Repaso de Cálculo de Probabilidades

Experimentos y sucesos aleatorios

Se dice que un experimento es aleatorio si se verifican las siguientes condiciones:

- Se puede repetir indefinidamente, siempre en las mismas condiciones;
- Antes de realizarlo, no se puede predecir el resultado que se va a obtener;
- El resultado que se obtenga, e, pertenece a un conjunto conocido previamente de resultados posibles. A este conjunto, de resultados posibles, lo denominaremos **espacio muestral** y lo denotaremos normalmente mediante la letra E. Los elementos del espacio muestral se denominan **sucesos elementales**.

$$e_1, e_2 \in E \Longrightarrow e_1, e_2 \text{ son sucesos elementales.}$$

Cualquier subconjunto de E se denominará **suceso aleatorio**, y se denotará normalmente con las letras $A, B, ...$

$$A, B \subset E \Longrightarrow A, B \text{ son sucesos aleatorios.}$$

Se puede observar que los sucesos elementales son sucesos aleatorios compuestos por un sólo elemento. Por supuesto los sucesos aleatorios son más generales que los elementales, ya que son conjuntos que pueden contener no a uno sólo, sino a una infinidad de sucesos elementales (y también no contener ninguno). Sucesos aleatorios que aparecen con gran frecuencia en el cálculo de probabilidades son los siguientes:

Suceso seguro: Es aquel que siempre se verifica después del experimento aleatorio, es decir, el mismo E

$$E \subset E \Longrightarrow E \text{ es el suceso seguro.}$$

Suceso imposible: Es aquel que nunca se verifica como resultado del experimento aleatorio. Como debe ser un subconjunto de E, la única posibilidad es que el suceso imposible sea el conjunto vacío: $\emptyset \subset E$.

Suceso contrario a un suceso A: También se denomina complementario de A, y es el suceso que se verifica si, como resultado del experimento aleatorio, no se verifica A. Se acostumbra a denotar con el símbolo $\overline{A}$ ó A^c. Así, $A^c = \{e \in E : e \notin A\}$.

Ejemplo

Si realizamos el experimento aleatorio de lanzar un dado al aire, tenemos:

$$\begin{array}{rcl} \text{Sucesos elementales} & \longrightarrow & 1,\ 2,\ 3,\ 4,\ 5,\ 6 \\ \text{Espacio muestral} & \longrightarrow & E = \{1,2,3,4,5,6\} \\ \text{Sucesos aleatorios} & \longrightarrow & \left\{ \begin{array}{c} \emptyset \text{ suceso imposible} \\ E \text{ suceso seguro} \\ \{1,2,3\} \\ \{4,5\} \\ \{2,4,6\} = \overline{\{1,2,3\}} \\ \cdots \end{array} \right. \end{array}$$

Para trabajar con el cálculo de probabilidades es necesario fijar previamente cierta terminología. Vamos a introducir parte de ella a continuación.

Operaciones básicas con sucesos aleatorios

Al ser los sucesos aleatorios nada más que subconjuntos de un conjunto E, espacio muestral, podemos aplicar las conocidas operaciones con conjuntos, como son la unión, intersección y diferencia:

Unión: Dados dos sucesos aleatorios $A, B \subset E$, se denomina suceso unión de A y B al conjunto formado por todos los sucesos elementales que pertenecen a A o bien pertenecen a B, incluyendo los que están en ambos simultáneamente, es decir

$$A \cup B = \{e \in E : e \in A \text{ ó } e \in B\}.$$

Como ejemplo, tenemos que la unión de un suceso cualquiera con su complementario es el suceso seguro.

Volviendo al ejemplo del lanzamiento de un dado, si $A = \{1,2,3\}$ y $B = \{3,4\}$, el suceso unión de A y B es $A \cup B = \{1,2,3,4\}$.

Intersección: Dados dos sucesos aleatorios $A, B \subset E$, se denomina suceso intersección de A y B al conjunto formado por todos los sucesos elementales que pertenecen a A y B a la vez, es decir,

$$A \cap B = \{e \in E : e \in A \text{ y además } e \in B\}.$$

Un ejemplo de intersección es la de un suceso aleatorio cualquiera, $A \subset E$, con su complementario, que es el suceso imposible.

Volviendo al ejemplo del dado, si $A = \{1, 2, 3\}$ y $B = \{3, 4\}$, el suceso intersección de A y B es $A \cap B = \{3\}$.

Diferencia: Dados dos sucesos aleatorios aleatorios $A, B \subset E$, se llama suceso diferencia de A y B, y se representa mediante $A \backslash B$, o bien $A - B$, al suceso formado por todos los sucesos elementales que pertenecen a A, pero no a B:

$$A - B = \{e \in E : e \in A \text{ y además } e \notin B\} = A \cap B^c.$$

Por ejemplo, si $A = \{1, 2, 3\}$ y $B = \{3, 4\}$, $A - B = \{1, 2\}$ y $B - A = \{4\}$

Obsérvese que el suceso contrario de un suceso A, puede escribirse como la diferencia del suceso seguro menos éste, o sea, $A^c = \{e \in E : e \notin A\} = E \backslash A$.

Diferencia simétrica: Si $A, B \in E$, se denomina suceso diferencia simétrica de A y B, y se representa mediante $A \triangle B$, al suceso aleatorio formado por todos los sucesos elementales que pertenecen a A y no a B, y los que están en B y no en A:

$$A \triangle B = (A \backslash B) \cup (B \backslash A) = (A \cup B) \setminus (B \cap A).$$

Por ejemplo, si $A = \{1, 2, 3\}$ y $B = \{3, 4\}$, $A \triangle B = \{1, 2, 4\} = B \triangle A$.

Hay ciertas propiedades que relacionan la unión, intersección y suceso contrario, que son conocidas bajo el nombre de **Leyes de Morgan**:

$$\begin{aligned} \overline{A \cup B} &= \overline{A} \cap \overline{B} \\ \overline{A \cap B} &= \overline{A} \cup \overline{B} \end{aligned}$$

Experimentos aleatorios y probabilidad

Se denominan experimentos deterministas aquellos que realizados de una misma forma y con las mismas condiciones iniciales, ofrecen siempre el mismo resultado. Como ejemplo,

tenemos que un objeto de cualquier masa partiendo de un estado inicial de reposo, y dejado caer al vacío desde una torre, llega siempre al suelo con la misma velocidad $v = \sqrt{2gh}$.

Cuando en un experimento no se puede predecir el resultado final, hablamos de experimento aleatorio. Este es el caso cuando lanzamos un dado y observamos su resultado.

En los experimentos aleatorios se observa que cuando el número de experimentos aumenta, las frecuencias relativas con las que ocurre cierto suceso e, $f_n(e)$,

$$f_n(e) = \frac{\text{número de ocurrencias de } e}{n}$$

tiende a converger hacia cierta cantidad que se puede interpretar como la *probabilidad* de ocurrir e:

$$P(E) = \lim_{n\to\infty} f_n(e).$$

Definición axiomática de probabilidad

Para hacer una definición rigurosa de la probabilidad, necesitamos precisar ciertas leyes o axiomas que debe cumplir una función de probabilidad. Intuitivamente, estos axiomas deberían implicar, entre otras, las siguientes cuestiones, que son lógicas en términos de lo que se puede esperar de una función de probabilidad:

Concepto axiomático de probabilidad

Toda medida de incertidumbre debe cumplir los siguientes axiomas para ser una *medida de probabilidad*:

Axioma 1 La probabilidad es una función que sólo toma valores positivos comprendidos entre 0 y 1.

Axioma 2. La probabilidad del suceso seguro es 1: $P(E) = 1$.

Axioma 3. La probabilidad de la unión numerable de sucesos disjuntos es la suma de sus probabilidades (ver figura).

El tercer axioma de probabilidad indica que si $A = A_1 \cup A_2 \cup \cdots$, con $A_i \cap A_j = \emptyset$, entonces $P(A) = P(A_1) + P(A_2) + \cdots$

Probabilidad condicionada e independencia de sucesos

En numerosas ocasiones, tendremos que modelizar una situación en la que se dispone de información adicional, debiendo condicionarse a sucesos o circunstancias. Formalmente, suponemos que estamos interesados en un suceso A; hemos asignado $P(A)$ y nos informan de que ha ocurrido B y queremos saber cómo cambian mis creencias sobre A.

Obviamente, en algunos casos no cambiarán tales creencias. Por ejemplo, si nos dicen que $A = E$ (esto es, no nos dicen nada nuevo, no aportan información), $P(B)$ no debe cambiar. En la mayor parte de los casos, sin embargo, el aporte de nueva información modifica la probabilidad.

El concepto básico para modelizar tales ideas es la *probabilidad condicionada* $P(B|A)$. Su definición es la siguiente.

