Probabilidad I

Grado en Matemáticas

Estructura del curso

Javier Cárcamo

Departamento de Matemáticas
Universidad Autónoma de Madrid
javier.carcamo@uam.es

Información de contacto

Profesor: Javier Cárcamo

Correo electrónico: javier.carcamo@uam.es

Teléfono: 91 497 7635

Despacho: Módulo 17 - Despacho 412

Página web: http://verso.mat.uam.es/~javier.carcamo/

Tutorías: Bajo petición

Probabilidad I

Tema 1. Espacios de probabilidad

Tema 2. Variables aleatorias

Tema 3. Vectores aleatorios

Tema 4. Esperanza y momentos

Tema 5. Funciones características

Tema 6. Teoremas límite

- Desarrollar una intuición sobre los *fenómenos aleatorios*.
- Comprender y manejar los principios básicos del cálculo de probabilidades.
- Conocer los resultados básicos de la probabilidad, incluida su demostración, al menos en situaciones sencillas.
- Manejar funciones de distribución y de densidad en una y varias variables.
- Familiarizarse con las distribuciones de probabilidad más usuales y sus aplicaciones.
- Ser capaz de modelar fenómenos aleatorios sencillos.
- Entender y saber utilizar la convergencia de sucesiones de variables aleatorias, la ley de los grandes números y el teorema central del límite.

Importante: Este curso es una introducción elemental a la Teoría de la probabilidad. La formalización de muchas de las ideas que se verán en este curso se dará en Probabilidad II.

Transparencias: Se pueden descargar en moodle:

Apuntes Probabilidad 1 (1on1)
Apuntes Probabilidad 1 (2on1)
Apuntes Probabilidad 1 (4on1)

Relaciones de problemas: Dado el limitado tiempo, sólo haremos en clase los problemas en los que se tengan dificultades. Estos ejercicios se pedirán por adelantado al profesor para desarrollarlos en las clases de problemas.

Página web: En el Moodle de grado de la asignatura se encontrará disponible el resto del material del curso: Guía docente; horarios y fechas de exámenes; calificaciones (cuando las haya); información acerca de las revisiones; otros materiales complementarios; etc.

Exámenes parciales: Se realizará 2 exámenes parciales de 2 horas. La fechas previstas son los jueves 14 de marzo (parcial 1) y 25 de abril (parcial 2) de 2019.

Examen final: lunes 20 de mayo por la tarde (15 h.).

Calificación: La nota final será:

$$\text{Nota} = \text{máx}\{0{,}3 \times \text{Nota parciales} + 0{,}7 \times \text{Nota final}, \text{Nota final}\}.$$

Atención: Para poder aprobar usando los parciales es necesario sacar al menos un 4 en el examen final. La asistencia a clase y la entrega voluntaria de algunos problemas propuestos pueden ayudar a mejorar la nota.

Evaluación extraordinaria: Aquellos alumnos que no hayan superado la convocatoria ordinaria podrán presentarse al examen extraordinario el martes 18 de junio por la mañana (10 h.). La nota de los exámenes parciales no se conservará para esta convocatoria.

Calificaciones: opción 2 (aprobar por parciales)

Exámenes parciales: Se realizará 2 exámenes parciales de 2 horas. La fechas previstas son el 14 de marzo (parcial 1) y 25 de abril (parcial 2) de 2019.

Examen final: lunes 20 de mayo por la tarde (15 h.).

Calificación final:

(1) $C = (NP1 + NP2)/2$ si $NP1$ y $NP2 \geq 4$ y $C \geq 5$.

(2) NEF (Nota del examen final) si no se verifica (1).

(3) Subir nota: $\text{máx}\{5, NEF\}$ si $(NP1 + NP2)/2 \geq 5$.

Atención: La asistencia a clase y la entrega voluntaria de algunos problemas propuestos pueden ayudar a mejorar la nota.

Evaluación extraordinaria: Aquellos alumnos que no hayan superado la convocatoria ordinaria podrán presentarse al examen extraordinario el martes 18 de junio por la mañana (10 h.).

Bibliografía

- Grimmet, G. y Welsh, D. (1986). Probability: An Introduction. Oxford University Press (libro de texto).
- Pitman, J. (1993). Probability. Springer.
- Spiegel, M. *et al.* (2010). Probabilidad y estadística. Schaum. McGraw-Hill.
- Freedman, D. *et al.* (1991, 1998, 2007). Statistics. Norton (traducción: Estadística. Antoni Bosch, 1993).

