-2}^5 f(x) dx.$$

Notemos además que si X es continua, entonces $P(X = a) = \int_a^a f(t) dt$, y por lo tanto $P(X = a) = 0$. En particular, f y la función de distribución acumulada F se relacionan por:

$$F(a) = P(X \leq a) = \int_{-\infty}^a f(x) dx.$$

Derivando con respecto a a tenemos que $F'(a) = f(a)$, para todo $a \in \mathbb{R}$ tal que F es derivable en a .

Además, si $\epsilon > 0$, entonces

$$P(a - \epsilon/2 \leq X \leq a + \epsilon/2) = \int_{a-\epsilon/2}^{a+\epsilon/2} f(x) dx,$$

es decir que para valores de ϵ cercanos a 0 esta probabilidad es aproximadamente $f(a) \cdot \epsilon$. De esta manera $f(a)$ da una medida de la probabilidad de que la variable aleatoria tome valores cercanos a a .

1.2.1. Distribución conjunta

Si X e Y son variables aleatorias sobre un espacio de probabilidad (S, P) , se llama **función de distribución acumulada conjunta** de X e Y a la función $F : \mathbb{R} \times \mathbb{R} \mapsto [0, 1]$ dada por

$$F(a, b) = P(X \leq a, Y \leq b).$$

En particular, si X e Y son variables aleatorias discretas se define la **función de masa de probabilidad conjunta** de X e Y como:

$$p(a, b) = P(X = a, Y = b).$$

Dos variables aleatorias X e Y se dicen **conjuntamente continuas** si existe una función f llamada **función de densidad conjunta** tal que

$$P(X \in C, Y \in D) = \int \int_{x \in C, y \in D} f(x, y) dx dy.$$

Si F es función de distribución conjunta de X e Y , entonces pueden calcularse las distribuciones de probabilidad de X e Y a partir de F , también llamadas **distribuciones marginales** F_X y F_Y . Esto es:

$$F_X(a) = P(X \leq a) = F(a, \infty), \quad F_Y(b) = P(Y \leq b) = F(\infty, b).$$

Si X e Y son variables aleatorias discretas con función de masa conjunta p , entonces las funciones de probabilidad de masa marginales de X e Y están dadas respectivamente por:

$$p_X(a) = \sum_b p(a, b), \quad p_Y(b) = \sum_a p(a, b).$$

Aquí los subíndices de la sumatoria b y a toman todos los valores posibles en el rango de Y y X , respectivamente.

Si X e Y son conjuntamente continuas con función de densidad conjunta f , las distribuciones marginales están dadas por:

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

$$F_Y(b) = \int_{-\infty}^{\infty} \int_{-\infty}^b f(x, y) dy dx = \int_{-\infty}^b f_Y(y) dy.$$

Notemos que las correspondientes densidades marginales f_X y f_Y están dadas por:

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

Así, a partir de la función de distribución conjunta es posible obtener las distribuciones marginales de X e Y . La recíproca no es cierta en general.

1.2.2. Distribución condicional

Si X e Y son variables aleatorias discretas, entonces la **probabilidad de masa condicional** $p_{X|Y}$ se define como

$$p_{X|Y}(x | y) = P(X = x | Y = y).$$

En particular, tenemos que:

$$p_{X|Y}(x | y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p(x, y)}{p_Y(y)}.$$

Si X e Y son conjuntamente continuas, se define la **función de densidad condicional** como:

$$f_{X|Y}(x | y) = \frac{f(x, y)}{f_Y(y)}.$$

En particular, la probabilidad condicional $P(X \leq x | Y = y)$ está dada por

$$P(X \leq a | Y = y) = \int_{-\infty}^a \frac{f(x, y)}{f_Y(y)} dx = \int_{-\infty}^a \frac{f(x, y)}{\int_{-\infty}^{\infty} f(s, y) ds} dx.$$

En particular tenemos una versión de la fórmula de Bayes que relaciona las distribuciones condicionales y marginales para dos variables aleatorias X e Y :

$$p_{X|Y}(x | y) = \frac{p_{Y|X}(y | x)p_X(x)}{p_Y(y)}, \quad f_{X|Y}(x | y) = \frac{f_{Y|X}(y | x)f_X(x)}{f_Y(y)}.$$

Ejemplo 1.11. Consideremos X e Y variables aleatorias continuas con función de densidad conjunta

$$f_{X,Y}(x, y) = \begin{cases} \frac{e^{-x/y} e^{-y}}{y} & x \geq 0, 0 < y < \infty \\ 0 & \text{en otro caso.} \end{cases}$$