Sea $B \subset E$ un suceso aleatorio de probabilidad no nula, $P(B) > 0$. Para cualquier otro suceso $A \subset E$, se llama probabilidad condicionada de A respecto de B a la cantidad que representamos mediante $P(A|B)$, y se calcula como

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

Ejemplo

Se lanza un dado al aire ¿Cuál es la probabilidad de que salga el número 4? Si sabemos que el resultado ha sido un número par, ¿se modifica esta probabilidad?

Solución: El espacio muestral que corresponde a este experimento es $E = \{1, 2, 3, 4, 5, 6\}$ y se ha de calcular la probabilidad del suceso $A = \{4\}$. Si el dado no está trucado, todos los números tienen la misma probabilidad de salir y, siguiendo la definición de probabilidad de Laplace,

$$P(A) = \frac{\text{casos favorables}}{\text{casos posibles}} = \frac{1}{6}$$

Obsérvese que para calcular la probabilidad de A según la definición de Laplace hemos tenido que suponer previamente que todos los elementos del espacio muestral tienen la misma probabilidad de salir, es decir:

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6).$$

Por otro lado, si se sabe que ha salido un número par, de nuevo por la definición de probabilidad de Laplace tendríamos

$$P(A|\text{par}) = \frac{\text{casos favorables}}{\text{casos posibles}} = \frac{\text{número de elementos en } \{4\}}{\text{número de elementos en } \{2, 4, 6\}} = \frac{1}{3}.$$

Esta misma probabilidad se podría haber calculado siguiendo la definición de la probabilidad condicionada, ya que si escribimos

$$\begin{aligned} P(A) &= \frac{1}{6}, \\ P(\text{par}) &= \frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2} \\ P(A\cap \text{par}) &= \frac{1}{6}, \end{aligned}$$

y entonces

$$P(A|\text{par}) = \frac{P(A\cap \text{par})}{P(\text{par})} = \frac{1/6}{1/2} = \frac{1}{3},$$

que por supuesto coincide con el mismo valor que calculamos usando la definición de probabilidad de Laplace.

Observación

Obsérvese que según la definición de probabilidad condicionada, se puede escribir la probabilidad de la intersección de dos sucesos de probabilidad no nula como

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

o sea, la probabilidad de la intersección de dos sucesos, es la probabilidad de uno cualquiera de ellos, multiplicada por la probabilidad del segundo sabiendo que ha ocurrido el primero.

Si entre dos sucesos no existe ninguna relación cabe esperar que la expresión *sabiendo que* no aporte ninguna información. De este modo introducimos el concepto de independencia de dos sucesos A y B como:

$$A \text{ es independiente de } B \Longleftrightarrow P(A\cap B) = P(A)P(B).$$

Esta relación puede ser escrita de modo equivalente: dados dos sucesos de probabilidad no nula (de manera que $P(A)\neq 0\neq P(B)$) diremos que A es independiente de B si y sólo si $P(A)=P(A|B)$ ó equivalentemente $P(B)=P(B|A)$.

Así, se dice que dos experimentos son independientes si los resultados de uno son independientes de los del otro, para cualquier par de resultados que se escoja. Las definiciones se extienden, de forma inmediata, al caso de independencia de tres o más sucesos o experimentos.

Hay algunos resultados importantes del cálculo de probabilidades que son conocidos bajo los nombres de *teorema de la probabilidad compuesta*, *teorema de la probabilidad total*

y *teorema de Bayes*. Antes de mostrarlos, se enuncian, a modo de recopilación, una serie de resultados elementales:

Proposición

Sean $A, B \subset E$ no necesariamente disjuntos. Se verifican entonces las siguientes propiedades:

– Probabilidad de la unión de sucesos:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

– Probabilidad de la intersección de sucesos:

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

– Probabilidad del suceso contrario:

$$P(A^c) = 1 - P(A).$$

– Probabilidad condicionada del suceso contrario:

$$P(A^c|B) = 1 - P(A|B).$$

Teorema (Probabilidad compuesta)

Sea $A_1, A_1, \ldots, A_n \subset E$ una colección de sucesos aleatorios, entonces

$$P(A_1A_2 \ldots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1A_2 \ldots A_{n-1}).$$

Demostración:

$$\begin{aligned} P(A_1A_2 \ldots A_n) &= P\left((A_1A_2 \ldots A_{n-1}) \cap A_n\right) = \\ &= P(A_1A_2 \ldots A_{n-1}) \cdot P(A_n|A_1A_2 \ldots A_{n-1}) = \\ &= P(A_1A_2 \ldots A_{n-2}) \cdot P(A_{n-1}|A_1A_2 \ldots A_{n-2}) \cdot P(A_n|A_1A_2 \ldots A_{n-1}) = \cdots = \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1A_2 \ldots A_{n-1}). \end{aligned}$$

Los teoremas que restan nos dicen cómo calcular las probabilidades de sucesos cuando tenemos que el suceso seguro está descompuesto en una serie de sucesos incompatibles de

los que conocemos su probabilidad. Para ello necesitamos introducir un nuevo concepto: se dice que la colección $A_1, A_1, \ldots, A_n \subset E$ es un sistema exhaustivo y excluyente de sucesos si se verifica:

$$\begin{aligned} \bigcup_{i=1}^{n} A_i &= E \\ A_i \cap A_j &= \emptyset, \forall i \neq j \end{aligned}$$

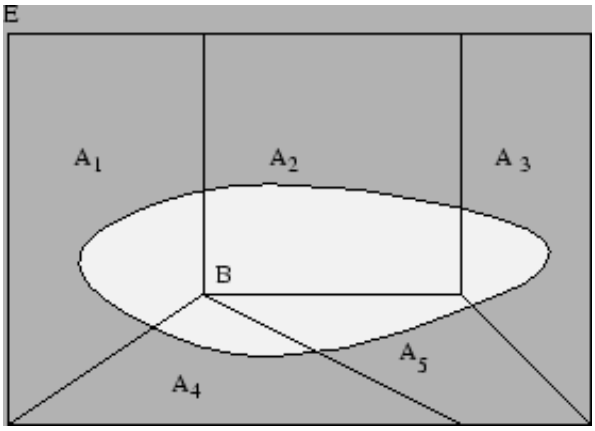
Teorema de la Probabilidad total

Sea $A_1, A_1, \ldots, A_n \subset E$ un sistema exhaustivo y mutuamente excluyente de sucesos. Entonces, $\forall B \subset E$, se verifica que

$$P(B) = \sum_{i=1}^{n} P\left(B|A_i\right) P(A_i).$$

Demostración:

Observando la figura



se deduce que los sucesos A_i forman un sistema exhaustivo y excluyente de sucesos, y se puede calcular la probabilidad de B a partir de las cantidades $P\left(B \cap A_i\right)$, o lo que es lo mismo, $P\left(B|A_i\right) \cdot P\left(A_i\right)$:

$$\begin{aligned} P(B) &= P(B \cap E) = P\left(B \cap \left(\bigcup_{i=1}^{n} A_i\right)\right) = \\ &= P\left(\bigcup_{i=1}^{n} (B \cap A_i)\right) = \sum_{i=1}^{n} P\left(B \cap A_i\right) = \sum_{i=1}^{n} P\left(B|A_i\right) \cdot P\left(A_i\right). \end{aligned}$$

Ejemplo

Se tienen dos urnas, y cada una de ellas contiene un número diferente de bolas blancas y rojas:

- Primera urna, U_1: 3 bolas blancas y 2 rojas;
- Segunda urna, U_2: 4 bolas blancas y 2 rojas.

Se realiza el siguiente experimento aleatorio: Se tira una moneda al aire y si sale cara se elige una bola de la primera urna, y si sale cruz de la segunda. ¿Cuál es la probabilidad de que salga una bola blanca?

Solución: La situación que tenemos puede ser esquematizada como

$3B$	$2R$	U_1	$P(U_1) = 1/2$	$P(B\|U_1) = 3/5$
$4B$	$2R$	U_2	$P(U_2) = 1/2$	$P(B\|U_2) = 4/6$

Como U_1 y U_2 forman un sistema incompatible y excluyente de sucesos (la bola resultado debe provenir de una de esas dos urnas y de una sólo de ellas), el teorema de la probabilidad total permite afirmar entonces que

$$P(B) = P(B|U_1) \cdot P(U_1) + P(B|U_2) \cdot P(U_2) = \frac{3}{5} \cdot \frac{1}{2} + \frac{4}{6} \cdot \frac{1}{2} = \frac{19}{30}.$$

Teorema de Bayes

Sea $A_1, A_1, \ldots, A_n \subset E$ un sistema exhaustivo y excluyente de sucesos. Sea $B \subset E$ un suceso del que conocemos las siguientes probabilidades: $P(B|A_i)$ para todo $i = 1, \ldots n$, a las que denominamos verosimilitudes, entonces se verifica, para todo $j = 1, \ldots n$,

$$P\left(A_j|B\right) = \frac{P\left(B|A_j\right) \cdot P\left(A_j\right)}{\sum_{i=1}^{n} P\left(B|A_i\right) \cdot P\left(A_i\right)}.$$

Demostración:

Es una consecuencia de la definición de probabilidad condicionada en términos de la intersección, y del teorema de la probabilidad total:

$$P\left(A_j|B\right) = \frac{P\left(A_j \cap B\right)}{P\left(B\right)} = \frac{P\left(B|A_j\right) \cdot P\left(A_j\right)}{\sum_{i=1}^{n} P\left(B|A_i\right) \cdot P\left(A_i\right)}. \quad \square$$

Ejemplo

Se tienen tres urnas. Cada una de ellas contiene un número diferente de bolas blancas y rojas:

- Primera urna, U_1: 3 bolas blancas y 2 rojas;
- Segunda urna, U_2: 4 bolas blancas y 2 rojas;
- Tercera urna, U_3: 3 bolas rojas.