Probabilidad I

Grado en Matemáticas

Tema 1
Espacios de probabilidad

Javier Cárcamo

Departamento de Matemáticas
Universidad Autónoma de Madrid
javier.carcamo@uam.es

Descripción del tema

1. Introducción: Fenómenos y experimentos aleatorios.
2. Álgebra de sucesos.
3. Espacios de probabilidad.
4. Probabilidad condicional.
5. Algunos resultados clásicos.
6. Independencia de sucesos.
7. Ejemplos.
8. Apéndice: Combinatoria elemental.

Objetivos principales

- Comprender la naturaleza de los experimentos aleatorios.
- Entender la probabilidad como medida de incertidumbre.
- Familiarizarse con la notación de la Teoría de la probabilidad.
- Entender que el conocimiento de nueva información sobre el modelo de probabilidad afecta al cálculo de las probabilidades.
- Manejar con soltura los resultados clásicos relativos a la probabilidad condicional.
- Familiarizarse con la noción de independencia estocástica.

Tipos de fenómenos:

1. **Deterministas**: Se conoce desde el principio el resultado final.
2. **Aleatorios**: Muchas situaciones finales posibles.

La **Teoría de la probabilidad** estudia el comportamiento de los fenómenos o experimentos aleatorios.

Dado ϵ experimento aleatorio:

- Conocemos con antelación el conjunto de todos los posibles resultados finales Ω (**espacio muestral**).
- No es posible determinar que resultado se va a dar previa realización de ϵ.
- La Teoría de la Probabilidad nos permitirá "medir"o "cuantificar"la incertidumbre asociada a los posibles resultados finales de un experimento aleatorio.

Ejemplos: Experimentos aleatorios y sus espacios muestrales

1. ϵ: lanzar un dado al aire. $\Omega = \{1, 2, 3, 4, 5, 6\}$.
2. ϵ: lanzar una moneda al aire. $\Omega = \{C, +\}$.
   ($C =$"sacar cara", $+ =$"sacar cruz")
3. ϵ: jugar en bolsa. $\Omega = \{G, P\}$.
   ($G =$"ganar dinero", $P =$"perder dinero")
4. ϵ: lanzar dos dados al aire.

$$\begin{aligned}\Omega &= \{(1,1),(1,2),\dots,(1,6),\dots,(6,1),\dots,(6,6)\}\\ &= \{(a,b) : a, b \in \{1,2,3,4,5,6\}\}.\end{aligned}$$

Situaciones con resultados aleatorios de la vida real

1. Control de calidad en una fábrica (pieza defectuosa o no).
2. Características de muestras recogidas en el campo (especie de insecto, longitud del fémur).
3. Peso y altura de una persona.

Definiciones básicas

Ω es el **espacio muestral**: conjunto de todos los posibles resultados finales de un experimento aleatorio ϵ.

Se llama **suceso (aleatorio)** a un subconjunto del espacio muestral ($A \subseteq \Omega$).
(Concretaremos la definición de suceso con más detalle en breve.)

Si el suceso está formado por un sólo elemento $\omega \in \Omega$, es decir, $A = \{\omega\} \subseteq \Omega$, el suceso se llama **suceso elemental**. Si A no es elemental, A se dice **suceso compuesto**.

El suceso Ω se denomina **suceso seguro**.
(Ω siempre ocurre al realizar ϵ)

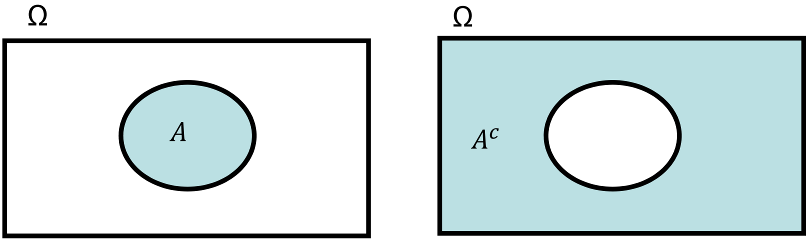
El suceso $\emptyset$ se denomina **suceso imposible**.
($\emptyset$ nunca ocurre al realizar ϵ)

Ejemplo: ϵ : lanzar un dado al aire.

El espacio muestral $\Omega = \{1, 2, 3, 4, 5, 6\}$.

- Un suceso A es cualquier subconjunto de $\{1, 2, 3, 4, 5, 6\}$.
- $P = \{2, 4, 6\}$ es un suceso (sacar un número par).
- $S_4 = \{4\}$ es un suceso elemental (sacar 4).
- $I = \{1, 3, 5\}$ es un suceso (sacar un número impar).
- $\Omega = \{1, 2, 3, 4, 5, 6\}$ es el suceso seguro (sacar cualquier número).