Entonces la densidad de Y está dada por:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_0^{\infty} \frac{e^{-x/y} e^{-y}}{y} dx = e^{-y}, \quad 0 < y < \infty,$$

y $f_Y(y) = 0$ en otro caso. Luego la función de densidad condicional de X dado Y está dada por:

$$f_{X|Y}(x | y) = \begin{cases} \frac{e^{-x/y}}{y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{en otro caso.} \end{cases}$$

Luego la probabilidad condicional de X dado Y se calcula como:

$$P(X \leq a | Y = y) = \int_0^a \frac{e^{-x/y}}{y} dx = 1 - e^{-a/y}, \quad 0 < a < \infty, 0 < y < \infty.$$

Dos variables aleatorias X e Y son **independientes** si para todo C, D subconjuntos de \mathbb{R} se cumple que

$$P(X \in C, Y \in D) = P(X \in C) \cdot P(Y \in D).$$

Esto es, si los conjuntos $\{X \in C\}$ y $\{Y \in D\}$ son eventos independientes.

En tal caso, si X e Y son variables aleatorias discretas se cumple que $p(a, b) = p_X(a) \cdot p_Y(b)$, y si X e Y son variables aleatorias conjuntamente continuas se cumple que $f(x, y) = f_X(x) \cdot f_Y(y)$. Por lo tanto, si X e Y son independientes, la distribución conjunta se obtiene a partir de las distribuciones de X e Y .

1.2.3. Convolución de distribuciones

Dadas dos variables aleatorias X e Y sobre un espacio de probabilidad (S, P) consideremos la distribución de la suma de estas variables, $X + Y$. Así, si X e Y son variables discretas, entonces $X + Y$ también es discreta y se tiene que $X + Y$ toma un valor a cada vez que X e Y toman valores x y $a - x$, y recíprocamente. Por lo tanto:

$$P(X + Y = a) = \sum_x P(X = x, Y = a - x) = \sum_x p(x, a - x),$$

donde $p(x, y)$ denota la probabilidad conjunta de X e Y . Si X e Y son conjuntamente continuas, entonces $X + Y \leq a$ cada vez que X toma un valor x e Y es menor o igual a $a - x$. Luego

$$P(X + Y \leq a) = \int_{-\infty}^{\infty} \int_{-\infty}^{a-x} f_{X,Y}(x, y) dy dx = \int_{-\infty}^{\infty} \int_{-\infty}^a f_{X,Y}(x, y - x) dy dx,$$

con $f_{X,Y}$ la densidad conjunta de X e Y .

Si las variables son independientes, entonces para el caso discreto resulta que la probabilidad de masa de la variable aleatoria $X + Y$ está dada por:

$$P(X + Y = a) = \sum_x p_X(x) p_Y(a - x).$$

Para el caso continuo tenemos que

$$P(X + Y \leq a) = \int_{-\infty}^{\infty} \int_{-\infty}^a f_X(x) f_Y(y - x) dy dx = \int_{-\infty}^a \int_{-\infty}^{\infty} f_X(x) f_Y(y - x) dx dy,$$

de donde vemos que la densidad de $X + Y$ está dada por $\int_{-\infty}^{\infty} f_X(x) f_Y(y - x) dx$. Aquí p_X , p_Y , f_X y f_Y son las probabilidades de masa y densidades marginales respectivamente.

Esto da lugar al concepto de **convolución** de funciones de probabilidad. En particular, para las probabilidades de masa la convolución $p_X * p_Y(a)$ se define por:

$$p_X * p_Y(a) = \sum_x p_X(x) p_Y(a - x).$$

Para las funciones de densidad, la convolución $f_X * f_Y(a)$ se define por:

$$f_X * f_Y(a) = \int_{-\infty}^{\infty} f_X(x) f_Y(a - x) dx.$$

Así, si X e Y son variables aleatorias discretas independientes, entonces la probabilidad de masa de $X + Y$ está dada por la convolución $p_X * p_Y$. Si son continuas e independientes, la densidad de $X + Y$ está dada por la convolución de las densidades marginales, $f_X * f_Y$.

Ejemplo 1.12. Consideremos X e Y dos variables aleatorias independientes con densidades f y g respectivamente, dadas por:

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad g(x) = \begin{cases} 3e^{-3x} & x > 0 \\ 0 & x \leq 0. \end{cases}$$

Como $f(x) = 0$ para valores de x negativos, y $g(x - y) = 0$ si $y > x$, tenemos que para $x > 0$:

$$\begin{aligned} (f * g)(x) &= \int_{-\infty}^{\infty} f(y) g(x - y) dy = \int_0^x f(y) g(x - y) dy \\ &= \int_0^x e^{-y} 3e^{-3(x-y)} dy \\ &= 3e^{-3x} \int_0^x e^{2y} dy \\ &= \frac{3}{2} e^{-3x} (e^{2x} - 1) = \frac{3}{2} (e^{-x} - e^{-3x}). \end{aligned}$$