Se realiza el siguiente experimento aleatorio: Alguien elige al azar y con la misma probabilidad una de las tres urnas, y saca una bola.

Si el resultado del experimento es que ha salido una bola blanca, ¿cuál es la probabilidad de que provenga de la primera urna? Calcular lo mismo para las otras dos urnas.

Solución: Vamos a representar en un esquema los datos de que disponemos:

$3B$	$2R$	U_1	$P(U_1)=1/3$	$P(B\|U_1)=3/5$
$4B$	$2R$	U_2	$P(U_2)=1/3$	$P(B\|U_2)=4/6$
$0B$	$3R$	U_3	$P(U_3)=1/3$	$P(B\|U_3)=0$

En este caso U_1, U_2 y U_3 forman un sistema incompatible y excluyente de sucesos (la bola resultado debe provenir de una de esas tres urnas y de una sólo de ellas), por tanto es posible aplicar el teorema de Bayes:

$$\begin{aligned} P\left(U_1|B\right) &= \frac{P\left(B|U_1\right)\cdot P\left(U_1\right)}{P\left(B|U_1\right)\cdot P\left(U_1\right)+P\left(B|U_2\right)\cdot P\left(U_2\right)+P\left(B|U_3\right)\cdot P\left(U_3\right)} = \\ &= \frac{\frac{3}{5}\cdot\frac{1}{3}}{\frac{3}{5}\cdot\frac{1}{3}+\frac{4}{6}\cdot\frac{1}{3}+0\cdot\frac{1}{3}} = \frac{9}{19}. \end{aligned}$$

Con respecto a las demás urnas es lo mismo:

$$\begin{aligned} P\left(U_2|B\right) &= \frac{P\left(B|U_2\right)\cdot P\left(U_2\right)}{P\left(B|U_1\right)\cdot P\left(U_1\right)+P\left(B|U_2\right)\cdot P\left(U_2\right)+P\left(B|U_3\right)\cdot P\left(U_3\right)} = \\ &= \frac{\frac{4}{6}\cdot\frac{1}{3}}{\frac{3}{5}\cdot\frac{1}{3}+\frac{4}{6}\cdot\frac{1}{3}+0\cdot\frac{1}{3}} = \frac{10}{19}. \end{aligned}$$

$$P(U_3|B) = \frac{P(B|U_3)\cdot P(U_3)}{P(B|U_1)\cdot P(U_1)+P(B|U_2)\cdot P(U_2)+P(B|U_3)\cdot P(U_3)} = \\ = \frac{0\cdot\frac{1}{3}}{\frac{3}{5}\cdot\frac{1}{3}+\frac{4}{6}\cdot\frac{1}{3}+0\cdot\frac{1}{3}} = 0.$$

Variables Aleatorias

Introducción

Normalmente, los resultados posibles (espacio muestral E) de un experimento aleatorio no son valores numéricos. Por ejemplo, si un experimento consiste en lanzar de modo ordenado tres monedas al aire, para observar el número de caras ($\mathcal{C}$) y cruces ($\mathcal{R}$) que se obtienen, el espacio muestral asociado a dicho experimento aleatorio sería:

$$E = \{\mathcal{CCC}, \mathcal{CCR}, \mathcal{CRC}, \mathcal{CRR}, \mathcal{RCC}, \mathcal{RCR}, \mathcal{RRC}, \mathcal{RRR}\}.$$

En Cálculo de Probabilidades resulta más fácil utilizar valores numéricos en lugar de trabajar directamente con los elementos de un espacio muestral como el anterior. Así, se prefiere identificar los sucesos $\{\mathcal{CRR}, \mathcal{RCR}, \mathcal{RRC}\}$ con el valor numérico 1 que representa el número de caras obtenidas al realizar el experimento.

De este modo, aparece el concepto de *variable aleatoria unidimensional* como el de una función

$$\begin{aligned} X \;:\; & E \longrightarrow \mathbb{R} \\ & e \longrightarrow X(e) \end{aligned}$$

que a cada suceso elemental e le atribuye un único número real del espacio muestral E.

En el ejemplo anterior, se puede definir la variable aleatoria $X \equiv$ *número de caras*, del siguiente modo:

$$X : E \longrightarrow \mathbb{R}$$

$$\begin{aligned} X(\mathcal{CCR}) &= X(\mathcal{CRC}) = X(\mathcal{RCC}) = 2 \\ X(\mathcal{RRC}) &= X(\mathcal{RCR}) = X(\mathcal{CRR}) = 1 \\ X(\mathcal{RRR}) &= 0 \end{aligned}$$

Observación: La variable X no recibe el calificativo de aleatoria por el hecho de que atribuya de modo imprevisible un valor cualquiera a un elemento $e \in E$ ya que este valor está definido de forma precisa (determinística). Lo que es aleatorio, en realidad, es que al realizar el experimento, no sabemos qué elemento de E puede ocurrir.

La composición de una función real con una variable aleatoria es también otra variable aleatoria, pues está definida sobre E y a cada elemento suyo le asocia un valor real.

$$\left.\begin{array}{c} X : E \longrightarrow \mathbb{R} \\ h : \mathbb{R} \longrightarrow \mathbb{R} \end{array}\right\} \implies h(X) = h \circ X : E \longrightarrow \mathbb{R}$$
$$x \longrightarrow h(X(e))$$

En función de los valores que tome la variable, ésta puede ser clasificada en discreta o continua del siguiente modo:

Variable aleatoria discreta: es aquella que sólo puede tomar un número finito o infinito numerable de valores. Por ejemplo,

$$X : E \longrightarrow \mathbb{N}$$

Variable aleatoria continua: es la que puede tomar un número infinito no numerable de valores.

$$X : E \longrightarrow \mathbb{R}$$

Observación: Si sobre los elementos de E existe una distribución de probabilidad, ésta se transmite a los valores que toma la variable X. Es decir, toda v.a. conserva la estructura probabilística del experimento aleatorio que describe, en el sentido de que si $\mathcal{P}$ es la función de probabilidad definida sobre el espacio muestral E, ésta induce otra función $\mathcal{P}^*$ definida sobre $\mathbb{R}$, de forma que conserva los valores de las probabilidades.

$$\begin{aligned} \mathcal{P}^*[X = x] &= \mathcal{P}[\{e \in E : X(e) = x\}] \\ \mathcal{P}^*[X \in (a,b)] &= \mathcal{P}[\{e \in E : X(e) \in (a,b)\}] \end{aligned}$$

De ahora en adelante omitiremos el asterisco, para simplificar la notación, y no diferenciaremos entre las probabilidades calculadas sobre el espacio muestral del experimento aleatorio original, E, y las calculadas sobre $\mathbb{R}$.

Vamos a estudiar los conceptos más importantes relacionados con la distribución de probabilidad de una v.a., diferenciando entre los casos de v.a. discreta y v.a. continua.

Definición

Dada la v.a. X sobre el espacio E, asociado a un experimento aleatorio, se define su distribución de probabilidad como

$$P(X \in A) = P\{e \in E : X(e) \in A\}$$

Variables aleatorias discretas

Dada una v.a. discreta $X : E \longrightarrow \mathbb{N}$, f, se define su función de masa de probabilidad , de modo que $f(x_i)$ es la probabilidad de que X tome ese valor concreto:

$$\begin{aligned} f: \quad & \mathbb{N} \longrightarrow [0,1] \\ & x_i \longrightarrow f(x_i) = P[X = x_i] = P[\{e, \text{ tal que } X(e) = x_i\}] \end{aligned}$$

Si x_i no es uno de los valores que puede tomar X, entonces $f(x_i) = 0$. La representación gráfica de la función de masa de probabilidad se realiza mediante un diagrama de barras análogo al de distribución de frecuencias relativas para variables discretas en Estadística Descriptiva.