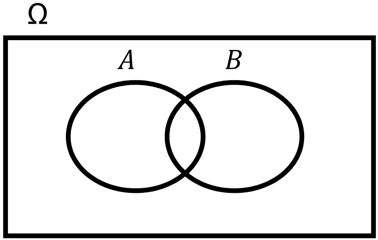
Si $A \subset \Omega$ es un suceso, A^c (el complementario del conjunto A) se denomina **suceso contrario** a A.

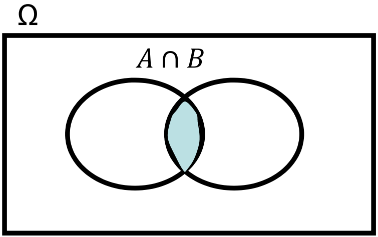


Ejemplo: ϵ : lanzar un dado al aire ($\Omega = \{1, 2, 3, 4, 5, 6\}$).

- Si $A = \{2, 4, 6\} \Rightarrow A^c = \{1, 3, 5\}$.
- Si $A = \{4\} \Rightarrow A^c = \{1, 2, 3, 5, 6\}$.
- Si $A = \Omega \Rightarrow A^c = \emptyset$.

Sean A y B dos sucesos. La **intersección** de los sucesos A y B, $A \cap B$, representa que se cumplan ambos sucesos simultáneamente.

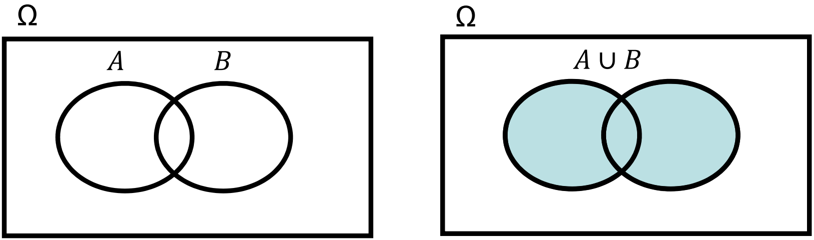




Ejemplo: ϵ : lanzar un dado al aire ($\Omega = \{1, 2, 3, 4, 5, 6\}$).

- Si $A = \{1, 3, 5\}$, $B = \{2, 4, 6\} \Rightarrow A \cap B = \emptyset$.
- Si $A = \{2, 4\}$, $B = \{1, 4, 5\} \Rightarrow A \cap B = \{4\}$.

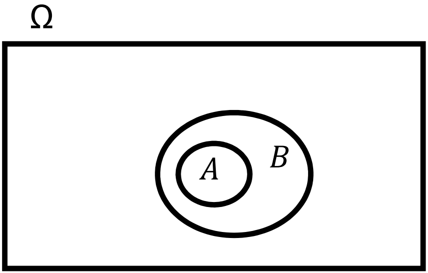
Sean A y B dos sucesos. La **unión** de los sucesos A y B, $A \cup B$, representa que se cumpla el suceso A o que se cumpla el suceso B.



Ejemplo: ϵ : lanzar un dado al aire ($\Omega = \{1, 2, 3, 4, 5, 6\}$).

- Si $A = \{1, 3, 5\}$, $B = \{2, 4, 6\} \Rightarrow A \cup B = \Omega$.
- Si $A = \{2, 4\}$, $B = \{1, 4, 5\} \Rightarrow A \cup B = \{1, 2, 4, 5\}$.

Sean A y B dos sucesos. La **inclusión** de A en B, $A \subseteq B$, representa que se si se cumple A, entonces también se cumple B.



Ejemplo: ϵ : lanzar un dado al aire ($\Omega = \{1, 2, 3, 4, 5, 6\}$).

- Si $A = \{2\}$, $B = \{2, 4, 6\} \Rightarrow A \subseteq B$.
  (Si sacamos un 2, entonces hemos sacado un número par)

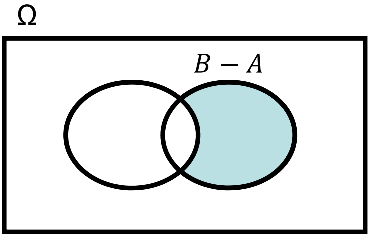
Sean A y B dos sucesos. La **diferencia** de B y A,

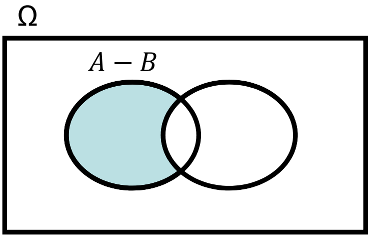
$$B - A = B \cap A^c,$$

representa que se cumple B y *no* se cumple A.