Para $x \leq 0$ se verifica $(f * g)(x) = 0$. Así, para $a > 0$ tenemos que:

$$P(X+Y \leq a) = \int_0^a \frac{3}{2} (e^{-x} - e^{-3x}) dx = \frac{3}{2} \left(1 - e^{-a} - \frac{1}{3} (1 - e^{-3a}) \right) = \frac{1}{2} (2 - 3e^{-a} + e^{-3a}).$$

1.2.4. Valor esperado

Dada una variable aleatoria discreta X que toma valores x_i , $i = 1, 2, \dots, n, \dots$, se llama **valor esperado** o **esperanza matemática** a la cantidad (si existe)

$$E[X] = \sum_i x_i P(X = x_i).$$

Si X es una variable aleatoria continua, su valor esperado se define por el valor de la integral (si existe):

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx.$$

Es importante notar que el valor esperado no es necesariamente un valor posible de X .

Ejemplo 1.13. Si X toma los valores 1, 2, 3 y 4, con $p(1) = p(2) = 0.3$, $p(3) = p(4) = 0.2$, entonces

$$E[X] = 1 \cdot 0.3 + 2 \cdot 0.3 + 3 \cdot 0.2 + 4 \cdot 0.2 = 2.3.$$

Sean X e Y dos variables aleatorias sobre un mismo espacio de probabilidad S . Entonces se cumplen las siguientes propiedades:

a) Si $g : \mathbb{R} \mapsto \mathbb{R}$, entonces $g(X)$ es una variable aleatoria y

$$\begin{aligned} E[g(X)] &= \sum_i g(x_i) p(x_i), & (\text{si } X \text{ es discreta}), \\ E[g(X)] &= \int_{-\infty}^{\infty} g(x) f(x) dx, & (\text{si } X \text{ es continua}). \end{aligned}$$

En particular, tomando $g(x) = ax + b$ se deduce que

$$E[aX + b] = aE[X] + b.$$

b) El valor esperado es un operador lineal. Esto es,

$$E[X + Y] = E[X] + E[Y].$$

1.2.5. Varianza

La **varianza** es una medida de la dispersión de X en torno a su valor esperado $E[X] = \mu$, y está definido por:

$$\text{Var}(X) = E[(X - \mu)^2].$$

Notemos que

$$\text{Var}(X) = E[(X^2 - 2X\mu + \mu^2)] = E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - \mu^2.$$

La varianza es siempre un número positivo a menos que la variable aleatoria sea siempre constante. Es importante notar que si a y b son números reales, entonces

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

La varianza no verifica la condición de linealidad. En efecto, si $E[X] = \mu$, $E[Y] = \theta$, entonces

$$\begin{aligned} \text{Var}(X + Y) &= E[((X + Y) - (\mu + \theta))^2] \\ &= E[(X - \mu)^2 + (Y - \theta)^2 + 2(X - \mu)(Y - \theta)] \\ &= E[(X - \mu)^2] + E[(Y - \theta)^2] + 2E[(X - \mu)(Y - \theta)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2E[(X - \mu)(Y - \theta)]. \end{aligned}$$

Si X e Y son variables aleatorias, se define la **covarianza** de X e Y por:

$$\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

De esta manera,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{cov}(X, Y).$$

Si X e Y son independientes entonces $\text{cov}(X, Y) = 0$ (la recíproca en general no es cierta). Esto puede verse observando que $\text{cov}(X, Y) = E[XY] - E[X]E[Y]$. Luego, si X e Y son independientes y conjuntamente continuas con densidad conjunta f , entonces $f(x, y) = f_X(x)f_Y(y)$ y por lo tanto:

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_X(x) y f_Y(y) dx dy = E[X]E[Y].$$

Si X e Y son discretas con probabilidad de masa conjunta p , entonces $p(x, y) = p_X(x)p_Y(y)$ y

$$E[XY] = \sum_i \sum_j x_i y_j p(x_i, y_j) = \sum_i \sum_j x_i p_X(x_i) y_j p_Y(y_j) = E[X]E[Y].$$

Por lo tanto, si X e Y son variables aleatorias independientes, se cumple que

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Otra medida de dispersión de una variable aleatoria es la **desviación estándar** $\sigma(X)$, definida por

$$\sigma(X) = \sqrt{\text{Var}(X)}.$$

Tiene la ventaja de mantener las mismas unidades (distancia, longitud, tiempo, etc.) que X y $E[X]$. A su vez, se define la **correlación** de dos variables aleatorias X e Y por:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}.$$

Se cumple que $\rho(X, Y)$ es un número entre -1 , y 1 y da una medida normalizada de la covarianza entre dos variables aleatorias.