Por ejemplo, si se considera el caso del lanzamiento de 3 monedas de forma que cada una de ellas tenga probabilidad 1/2 de obtener el resultado de cara o cruz, se tiene que

$$\begin{aligned} f(3) &= P[X = 3] = P[\{\mathcal{CCC}\}] = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} \\ f(2) &= P[X = 2] = P[\{\mathcal{RCC}, \mathcal{CCR}, \mathcal{CRC}\}] = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} \\ f(1) &= P[X = 1] = P[\{\mathcal{RRC}, \mathcal{RCR}, \mathcal{CRR}\}] = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} \\ f(0) &= P[X = 0] = P[\{\mathcal{RRR}\}] = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} \end{aligned}$$

Observación: Obsérvese que X está definido sobre el espacio muestral de sucesos E, mientras que f lo está sobre el espacio de números reales $\mathbb{R}$.

Las propiedades de la función de masa de probabilidad de v.a. se deducen de forma inmediata de los axiomas de probabilidad:

Proposición

Si $x_1, x_2, \ldots x_k$ son todos los valores admisibles de la v.a. X, entonces la suma de las probabilidades de todos los posibles valores de la variable es 1.

$$\sum_{i=1}^{k} f(x_i) = \sum_{i=1}^{k} P\left[X = x_i\right] = 1$$

y además todas las probabilidades son no negativas:

$$\begin{aligned} f(x_i) &\geq 0 \\ \forall i &= 1, \ldots, k \end{aligned}$$

Es evidente que si tenemos tres constantes $a < b < c$, los sucesos $A = \{a \leq X \leq b\}$ y $B = \{b < X \leq c\}$ son mutuamente excluyentes, es decir, $A \cap B = \emptyset$, luego $P\left[A \cap B\right] = 0$. Por ello, si se define $C = \{a \leq X \leq c\} = A \cup B$, se tiene que

$$\begin{aligned} P(C) &= P(A) + P(B) \Longleftrightarrow \\ P\left[a \leq X \leq c\right] &= P\left[a \leq X \leq b\right] + P\left[b < X \leq c\right] \end{aligned}$$

Otro concepto importante es el de **función de distribución** de una variable aleatoria discreta, F, que se define de modo que si $x_i \in \mathbb{R}$, $F(x_i)$ es igual a la probabilidad de que X tome un valor inferior o igual a x_i:

$$\begin{aligned} F: \quad & N \longrightarrow [0,1] \\ & x_i \longrightarrow F\left(x_i\right) = P\left[X \leq x_i\right] = P\left[\{e, \text{ tal que } X(e) \leq x_i\}\right] \end{aligned}$$

Esta función se representa gráficamente del mismo modo que la distribución de frecuencias relativas acumuladas. Volviendo al ejemplo de las tres monedas, se tiene que

$$\begin{aligned} F(0) &= P\left[X \leq 0\right] = P\left[X = 0\right] = f(0) = \frac{1}{8} \\ F(1) &= P\left[X \leq 1\right] = f(0) + f(1) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8} \\ F(2) &= P\left[X \leq 2\right] = f(0) + f(1) + f(2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8} \\ F(3) &= P\left[X \leq 3\right] = f(0) + f(1) + f(2) + f(3) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{8}{8} = 1 \end{aligned}$$

Se observa que a valores no admisibles de la variable aleatoria les pueden corresponder valores de F no nulos. Por ejemplo,

$$F\left(-1\right) = P\left[X \leq -1\right] = P\left[\emptyset\right] = 0$$

Es sencillo comprobar las siguientes propiedades de la función de distribución:

Proposición

La función de distribución F, es una función *monótona no decreciente*, es decir,

$$x_1 < x_2 \Longrightarrow F(x_1) \leq F(x_2)$$

Además, es continua a la derecha

$$\lim_{x \to a^+} F(x) = F(a)$$

y

$$\begin{aligned} F(-\infty) &= \lim_{x_i \to -\infty} F(x_i) = 0 \\ F(+\infty) &= \lim_{x_i \to +\infty} F(x_i) = 1 \end{aligned}$$

Nota: De lo anterior se deduce que

$$\begin{aligned} P\left(X > x\right) &= 1 - F(x) \\ P\left(x_1 < X \leq x_2\right) &= F(x_2) - F(x_1). \end{aligned}$$

Momentos de una variable aleatoria discreta

En ocasiones se está interesado en resumir una variable aleatoria, típicamente a través de sus momentos. Los principales son:

- La esperanza o media de la variable aleatoria X

$$E\left[X\right] = \mu = \sum_i x_i f(x_i),$$

que constituye una medida de localización.

Se puede observar en esta medida, así como en las que siguen, el paralelismo con los momentos muestrales, aunque no se deben confundir.

- La varianza

$$\sigma^2 = \sum_i \left(x_i - \mu\right)^2 f(x_i),$$

que constituye una medida de dispersión. Alternativamente, se tiene que

$$\begin{aligned} \sigma^2 &= \sum_i \left(x_i - \mu\right)^2 f(x_i) = \sum_i \left(x_i^2 - 2\mu x_i + \mu^2\right) f(x_i) = \\ &= \sum_i x_i^2 f(x_i) - \mu^2. \end{aligned}$$

- La desviación típica

$$\sigma = \sqrt{\sum_i (x_i - \mu)^2 f(x_i)}.$$

- La mediana $\widetilde{x}$ definida como el valor tal que

$$\begin{aligned} P(X \leq \widetilde{x}) &\geq 0{,}5 \\ P(X \geq \widetilde{x}) &\geq 0{,}5, \end{aligned}$$

que constituye otra medida de centralización.

Distribución de Bernoulli

Consiste en realizar un experimento aleatorio una sola vez y observar si cierto suceso ocurre o no, siendo p la probabilidad de *éxito* y $q = (1-p)$ de *fracaso* . En realidad, se trata de una variable dicotómica, que únicamente puede tomar dos valores. Podríamos, por tanto, definir este experimento mediante una v.a. discreta X que toma los valores $X = 0$ si el suceso no ocurre, y $X = 1$ en caso contrario, denotándose $X \sim Ber(p)$ si

$$X = \begin{cases} 0 \longrightarrow q = (1-p) = P(X=0) \\ \\ 1 \longrightarrow p = P(X=1) \end{cases}$$

Por ejemplo, se lanza una moneda y se considera la v.a. $X =$ *número de caras obtenidas,* $X = 1$ con $p = \frac{1}{2}$ y $X = 0$ con $q = \frac{1}{2}$.

Para una v.a. de *Bernouilli*, su función de masa de probabilidad es:

$$\begin{aligned} P(X=1) &= p \\ P(X=0) &= 1-p, \end{aligned}$$

y 0 en el resto.

Su función de distribución es

$$F(x) = \begin{cases} 0 \text{ si } x < 0 \\ q \text{ si } 0 \leq x < 1 \\ 1 \text{ si } x \geq 1 \end{cases}$$

Los principales momentos de la X son

$$\begin{aligned} E[X] &= 0 \cdot P(X=0) + 1 \cdot P(X=1) = p \\ E\left[X^2\right] &= 0^2 \cdot P(X=0) + 1^2 \cdot P(X=1) = p \\ Var[X] &= p - p^2 = p(1-p) = pq. \end{aligned}$$

Distribución binomial

Se utiliza la distribución binomial para modelizar el número de veces que se da un resultado al realizar varias pruebas idénticas e independientes de un experimento con dos resultados posibles.

Las hipótesis específicas que se hacen son:

- Se consideran n repeticiones *independientes* de un experimento.
- El experimento tiene dos resultados posibles: *éxito* o *fracaso*.
- La probabilidad p de éxito es la misma en cada repetición (o ensayo).

El espacio muestral del experimento es el conjunto E de n-uplas (éxito o fracaso). La variable X de interés es el número de éxitos obtenidos en esas n pruebas.

Entonces, se dice que X tiene distribución *binomial* de parámetros n y p, y se escribe como $X \sim Bin\,(n,p)$, siendo la distribución

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

Ejemplo Hay una probabilidad de 0.05 de que un dispositivo falle, bajo condiciones intensas de uso, en un día. Una central hidroeléctrica tiene 16 dispositivos operando en condiciones similares. ¿Cuál es la probabilidad de que a lo sumo fallen 2 dispositivos? ¿y que al menos fallen 4?

Solución:

Si X designa el número de fallos de los dispositivos en un día bajo condiciones intensas de uso, y se puede suponer que $X \sim Bin\,(16, 0{,}05)$.