Si $A \subseteq B$, $B - A$ se denomina **diferencia propia**.

Observar que $A^c = \Omega - A$.

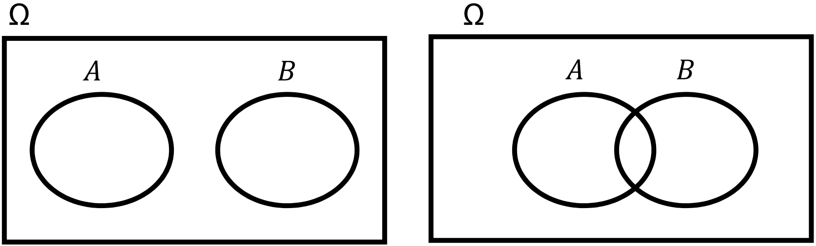




Ejemplo: ϵ : lanzar un dado al aire ($\Omega = \{1, 2, 3, 4, 5, 6\}$).

- Si $A = \{2, 4, 6\}$, $B = \{2, 4\} \Rightarrow A - B = \{6\}$.
- Si $A = \{2, 4, 6\}$, $B = \Omega \Rightarrow A - B = \emptyset$.

Sean A y B dos sucesos. A y B se dicen **disjuntos** o **incompatibles** si $A \cap B = \emptyset$. En este caso la unión entre ellos se denota también $A \cup B = A \uplus B = A + B$ (unión disjunta).



Disjuntos **No disjuntos**

Ejemplo: ϵ : lanzar un dado al aire ($\Omega = \{1, 2, 3, 4, 5, 6\}$).

- $A = \{2, 4, 6\}$, $B = \{1, 3, 5\}$ son disjuntos. $A + B = \Omega$.
- $A = \{2, 4, 6\}$ y $B = \{2\}$ *no* son disjuntos.

Propiedades de las operaciones con sucesos

Conmutatividad:

- $A \cup B = B \cup A$ (unión).
- $A \cap B = B \cap A$ (intersección).

Asociatividad:

- $A \cup (B \cup C) = (A \cup B) \cup C$ (unión).
- $A \cap (B \cap C) = (A \cap B) \cap C$ (intersección).

Elementos neutros:

- $A \cup \emptyset = A$ (unión).
- $A \cap \Omega = A$ (intersección).

Distributividad:

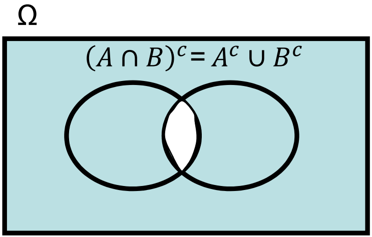
- De la unión respecto a la intersección:
  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
- De la intersección respecto a la unión:
  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

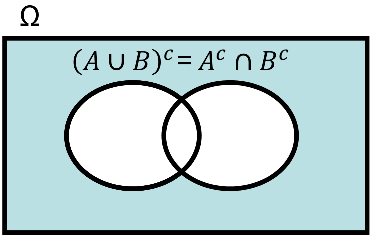
Propiedades de complementación

- $(A^c)^c = A$.
- $\Omega^c = \emptyset$ y $\emptyset^c = \Omega$.

Leyes de De Morgan

- $(A \cap B)^c = A^c \cup B^c$.
- $(A \cup B)^c = A^c \cap B^c$.





Queremos definir la probabilidad como *medida* de la incertidumbre.

Definiciones de la probabilidad en la historia

Clásica: Basada en los juegos de azar. La probabilidad se define como el cociente entre los casos favorables y los posibles.

Inconvenientes: ¿Qué ocurre cuando Ω es infinito o cuando los sucesos elementales no son equiprobables?

Frecuentista o empírica: La probabilidad de un suceso se define como el límite de las frecuencias relativas del suceso.

Inconvenientes: ¿Qué número de pruebas debemos realizar?, ¿qué ocurre con aquellos experimentos que se puedan repetir una sola vez?

Axiomática: Engloba a las anteriores y solventa los problemas mencionados. Es la que estudiaremos y se debe a Kolmogorov.

Un **espacio de probabilidad** es un triplete $(\Omega, \mathcal{F}, \mathrm{P})$, donde

(1) Ω es un conjunto no vacío llamado **espacio muestral**.

(2) $\mathcal{F}$ es una σ-**álgebra** o **tribu** de $\mathcal{P}(\Omega)$ (partes de Ω).