1.2.6. Desigualdad de Chebyshev

Si X toma sólo valores no negativos y $a > 0$, entonces se cumple la desigualdad

$$P(X \geq a) \leq \frac{E[X]}{a}. \quad (1.5)$$

Para probar (1.5) consideramos la variable aleatoria Y dada por

$$Y = \begin{cases} a & \text{si } X \geq a \\ 0 & \text{en caso contrario.} \end{cases}$$

Dado que $Y \leq X$, lo mismo vale para sus valores esperados: $E[Y] \leq E[X]$. Luego,

$$a P(X \geq a) = E[Y] \leq E[X],$$

de donde se deduce el resultado. De esta propiedad para variables aleatorias no negativas puede derivarse la Desigualdad de Chebyshev.

Teorema 1.1 (Desigualdad de Chebyshev). Si X es variable aleatoria con media μ y varianza σ^2 , entonces para $k > 0$

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Para la demostración, basta considerar la variable aleatoria

$$Y = \frac{|X - \mu|^2}{\sigma^2},$$

que toma valores positivos y su valor esperado es 1: $E[Y] = 1$. Aplicando la desigualdad (1.5) para un $k > 0$, tenemos que:

$$P(Y \geq k^2) \leq \frac{1}{k^2}.$$

Dado que los eventos $\{Y \geq k^2\}$ y $\{\frac{|X - \mu|}{\sigma} \geq k\}$ son iguales, se sigue el resultado.

Así por ejemplo, la probabilidad que los valores de una variable aleatoria estén a una distancia menor a 2 desviaciones estándar del valor esperado es mayor a $1 - 0.25 = 0.75$.

1.2.7. Leyes de los grandes números

Dos variables aleatorias se dicen **idénticamente distribuidas** si tienen una misma función de distribución acumulada.

Los siguientes resultados teóricos enuncian que si se tiene una sucesión de variables aleatorias independientes, idénticamente distribuidas, con valor esperado μ , entonces los promedios de estas variables *convergen*, en algún sentido de convergencia, al valor μ .

Específicamente, si $X_1, X_2, \dots, X_n, \dots$ son variables aleatorias independientes e idénticamente distribuidas, con media μ , se cumplen las siguientes propiedades:

- **Ley débil de los grandes números:**

$$P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \mu\right| > \epsilon\right) \rightarrow 0 \quad n \rightarrow \infty.$$

- **Ley fuerte de los grandes números:**

Con probabilidad 1 se cumple que:

$$\lim_{n \rightarrow \infty} \frac{X_1 + X_2 + \dots + X_n}{n} = \mu.$$

1.3. Distribuciones de probabilidad

En la literatura se puede encontrar un gran número de distribuciones de probabilidad teóricas. En esta sección consideraremos algunas de estas distribuciones que luego utilizaremos para modelado y simulación.

1.3.1. Variables aleatorias discretas

Distribución uniforme discreta: $U\{1, n\}$

Se dice que una variable aleatoria tiene **distribución uniforme** si todos sus valores son equiprobables. Con $U\{1, n\}$ denotaremos a la variable aleatoria que toma valores en el conjunto $\{1, 2, \dots, n\}$, todos con la misma probabilidad $\frac{1}{n}$.

$$p(i) = \frac{1}{n}, \quad i = 1, 2, \dots, n.$$

El valor esperado y la varianza están dados por:

$$E[X] = \frac{n+1}{2}, \quad \text{Var}(X) = \frac{n^2-1}{12}.$$

Distribución de Bernoulli: $B(p)$

Una variable aleatoria que toma dos valores con probabilidad p (éxito) y $1-p$ (fracaso), se dice **de Bernoulli**. La distribución de Bernoulli teórica se corresponde con la variable aleatoria que toma los valores 1 y 0:

$$X = \begin{cases} 1 & \text{con prob. } p \\ 0 & \text{con prob. } 1-p. \end{cases}$$

Su valor esperado y varianza están dados por:

$$E[X] = p, \quad \text{Var}(X) = p \cdot (1-p).$$

Distribución Binomial $B(n, p)$

Si consideramos un experimento que consiste en n ensayos independientes, cada uno con probabilidad p de éxito, entonces el número de éxitos tiene una distribución binomial de parámetros n y p . El rango de la variable aleatoria es el conjunto $\{0, 1, 2, \dots, n\}$, y la función de masa de probabilidad está dada por:

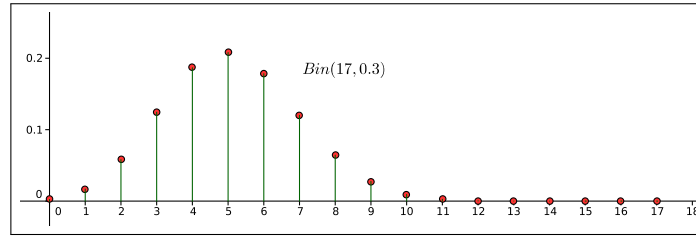
$$p(i) = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad 0 \leq i \leq n.$$