Nos piden primero

$$P\,(X \leq 2) = \binom{16}{0} 0{,}05^0 0{,}95^{16} + \binom{16}{1} 0{,}05^1 0{,}95^{15} + \binom{16}{2} 0{,}05^2 0{,}95^{14} = 0{,}957\,06$$

Después,

$$\begin{aligned} & P\,(X \geq 4) = 1 - P\,(X \leq 3) = \\ = \;& 1 - \left(\binom{16}{0} 0{,}05^0 0{,}95^{16} + \binom{16}{1} 0{,}05^1 0{,}95^{15} + \binom{16}{2} 0{,}05^2 0{,}95^{14} + \binom{16}{3} 0{,}05^3 0{,}95^{13} \right) = \\ = \;& 0{,}00701 \end{aligned}$$

Proposición

La media y la varianza de una $Bin(n,p)$ es

$$\begin{aligned} E[X] &= \mu = np \\ Var[X] &= \sigma^2 = np(1-p). \end{aligned}$$

Demostración:

Como la esperanza o media de una variable aleatoria X es

$$E[X] = \mu = \sum_i x_i f(x_i),$$

entonces, en este caso,

$$E[X] = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} =$$

$$np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} = np \left[\sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y (1-p)^{n-1-y} \right] =$$

$$= \quad np.$$

En el caso de la varianza, se tiene que

$$\sigma^2 = \sum_i x_i^2 f(x_i) - \mu^2 \quad \text{y} \quad \sum_i x_i^2 f(x_i) = \sum_i x_i (x_i - 1) f(x_i) + \sum_i x_i f(x_i).$$

Aplicando esto al caso de la binomial, se tiene que

$$\sum_{x=0}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \sum_{x=2}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} =$$

$$= \quad n(n-1) p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} (1-p)^{(n-2)-(x-2)} = n(n-1) p^2.$$

Así,

$$\sigma^2 = n(n-1) p^2 + np - n^2 p^2 = np(1-p).$$

Distribución de Poisson

Como primera motivación, estamos interesados en contar el número de éxitos en n pruebas del tipo éxito o fracaso cuando n es grande y el suceso es raro (esto es, la probabilidad de éxito es pequeña).

Específicamente, supongamos que $X \sim Bin(n,p)$, esto es,

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$$

con $n \longrightarrow \infty$, $np = \lambda$, esto es, $p = \frac{\lambda}{n} \longrightarrow 0$. Sustituyendo, se tiene

$$\begin{aligned} \frac{n!}{x!\,(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} &= \frac{n\,(n-1)\cdots(n-x+1)}{n\cdot n\cdots n}\frac{1}{x!}\lambda^x \left(1-\frac{\lambda}{n}\right)^{n-x} = \\ &= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}{x!}\lambda^x \left(1-\frac{\lambda}{n}\right)^{n-x}, \end{aligned}$$

Por otro lado, se tiene que

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right) \underset{n\to\infty}{\longrightarrow} 1$$

$$\left(1-\frac{\lambda}{n}\right)^{n-x} = \left(\left(1-\frac{\lambda}{n}\right)^{\frac{n}{\lambda}}\right)^{\lambda}\left(1-\frac{\lambda}{n}\right)^{-x} \underset{n\to\infty}{\longrightarrow} e^{-\lambda}.$$

De este modo, si $X \sim Bin(n,p)$, siendo $np = \lambda$,

$$P(X=x) \underset{n\to\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^x}{x!},$$

$x = 0, 1, 2, \ldots$

Proposición

Se dice que una variable aleatoria X tiene una distribución *de Poisson* de parámetro λ, $X \sim Po(\lambda)$ si

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!},$$

donde $k = 0, 1, 2, \ldots$

Observación: Se observa que

$$\sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}\cdot e^{\lambda} = 1.$$

Proposición

Si $X \sim Po(\lambda)$ entonces

$$\begin{aligned} E[X] &= \sum_{x=0}^{\infty} x\cdot\frac{\lambda^x e^{-\lambda}}{x!} = \lambda\sum_{x=1}^{\infty}\frac{\lambda^{x-1}e^{-\lambda}}{(x-1)!} = \\ &= \lambda e^{-\lambda}\sum_{y=0}^{\infty}\frac{\lambda^y}{y!} = \lambda\cdot e^{-\lambda}\cdot e^{\lambda} = \lambda. \end{aligned}$$

En forma similar se calcula

$$Var\left[X\right] = \lambda.$$

Ejemplo El 5 % de los objetos que produce una línea de producción sale con defectos. Calcular la probabilidad de que 2 de 100 objetos tengan defectos.
Solución:

Si X designa el número de objetos defectuosos de 100, se tiene que $X \sim Bin\left(100, 0{,}05\right)$. Se pide

$$P\left(X = 2\right) = \binom{100}{2} 0{,}05^2 \cdot 0{,}95^{98} = 0{,}081$$

Alternativamente, como n es grande y p es pequeño, y aplicamos la aproximación de Poisson, se tiene que $\lambda = np = 100 \cdot 0{,}05 = 5$, y

$$P\left(X = 2\right) \approx e^{-5}\frac{5^2}{2!} = 0{,}084$$

Observación: La aproximación de la distribución de Poisson a la binomial es buena cuando $n \geq 20$ y $p \leq 0{,}05$ y es muy buena cuando $n \geq 100$ y $np \leq 10$.

Variables aleatorias continuas

Se presentan ahora los conceptos básicos sobre variables aleatorias continuas. Se repiten las mismas ideas que en el caso discreto, con *funciones de densidad* en lugar de *funciones de masa*, e *integrales* en lugar de *sumas*.

En este caso, lo que generaliza de modo natural el concepto de suma es el de integral. Por otro lado, para variables continuas no tiene interés hablar de la probabilidad de que $X = x \in \mathbb{R}$, ya que ésta debe de valer siempre 0, para que la suma infinita no numerable de las probabilidades de todos los valores de la variable no sea infinita.

De este modo, es necesario introducir un nuevo concepto que sustituya en v.a. continuas, al de función de probabilidad de una v.a. discreta.

Este concepto es el de **función de densidad** de una v.a. continua, que se define como una función $f : \mathbb{R} \longrightarrow \mathbb{R}$, integrable no negativa sobre la recta real, que verifica las dos

propiedades siguientes:

$$\begin{aligned} f(x) &\geq 0, \\ \int_{-\infty}^{+\infty} f(x)dx &= 1, \\ P\left(X \in A\right) &= \int_A f(x)dx. \end{aligned}$$

Así, dados $a < b$, se tiene que

$$P\left[a \leq X \leq b\right] = \int_a^b f(x)dx.$$

Se puede observar que, por ser f una función integrable, la probabilidad de un punto es nula:

$$P\left[X = a\right] = P\left[a \leq X \leq a\right] = \int_a^a f(x)dx = 0,$$

y por ello, calcular la probabilidad de un intervalo no afecta nada el que éste sea abierto o cerrado por cualquiera de sus extremos, pues estos son puntos y por tanto de probabilidad nula:

$$P\left[a \leq X \leq b\right] = P\left[a \leq X < b\right] = P\left[a < X \leq b\right] = P\left[a < X < b\right]$$

Función de distribución

La función de distribución se define de manera que dado $x \in \mathbb{R}$, $F(x)$ es la probabilidad de que X sea menor o igual que x, es decir

$$F(x) = P\left(X \leq x\right) = \int_{-\infty}^{x} f(t)dt.$$

Se verifican las propiedades ya mencionadas en el caso discreto:

(i) $P\left[a < X \leq b\right] = F(b) - F(a)$

(ii) F es monótona no decreciente

(iii) $F\left(-\infty\right) = 0$

(iv) $F\left(+\infty\right) = 1$

(v) F es continua por la derecha.

Además, por el teorema fundamental del Cálculo

$$\frac{dF(x)}{dx} = f(x),$$

si existe la derivada.

Momentos

Como en el caso discreto, en ocasiones se está interesado en resumir la distribución de X a través de sus momentos.

La esperanza de una variable X es

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

y la varianza es

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

Ejemplo Para la v.a. del ejemplo anterior se tiene que

$$\begin{aligned} \mu &= \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{\infty} x 2e^{-2x} dx = \frac{1}{2}, \\ \sigma^2 &= \int_{-\infty}^{\infty} x^2 2e^{-2x} dx - \frac{1}{4} = \frac{1}{4}. \end{aligned}$$

Como en el caso de las variables discretas, se muestran a continuación las distribuciones continuas más importantes en Ciencias.