Más adelante resultará útil la siguiente fórmula recursiva para las probabilidades de masa:

$$p(0) = (1-p)^n, \quad p(i+1) = \frac{n-i}{i+1} \frac{p}{(1-p)} p(i), \quad 0 \leq i \leq n-1.$$

El valor esperado y la varianza están dados por:

$$E[X] = np, \quad \text{Var}(X) = np(1-p).$$

Figura 1.1: Distribución binomial. $\text{Bin}(17, 0.3)$ **Distribución de Poisson: $\mathcal{P}(\lambda)$**

Una variable aleatoria se dice que es de **Poisson con parámetro λ** si toma valores en $\mathbb{N} \cup 0$ con probabilidad de masa

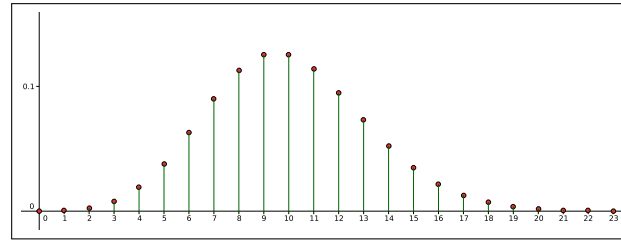
$$p(i) = P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}, \quad i \geq 0.$$

Una fórmula recursiva para estas probabilidades está dada por:

$$p(0) = e^{-\lambda}, \quad p(i+1) = \frac{\lambda}{i+1} p(i), \quad i \geq 0.$$

El valor esperado y la varianza están dados por:

$$E[X] = \lambda, \quad \text{Var}(X) = \lambda.$$

Figura 1.2: Distribución de Poisson. $\lambda = 10$ **Distribución Geométrica: $\text{Geom}(p)$**

Dada una sucesión de ensayos independientes con probabilidad p de éxito, la variable aleatoria geométrica cuenta el número de ensayos independientes hasta obtener el primer éxito. Su rango es el conjunto de números naturales, \mathbb{N} .

La función de probabilidad de masa está dada por:

$$p(n) = P(X = n) = p(1-p)^{n-1}, \quad n \geq 1.$$

El valor esperado y la varianza están dados por:

$$E[X] = \frac{1}{p}, \quad \text{Var}(X) = \frac{1-p}{p^2}$$

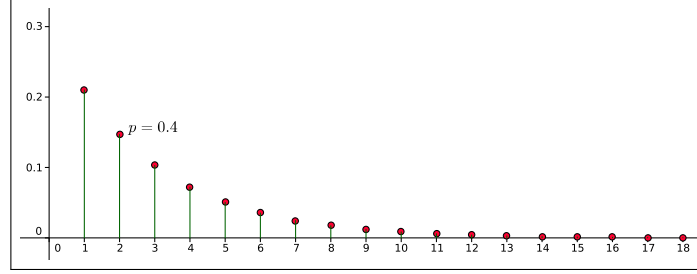


Figura 1.3: Distribución geométrica. $p = 0.4$

Distribución Binomial Negativa o Pascal: $\text{Bn}(r, p)$

Esta distribución se corresponde con la variable aleatoria que mide el número de ensayos independientes con probabilidad de éxito p , hasta obtener r éxitos. La variable toma valores en el intervalo $\{r, r+1, r+2, \dots\} = \{n \in \mathbb{N} \mid n \geq r\}$. La función de probabilidad de masa está dada por

$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n \geq r.$$

El valor esperado y la varianza están dados por:

$$E[X] = \frac{r}{p}, \quad \text{Var}(X) = \frac{r(1-p)}{p^2}.$$

Distribución Hipergeométrica $H(n, N, M)$

Esta distribución se corresponde con la variable aleatoria que mide el número de éxitos en una muestra de tamaño n extraída de un conjunto de $N+M$ elementos, donde un éxito equivale a extraer un elemento del subconjunto de cardinal N .

El rango de esta distribución es $\{0, 1, 2, \dots, n\}$. La función de probabilidad de masa está dada por:

$$p(i) = P(X = i) = \frac{\binom{N}{i} \binom{M}{n-i}}{\binom{N+M}{n}}.$$

La probabilidad $p(i)$ es 0 si $i > n$ o $n-i > M$. El valor esperado y la varianza están dadas por:

$$E[X] = \frac{nN}{N+M}, \quad \text{Var}(X) = \frac{nNM}{(N+M)^2} \left(1 - \frac{n-1}{N+M-1}\right).$$

1.3.2. Variables aleatorias continuas

Denotaremos con I_A a la **función indicadora del conjunto** A , dada por

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}.$$

Distribución uniforme $U(a, b)$

Una variable aleatoria continua X se dice uniformemente distribuida en (a, b) si su función de densidad está dada por

$$f(x) = \frac{1}{b-a} \mathbb{I}_{(a,b)}(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{c.c.} \end{cases}$$

Su función de distribución acumulada está dada por:

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

y la varianza y su valor esperado son:

$$E[X] = \frac{a+b}{2}, \quad \text{Var}(X) = \frac{1}{12}(b-a)^2.$$

Distribución Normal $N(\mu, \sigma)$

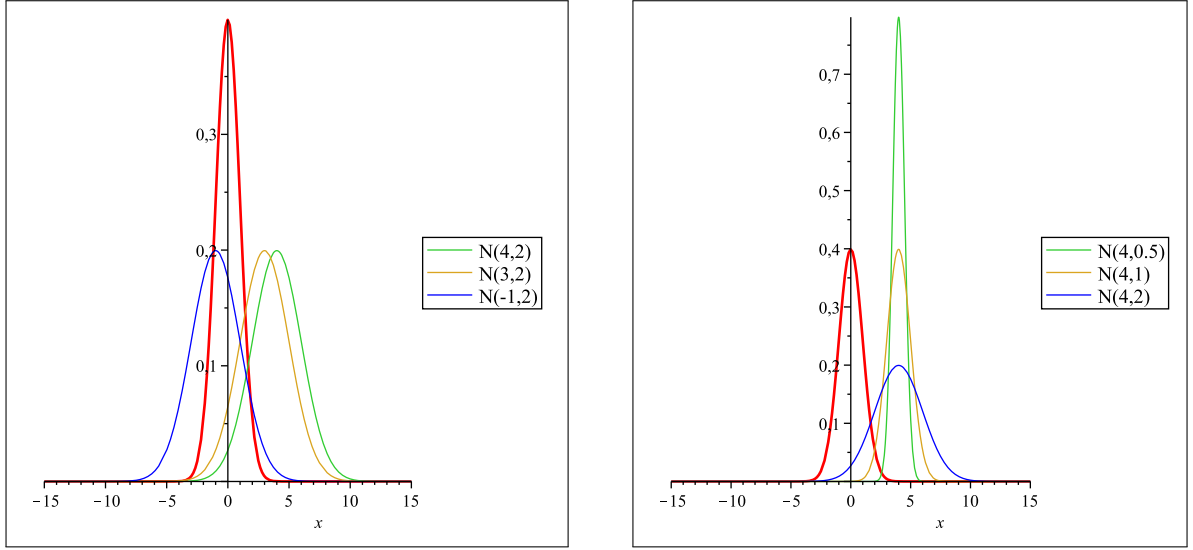
Una variable aleatoria continua X se dice normalmente distribuida con media μ y varianza σ^2 si su función de densidad de probabilidad está dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

En tal caso usamos la notación $X \sim N(\mu, \sigma)$. Si $Z \sim N(0, 1)$ se dice que su distribución es **normal estándar**. La función de densidad es entonces:

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

La Figura 1.3.2 muestra distintas distribuciones normales, comparadas con la distribución normal estándar. En el primer gráfico la varianza es constante y en el segundo es constante la media.

Figura 1.4: Ejemplos de distribución normal, variando μ y σ

Si $X \sim N(\mu, \sigma)$, entonces $\frac{X-\mu}{\sigma}$ tiene distribución normal estándar. Denotemos con Φ a la función de distribución acumulada de una variable aleatoria normal estándar, $Z \sim N(0, 1)$:

$$\Phi(x) = P(Z \leq x) = \frac{1}{2\pi} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt. \quad (1.6)$$

La integral $\Phi(x)$ dada en (1.6) no tiene una fórmula cerrada, por lo cual es usual utilizar valores tabulados para el cálculo o interpolación de valores de Φ . De la expresión para la función de densidad, se observa que es una función par, esto es, $f(x) = f(-x)$. Luego se cumple que

$$\Phi(x) = 1 - \Phi(-x), \quad \text{para todo } x \in \mathbb{R}.$$

Si X es normal con media μ y varianza σ^2 , entonces su función de distribución acumulada puede expresarse en términos de Φ :

$$F(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Algunos valores importantes a recordar son las probabilidades de que los valores de una variable aleatoria normal se distribuyan alrededor de la media, a una distancia menor de $k\sigma$, para ciertos valores de k . Para $k = 1, 2, 3$ tenemos:

$$P(|X - \mu| < \sigma) \simeq 68 \%, \quad P(|X - \mu| < 2\sigma) \simeq 95 \%, \quad P(|X - \mu| < 3\sigma) \simeq 99.7 \%.$$

Si $Z \sim N(0, 1)$ y α es un número entre 0 y 1, se suele denotar z_α al número real tal que

$$P(Z > z_\alpha) = \alpha.$$

Así por ejemplo,

$$z_{0.05} = 1.64, \quad z_{0.025} = 1.96, \quad z_{0.01} = 2.33, \quad z_{0.005} = 2.58.$$

Debido a la simetría de la densidad de la normal estándar, estos valores indican en particular que:

$$P(-1.64 \leq Z \leq 1.64) = 90 \%, \quad P(-2.58 \leq Z \leq 2.58) = 99 \%,$$

y resultados similares para el 95 % y el 98 %.