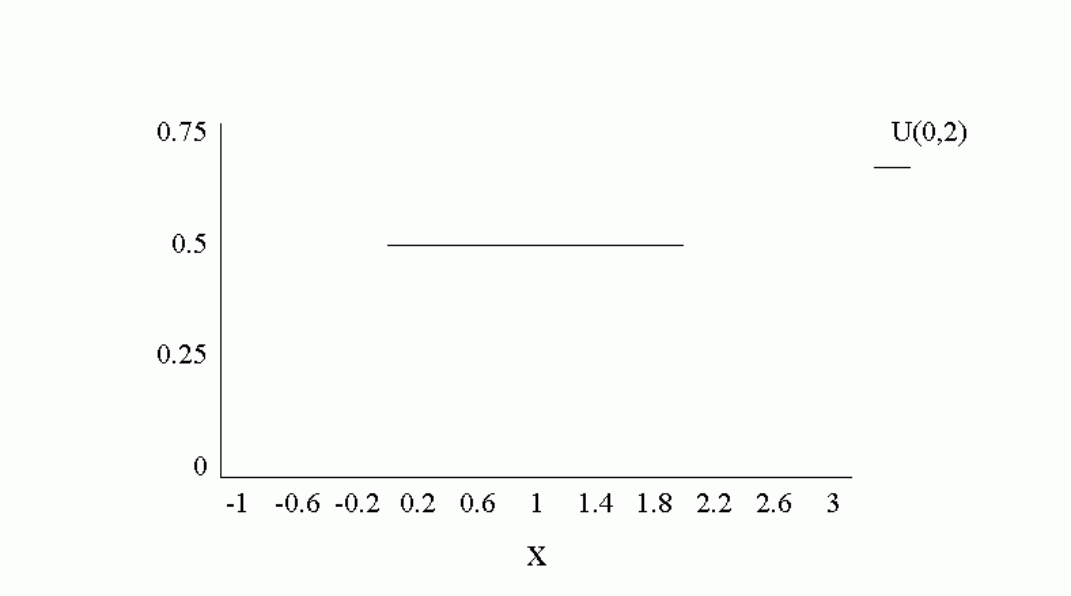
La distribución uniforme

Muchos fenómenos pueden asociarse a situaciones de incertidumbre uniforme.

Definición

Una v.a. tiene distribución uniforme en (a, b) (y se representa como $X \sim U(a, b)$) si su función de densidad es

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ \\ 0 \quad \text{resto} \end{cases}$$



Su función de distribución

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

Con esta ley de probabilidad, la probabilidad de que al hacer un experimento aleatorio el valor de X esté comprendido en un cierto subintervalo de (a, b) depende únicamente de la longitud del mismo, no de su posición. Cometiendo un abuso en el lenguaje, se puede decir que en una distribución uniforme la probabilidad de todos los puntos del soporte es la misma.

Sus momentos son

$$\begin{aligned} \mu &= \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{b^2-a^2}{2} = \frac{a+b}{2}, \\ E\left(X^2\right) &= \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \frac{b^3-a^3}{3} = \frac{1}{3}\left(b^2+ab+a^2\right). \end{aligned}$$

Resulta así que

$$\sigma^2 = \frac{1}{3}\left(b^2+ab+a^2\right) - \left(\frac{a+b}{2}\right)^2 = \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}.$$

La mediana es

$$\widetilde{x} = \frac{a+b}{2}.$$

La distribución exponencial

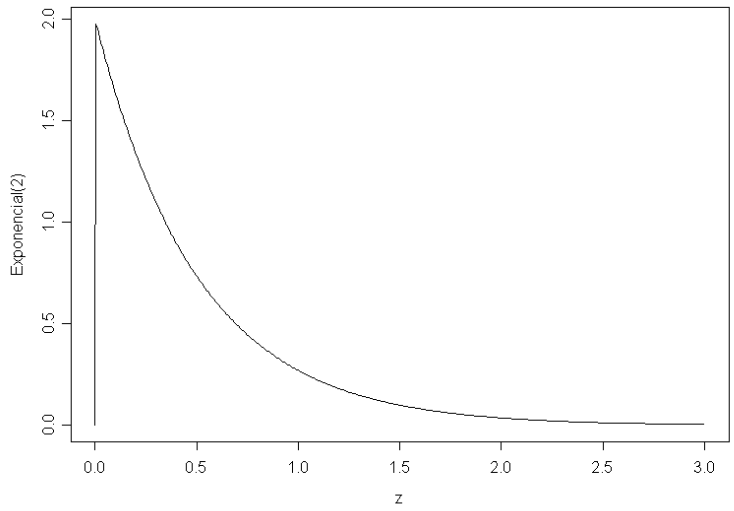
La distribución exponencial describe procesos en los que interesa saber el tiempo hasta que ocurre determinado evento, sabiendo que el proceso no tiene *memoria* (no depende de lo que ha ocurrido anteriormente).

Definición

Una v.a. tiene distribución exponencial de parámetro α (y se representa como $X \sim Exp(\alpha)$), si su función de densidad es

$$f(x|\alpha) = \begin{cases} \alpha e^{-\alpha x} & \text{para} \quad x > 0 \\ 0 & \text{resto} \end{cases}$$

con $\alpha > 0$.



Se comprueba fácilmente que los momentos de la exponencial son

$$\begin{aligned} \mu &= \frac{1}{\alpha} \\ Var(X) &= \frac{1}{\alpha^2} \end{aligned}$$

y que su función de distribución es

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\alpha x} & \text{para} \quad x > 0 \end{cases}$$

Observación: Una propiedad importante, y característica de esta distribución, es su *pérdida de memoria*. Se tiene que

$$\begin{aligned} P\left(T > t+h \middle| T > t\right) &= \frac{P\left(T > t+h, T > t\right)}{P\left(T > t\right)} = \frac{P\left(T > t+h\right)}{P\left(T > t\right)} = \\ &= \frac{e^{-\alpha(t+h)}}{e^{-\alpha t}} = e^{-\alpha h} = P\left(T > h\right). \end{aligned}$$

Esta propiedad hace útil a la exponencial en algunas aplicaciones de *Fiabilidad de Sistemas*.

La distribución gamma

La distribución exponencial, en realidad, está incluída dentro de una familia más amplia: la distribución gamma.

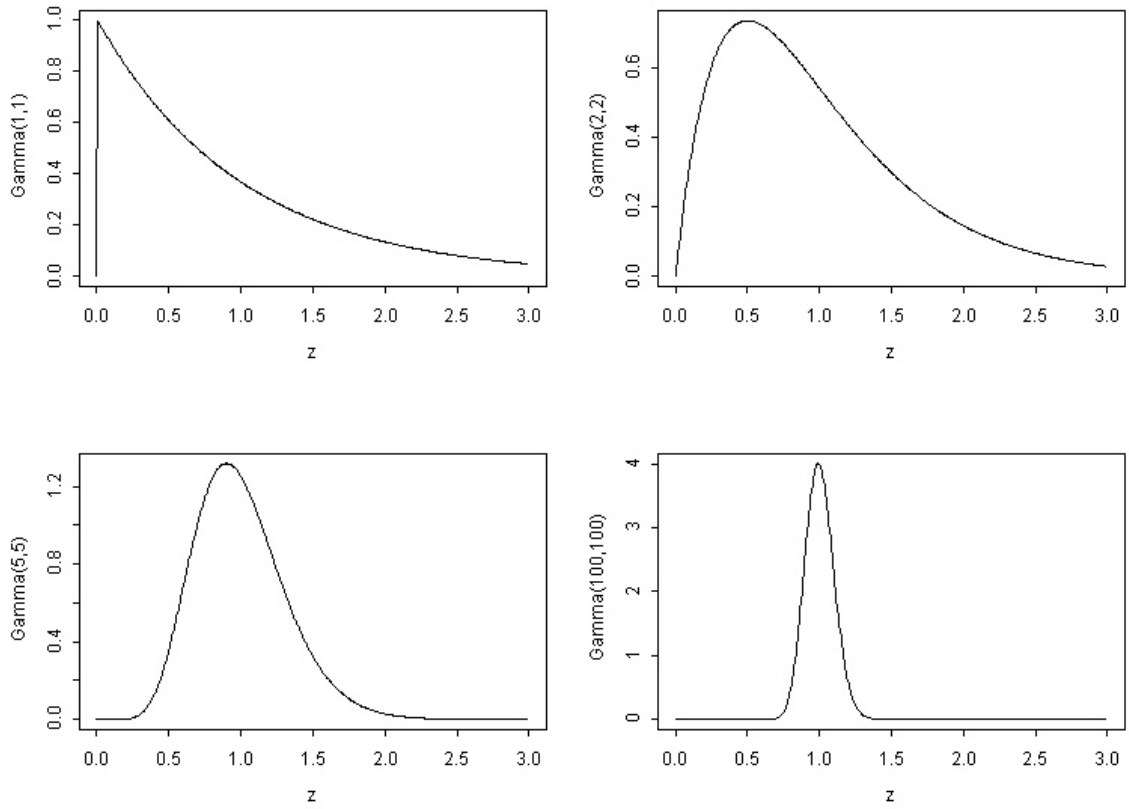
Se dice que una v.a. tiene una distribución gamma de parámetro de forma $p > 0$ y parámetro de escala $s > 0$, si su función de densidad es

$$f(x; a, p) = \frac{1}{s^p \Gamma(p)} \exp\left[-\frac{x}{s}\right] \cdot x^{p-1}$$

donde $x > 0$.