Un resultado importante en la teoría de probabilidad es el llamado Teorema Central del Límite. Este teorema establece que la suma de n variables aleatorias independientes, igualmente distribuidas, todas con media μ y varianza σ^2 , tiene una distribución aproximadamente normal, con media $n\mu$ y varianza $n\sigma^2$. Más precisamente:

Teorema 1.2 (Teorema Central del límite). Sean X_1, X_2, \dots , variables aleatorias independientes igualmente distribuidas con media μ y varianza σ^2 . Entonces

$$\lim_{n \rightarrow \infty} P\left(\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} < x\right) = \Phi(x).$$

Este teorema resultará útil para estimaciones estadísticas de intervalos de confianza a partir de una muestra de tamaño n .

Distribución exponencial $\mathcal{E}(\lambda)$

Una variable aleatoria X tiene distribución exponencial con parámetro λ , $\lambda > 0$, si su función de densidad está dada por

$$f_\lambda(x) = \lambda e^{-\lambda x} \mathbb{I}_{(0, \infty)}(x).$$

Su valor esperado y varianza están dados por:

$$E[X] = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}.$$

Si $X \sim \mathcal{E}(\lambda)$, entonces $cX \sim \mathcal{E}(\frac{1}{c}\lambda)$.

Se dice que una variable aleatoria X tiene **falta de memoria** si

$$P(X > s + t \mid X > s) = P(X > t),$$

para todo $s, t \in \mathbb{R}$. En efecto, si X tiene distribución exponencial con parámetro λ , entonces para cada t se cumple:

$$P(X > t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}.$$

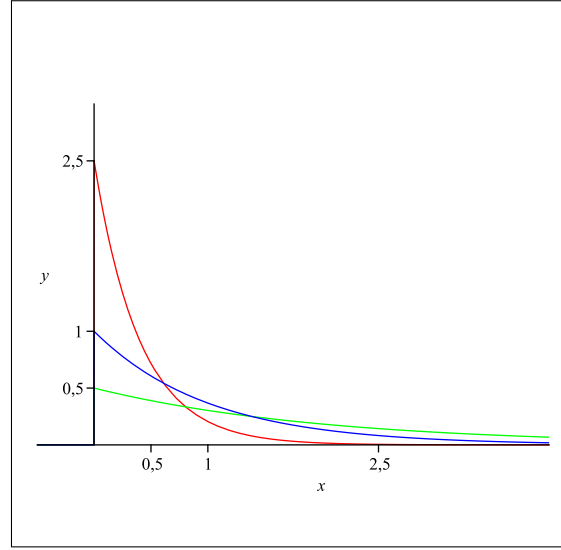


Figura 1.5: Función de densidad de variables exponenciales

Luego,

$$\begin{aligned}
 P(X > s + t \mid X > s) &= \frac{P(X > s + t, X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)} \\
 &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} \\
 &= P(X > t).
 \end{aligned}$$

Las variables aleatorias con distribución exponencial son las únicas variables aleatorias continuas con la propiedad de falta de memoria. El análogo en el caso discreto son las variables aleatorias geométricas.

Sean X_1, X_2, \dots, X_n son variables aleatorias independientes con función de distribución acumulada F_1, F_2, \dots, F_n , y sea M el mínimo entre estas variables:

$$M = \min_{1 \leq i \leq n} \{X_1, X_2, \dots, X_n\}.$$

Entonces M es una variable aleatoria, y su función de distribución acumulada cumple que:

$$1 - F_M(x) = P(M > x) = (1 - F_1(x)) \cdot (1 - F_2(x)) \cdots (1 - F_n(x)).$$

En particular, si las variables aleatorias tienen distribución exponencial:

$$X_i \sim \mathcal{E}(\lambda_i),$$

entonces su distribución satisface:

$$1 - F_M(x) = e^{-\lambda_1 x} \cdot e^{-\lambda_2 x} \cdots e^{-\lambda_n x} = e^{-(\sum_i \lambda_i) x}.$$

Por lo tanto, la distribución del **mínimo entre n variables aleatorias exponenciales independientes** es exponencial:

$$M \sim \mathcal{E}(\lambda_1 + \lambda_2 + \cdots + \lambda_n).$$

Distribución Lognormal $LN(\mu, \sigma)$

Una variable aleatoria continua X se dice **lognormal** si su logaritmo $\ln(X)$ tiene distribución normal. Denotaremos $X \sim LN(\mu, \sigma)$ si el logaritmo de X tiene media μ y desviación estándar σ . En tal caso su función de densidad está dada por:

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-(\log(x)-\mu)^2/(2\sigma^2)} \cdot \mathbb{I}_{(0,\infty)}(x).$$

Su valor esperado y varianza están dados por:

$$E[X] = e^{\mu+\sigma^2/2}, \quad Var(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

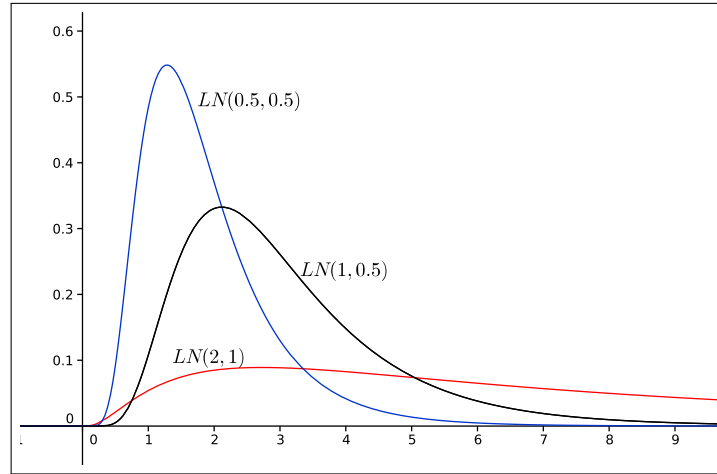


Figura 1.6: Distribución lognormal

Distribución Gamma $\Gamma(\alpha, \beta)$

Una variable aleatoria Gamma con parámetros α y β tiene una función de densidad dada por:

$$f(x) = \frac{1}{\Gamma(\alpha)} \beta^{-\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad x > 0.$$

En la Figura 1.7 se muestran los gráficos de $\Gamma(\alpha, 1)$, para $\alpha = 0.5, 1, 2$ y 3 . Notar de la definición de f que $\alpha = 1$ corresponde a la distribución exponencial $\mathcal{E}(1)$. Su valor esperado y varianza están dados por:

$$E[X] = \alpha\beta, \quad Var(X) = \alpha\beta^2.$$

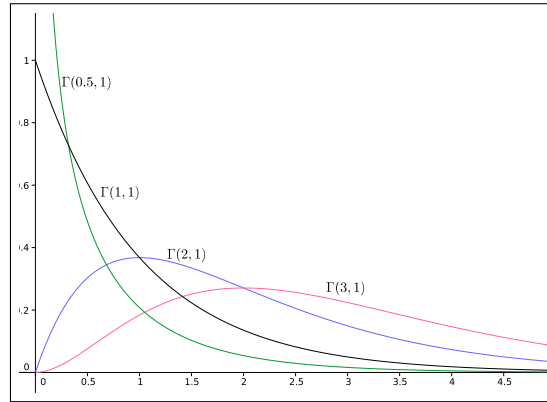


Figura 1.7: Distribuciones Gamma

Distribución Weibull (α, β)

Una variable aleatoria continua Weibull con parámetros α y β , con $\alpha > 0$ y $\beta > 0$, tiene función de densidad dada por:

$$f(x) = \alpha \beta^{-\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} \mathbb{I}_{(0, \infty)}(x).$$

En la Figura 1.8 se muestran gráficos para $\beta = 1$ y $\alpha = 0.5, 1, 2$ y 3 . Su valor esperado y varianza son:

$$E[X] = \frac{\beta}{\alpha} \Gamma\left(\frac{1}{\alpha}\right) \quad \text{Var}(X) = \frac{\beta^2}{\alpha} \left[2\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha} \left(\Gamma\left(\frac{1}{\alpha}\right) \right)^2 \right].$$

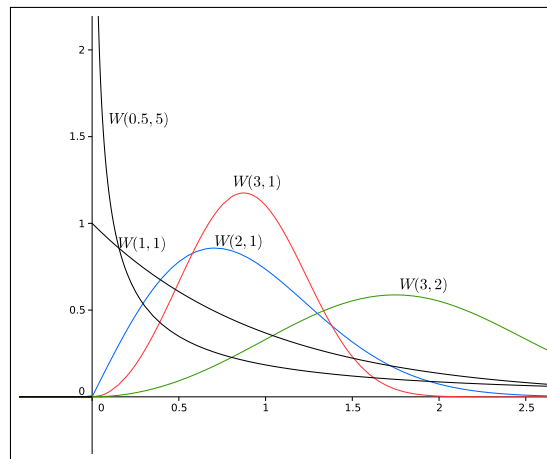


Figura 1.8: Distribución Weibull