Los momentos de la distribución gamma son

$$\begin{aligned} \mu &= sp \\ Var\left(X\right) &= s^2 p \end{aligned}$$



Una propiedad importante es la siguiente:

Propiedad

Si $X_1, X_2, \ldots, X_n$ son v.a. independientes $\gamma(p_1, s), \gamma(p_2, s), \ldots, \gamma(p_n, s)$ respectivamente entonces la suma de variables

$$T = X_1 + X_2 + \ldots + X_n$$

se distribuye como otra distribución gamma

$$\gamma\left(\sum_{i=1}^{n} p_i,\ s\right).$$

La distribución normal

La distribución normal ocupa un lugar central en la Estadística. Su origen está en el descubrimiento de regularidades en los errores de medición por parte del matemático

Gauss: los patrones que se observan se aproximan por la denominada *curva normal de errores*.

En Ciencias tendrá importancia por tres razones básicas:

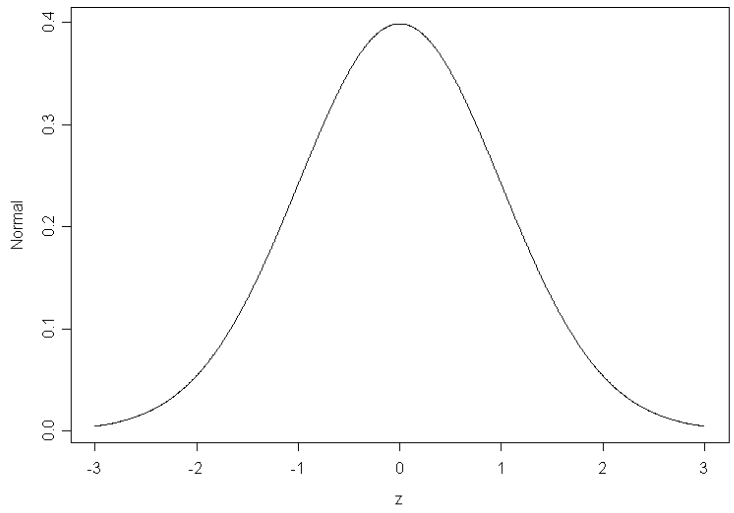
- Estaremos interesados en mediciones de diversas magnitudes, sometidas a errores que, típicamente, modelizaremos como una distribución normal.
- Por sus propiedades estadísticas y probabilísticas, los cálculos de interés de carácter estadístico son relativamente sencillos.
- El *Teorema Central del Límite* nos asegura que, para muestras grandes y bajo condiciones apropiadas, muchas funciones de las observaciones se aproximan, en distribución, mediante la normal

Definición

Una v.a. tiene distribución normal de parámetros μ y σ (y se representa como $X \sim N(\mu, \sigma)$), si su función de densidad es

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

donde $-\infty < x < \infty$.



Se comprueba fácilmente que

$$\int_{-\infty}^{\infty} f(x|\mu,\sigma)dx = 1$$
$$E(X) = \int_{-\infty}^{\infty} xf(x|\mu,\sigma)dx = \mu$$
$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x|\mu,\sigma)dx = \sigma^2,$$

esto es, el parámetro μ es la media y σ^2 es la varianza. Además la distribución es simétrica (respecto de μ) por lo que la mediana es también μ.

Se considera ahora, específicamente, la denominada distribución z normal estándar, con $\mu = 0$ y $\sigma = 1$, es decir,

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right),$$

donde $-\infty < z < \infty$.

Su función de distribución es

$$\phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

No es posible, sin embargo, la determinación de esta integral en forma concreta. Dada su importancia se dispone de tablas que permiten calcular $\phi(z)$ para cualquier valor de interés.

Para calcular probabilidades de distribuciones normales no estándar, se usa el hecho de que si $X \sim N(\mu,\sigma)$, entonces,

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1),$$

o, alternativamente, si $Z \sim N(0,1)$,

$$X = \mu + \sigma Z \sim N(\mu,\sigma).$$

Distribuciones Conjuntas

Se va a estudiar ahora el caso en el que se observan dos o más variables y/o se transforman de dos o más maneras las observaciones del experimento aleatorio básico. Primero se considera el caso discreto.

Distribuciones discretas multivariantes

Supongamos que se estudia los tipos de viajes contratados por una agencia de viajes, que pueden ser de tres tipos (0, 1, 2), bajo dos modos de operación (0, 1). Se dispone de las probabilidades de tipo (X_1), bajo el modo (X_2) de operación indicado, en un tiempo dado, que se expresa en la siguiente tabla.

		X_1			
		0	1	2	
X_2	0	0.1	0.4	0.1	0.6
	1	0.2	0.2	0	0.4
		0.3	0.6	0.1	1

Así, se tiene que $P(X_1 = 1, X_2 = 0) = 0{,}4$

La tabla define la **distribución conjunta** de (X_1, X_2) mediante su función de masa de probabilidad.

$$f(X_1, X_2) = P(X_1 = x_1, X_2 = x_2).$$

A partir de ella podemos definir diversas distribuciones de interés. Por ejemplo, se pueden definir:

- La distribución marginal de X_1 mediante

$$f_1(X_1) = P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2) = \sum_{x_2} f(X_1, X_2)$$

- La marginal de X_2 mediante

$$f_2(X_2) = P(X_2 = x_2) = \sum_{x_1} P(X_1 = x_1, X_2 = x_2) = \sum_{x_1} f(X_1, X_2)$$

- La distribución de X_2 condicionada a $X_1 = x_1$ mediante

$$\begin{aligned} f_2(X_2|X_1) &= P(X_2 = x_2|X_1 = x_1) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1)} = \\ &= \frac{f(X_1, X_2)}{f_1(X_1)} \end{aligned}$$

- La distribución de X_1 condicionada a $X_2 = x_2$ mediante

$$\begin{aligned} f_1(X_1|X_2) &= P(X_1 = x_1|X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)} = \\ &= \frac{f(X_1, X_2)}{f_2(X_2)} \end{aligned}$$

– Se dice que X_1 y X_2 son independientes si

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2),$$

para x_1 y x_2 cualesquiera, o equivalentemente,

$$f(X_1, X_2) = f_1(X_1) \cdot f_2(X_2).$$

Ejemplo En el ejemplo anterior se tiene

$$\begin{aligned} f_1(0) &= P(X_1 = 0) = P(X_1 = 0, X_2 = 0) + P(X_1 = 0, X_2 = 1) = \\ &= 0{,}1 + 0{,}2 = 0{,}3 \end{aligned}$$

$$f_2(0) = 0{,}1 + 0{,}4 + 0{,}1 = 0{,}6$$

$$f(0,0) = 0{,}1 \neq 0{,}3 \cdot 0{,}6 = f_1(0) \cdot f_2(0),$$

por lo que X_1 y X_2 no son independientes.

Estas ideas se pueden extender al caso k-variante. La distribución conjunta será

$$f(X_1, \ldots, X_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

y, por ejemplo, la distribución marginal de X_1 es

$$f_1(X_1) = \sum_{x_2, \ldots x_k} f(X_1, \ldots, X_n)$$

Distribuciones multivariantes continuas

Las ideas anteriores se pueden extender fácilmente al caso continuo. Para un vector aleatorio de variables aleatorias continuas $(X_1, \ldots, X_k)$ se tendrá una función de densidad multivariante $f(x_1, \ldots, x_k)$ que satisface

$$\begin{aligned} f(x_1, \ldots, x_k) &\geq 0 \\ \int \int \cdots \int_{\mathbb{R}^k} f(x_1, \ldots, x_k)\, dx_1 \ldots dx_k &= 1 \end{aligned}$$

y

$$P(a_1 \leq X_1 \leq b_1, \ldots, a_k \leq X_k \leq b_k) = \int_{a_k}^{b_k} \cdots \int_{a_1}^{b_1} f(x_1, \ldots, x_k)\, dx_1 \ldots dx_k$$

La función de distribución es

$$\begin{aligned} F(x_1,\ldots,x_k) &= P(X_1 \leq x_1,\ldots,X_k \leq x_k) = \\ &= \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_1} f(t_1,\ldots,t_k)\, dt_1 \ldots dt_k. \end{aligned}$$

Ejemplo Consideramos la función de densidad

$$f(x_1,x_2) = \begin{cases} 6e^{-2x_1}e^{-3x_2} & x_1 > 0, x_2 > 0 \\ 0 & \text{en el resto} \end{cases}$$

Se tiene

$$\int\int f(x_1,x_2)\, dx_2 dx_1 = \int_0^{\infty}\int_0^{\infty} 6e^{-2x_1}e^{-3x_2} dx_2 dx_1 = 1$$

por lo que se demuestra que es una función de densidad. Para calcular probabilidades se hace, por ejemplo,

$$\begin{aligned} P(1 \leq X_1 \leq 2, 2 \leq X_2 \leq 3) &= \int_2^3\int_1^2 6e^{-2x_1}e^{-3x_2} dx_1 dx_2 = \\ &= \left(e^{-4} - e^{-2}\right)\left(e^{-9} - e^{-6}\right) = 0{,}0003 \end{aligned}$$

$$\begin{aligned} P(0 \leq X_1 \leq 2, 2 \leq X_2 \leq \infty) &= \int_0^2\int_2^{\infty} 6e^{-2x_1}e^{-3x_2} dx_2 dx_1 = \\ &= \left(1 - e^{-4}\right)e^{-6} = 0{,}0024 \end{aligned}$$

La función de distribución es

$$\begin{aligned} F(x_1,x_2) &= \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} 6e^{-2t_1}e^{-3t_2} dt_2 dt_1 = \\ &\left(1 - e^{-2x_1}\right)\left(1 - e^{-3x_2}\right). \end{aligned}$$

La **marginal** de X_1 se calcula como

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1,\ldots,x_k)\, dx_k \ldots dx_2.$$

Se dice que $X_1,\ldots,X_k$ son **independientes** si

$$F(x_1,\ldots,x_k) = F_1(x_1) \cdot F_2(x_2) \ldots F_k(x_k)$$

o, equivalentemente,

$$f(x_1,\ldots,x_k) = f_1(x_1) \cdot f_2(x_2) \ldots f_k(x_k).$$

Ejemplo En el caso anterior, la marginal de X_1 es

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2 = \begin{cases} 0 & \text{para } x_1 \leq 0 \\ \\ \int_0^{\infty} 6e^{-2x_1}e^{-3x_2}dx_2 = 2e^{-2x_1} & \text{para } x_1 > 0 \end{cases}$$

Análogamente,

$$f_2(x_2) = \begin{cases} 0 & \text{para } x_2 \leq 0 \\ \\ 3e^{-3x_1} & \text{para } x_2 > 0 \end{cases}$$

Así,

$$F_1(x_1) = \begin{cases} 0 & \text{para } x_1 \leq 0 \\ \\ (1 - e^{-2x_1}) & \text{para } x_1 > 0 \end{cases}$$

$$F_2(x_2) = \begin{cases} 0 & \text{para } x_2 \leq 0 \\ \\ (1 - e^{-3x_2}) & \text{para } x_2 > 0 \end{cases}$$

Como se tenía que

$$F(x_1, x_2) = \begin{cases} 0 & \text{para } x_1 \leq 0 \text{ ó } x_2 \leq 0 \\ \\ (1 - e^{-2x_1})(1 - e^{-3x_2}) & \text{para } x_1, x_2 > 0 \end{cases}$$

de modo que $F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2)$, con lo que se puede afirmar que X_1 y X_2 son independientes.

De manera similar al caso discreto se introducen las distribuciones condicionadas. Por ejemplo, $X_1|X_2 = x_2$ con función de densidad

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}$$

para $f_2(x_2) \neq 0$.

Ejemplo Consideramos la densidad bivariante

$$f(x_1, x_2) = \begin{cases} \frac{2}{3}(x_1 + 2x_2) & \text{para } 0 < x_1 < 1, \quad 0 < x_2 \leq 1 \\ \\ 0 & \text{en otro caso} \end{cases}$$

La marginal de X_2 es

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1 = \begin{cases} 0 & x_2 \notin (0,1) \\ \int_0^1 \frac{2}{3}(x_1 + 2x_2)\, dx_1 = \frac{1}{3}(1 + 4x_2) & \text{para } 0 < x_2 < 1 \end{cases}$$

La densidad de X_1 condicionada a $X_2 = x_2$ (siendo $x_2 \in (0,1)$) es

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = \frac{\frac{2}{3}(x_1 + 2x_2)}{\frac{1}{3}(x_1 + 4x_2)} = \frac{2x_1 + 4x_2}{1 + 4x_2}$$

donde $0 < x_1 < 1$.

En el caso de independencia se tiene que

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = \frac{f_1(x_1) f_2(x_2)}{f_2(x_2)} = f_1(x_1).$$

Distribución normal multivariante

Es una generalización para vectores continuos del modelo normal. En el caso bivariante, la distribución normal de un vector (X, Y) de media $\mu = (\mu_1, \mu_2)$ y matriz de covarianzas

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \text{cov}(X,Y) \\ \text{cov}(X,Y) & \sigma_2^2 \end{pmatrix}$$

tiene como función de densidad

$$f(x,y) = \frac{1}{\left(\sqrt{2\pi}\right)^2 \sqrt{|\Sigma|}} \exp\left\{ -\frac{1}{2} (x - \mu_1, \quad y - \mu_2)\, \Sigma^{-1} \begin{pmatrix} x - \mu_1 \\ y - \mu_2 \end{pmatrix} \right\},$$

y se representa como $N(\mu, \Sigma)$,

Esta expresión se generaliza de modo trivial al caso de un vector con n componentes.

Propiedades

1. La distribución marginal de X es $N(\mu_1, \sigma_1)$
2. La distribución marginal de Y es $N(\mu_2, \sigma_2)$
3. La distribución de Y condicionada por $X = x$ es

   $$N\left(\mu_2 + \frac{\text{cov}(X,Y)}{\sigma_1^2}(x - \mu_1); \quad \sigma_2\sqrt{1 - \rho^2}\right)$$

   donde ρ es el coeficiente de correlación

   $$\rho = \frac{\text{cov}(X,Y)}{\sigma_1 \sigma_2}$$

4. Si un vector aleatorio (X, Y) tiene distribución $N(\mu, \Sigma)$ y $\text{cov}(X, Y) = 0$ entonces X e Y son independientes. Como

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix},$$

sustituyendo en la expresión de la función de densidad, se obtiene que

$$f(x, y) = f(x) \cdot f(y)$$

Propiedades del operador esperanza

Dada una v.a. X se puede definir la transformación $g(X)$ y calcular su esperanza definida mediante

$$E[g(X)] = \begin{cases} \int_{-\infty}^{\infty} g(x) f(x) dx & \text{en el caso continuo} \\ \sum_i g(x_i) f(x_i) & \text{en el caso discreto} \end{cases}$$

Para $Y = g(X_1, \ldots, X_k)$ se tiene que

$$E[Y] = \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_k) f(x_1, \ldots, x_k) dx_k \ldots dx_1 & \text{en el caso continuo} \\ \sum_{x_1} \cdots \sum_{x_k} g(x_1, \ldots, x_k) f(x_1, \ldots, x_k) & \text{en el caso discreto} \end{cases}$$

Un caso particular es

$$g(x_1, x_2) = (x_1 - \mu_1)(x_2 - \mu_2).$$

La esperanza de $g(x_1, x_2)$ es

$$E[(X_1 - \mu_1)(X_2 - \mu_2)] = Cov(X_1, X_2),$$

se llama **covarianza** y es una medida de la relación de crecimiento conjunto de ambas variables. Cuando toma valor positivo es porque predominan valores de X_1, X_2 grandes a la vez, o ambos pequeños a la vez. Cuando es negativo, resulta que X_1 es grande y X_2 pequeño a la vez, o a la inversa.

En el caso de independencia se tiene

$$Cov(X_1, X_2) = 0.$$

Propiedades útiles

Se pueden considerar las siguientes:

1. $E[kX] = kE[X]$.
2. $E[X+Y] = E[X] + E[Y]$.
3. $Var[kX] = k^2 Var[X]$.
4. Si X e Y son incorreladas entonces $E[XY] = E[X]E[Y]$.
5. $Var[X_1 + \cdots + X_n] = Var[X_1] + \cdots + Var[X_n] + 2\sum_{i<j} Cov[X_i, X_j]$.
6. Si $X_1, \cdots, X_n$ son incorreladas, entonces $Var[X_1 + \cdots + X_n] = Var[X_1] + \cdots + Var[X_n]$.
7. $Var[X-Y] = Var[X] + Var[Y] + 2Cov[X,Y]$.
8. Si X e Y son incorreladas, entonces $Var[X-Y] = Var[X] + Var[Y]$.

Ejemplo Supongamos que $X_1, \ldots, X_n$ son v.a. independientes con media μ y varianza σ^2. Definimos

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

Se tiene

$$\begin{aligned} E\left(\overline{X}\right) &= E\left(\sum_i \frac{1}{n} X_i\right) = \sum_i \frac{1}{n} E(X_i) = \sum_i \frac{1}{n}\mu = \mu \\ Var\left(\overline{X}\right) &= \sum_i \frac{1}{n^2} Var(X_i) = \frac{1}{n^2}\sum_i Var(X_i) = \frac{\sigma^2}{n} \end{aligned}$$

Teorema Central del Límite

Si $X_1, \ldots, X_n$ son v.a. independientes con media μ y varianza común $\sigma^2 < \infty$, la v.a.Z definida como

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

es una v.a. cuya función de densidad se aproxima a la distribución normal cuando n es grande:

$$Z \sim N(0, 1)$$

esto es, para n grande

$$\frac{X_1 + \cdots + X_n}{n} = \overline{X} \simeq N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$