(3) P es una **medida de probabilidad** sobre $\mathcal{F}$, es decir,

$$\begin{aligned} \mathrm{P} : \mathcal{F} &\longrightarrow [0,1] \\ A &\longmapsto \mathrm{P}(A) \end{aligned}$$

verificando los **axiomas de probabilidad de Kolmogorov**:

(A1) $\mathrm{P}(\Omega) = 1$.

(A2) σ-**aditividad** o **aditividad numerable**: $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$ disjuntos dos a dos (es decir, $A_i \cap A_j = \emptyset$, si $i \neq j$), entonces:

$$\mathrm{P}\left(\biguplus_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mathrm{P}(A_i).$$

Nota: Como se verá en otras asignaturas (Teoría de la integración y la medida, Probabilidad II), en algunas ocasiones no es posible calcular la probabilidad (o la medida) de cualquier subconjunto de Ω.

Las colecciones de conjuntos adecuadas para calcular probabilidades son las σ-álgebras.

Una colección $\mathcal{F} \subset \mathcal{P}(\Omega)$ se dice que es una σ-**álgebra** o **tribu** si

(1) $\Omega \in \mathcal{F}$.

(2) $\mathcal{F}$ es cerrada o estable para la complementación.

$$\text{Si } A \in \mathcal{F}, \text{ entonces } A^c \in \mathcal{F}.$$

(3) $\mathcal{F}$ es estable para la unión numerable.

$$\text{Si } \{A_i\}_{i=1}^{\infty} \subset \mathcal{F}, \text{ entonces } \cup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

Los elementos de $\mathcal{F}$ se llaman **sucesos (aleatorios)**.

Propiedades de las σ-álgebras

(1) $\emptyset \in \mathcal{F}$.

(2) Si $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$, entonces $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

(3) Si $\{A_i\}_{i=1}^{n} \subset \mathcal{F}$, entonces $\bigcap_{i=1}^{n} A_i \in \mathcal{F}$ y $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$.

- Si $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$, decimos que A_n **crece** hasta A, $A_n \uparrow A$, si $A_1 \subset A_2 \subset A_3 \subset \cdots$ y $A = \bigcup_{i=1}^{\infty} A_i$.
- Si $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$, decimos que A_n **decrece** hasta A, $A_n \downarrow A$, si $A_1 \supset A_2 \supset A_3 \supset \cdots$ y $A = \bigcap_{i=1}^{\infty} A_i$.

(4) $\mathcal{F}$ es estable para límites crecientes y decrecientes:
Si $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$ y $A_n \uparrow A$ o $A_n \downarrow A$, entonces $A \in \mathcal{F}$.

Nota: Para nuestros intereses y por simplicidad, podemos suponer que todos los subconjuntos de Ω están en la σ-álgebra $\mathcal{F}$. Por tanto, supondremos que podemos calcular la probabilidad de cualquier conjunto.

Propiedades de la probabilidad

1. $\mathrm{P}(\emptyset) = 0$.
2. $\{A_i\}_{i=1}^n \subset \mathcal{F}$ disjuntos, entonces $\mathrm{P}\left(\biguplus_{i=1}^n A_i\right) = \sum_{i=1}^n \mathrm{P}(A_i)$.
3. Si A, $B \in \mathcal{F}$, entonces $\mathrm{P}(B - A) = \mathrm{P}(B) - \mathrm{P}(A \cap B)$.
4. A, $B \in \mathcal{F}$ y $A \subset B$. $\mathrm{P}(B - A) = \mathrm{P}(B) - \mathrm{P}(A)$ y $\mathrm{P}(A) \leq \mathrm{P}(B)$.
5. $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$.
6. **Principio de inclusión-exclusión:**

$$\begin{aligned} \mathrm{P}(A \cup B) &= \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B) \\ \mathrm{P}(A \cup B \cup C) &= \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) \\ &\quad - (\mathrm{P}(A \cap B) + \mathrm{P}(A \cap C) + \mathrm{P}(B \cap C)) \\ &\quad + \mathrm{P}(A \cap B \cap C). \end{aligned}$$

$$\mathrm{P}\left(\cup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathrm{P}(A_i) - \sum_{i<j} \mathrm{P}(A_i \cap A_j) + \cdots + (-1)^{n+1} \mathrm{P}\left(\cap_{i=1}^n A_i\right).$$

7. $\{A_i\}_{i=1}^n \subset \mathcal{F}$, entonces $\mathrm{P}(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathrm{P}(A_i)$.

Propiedades de la probabilidad

1. $\mathrm{P}(\emptyset) = 0$.
2. $\{A_i\}_{i=1}^n \subset \mathcal{F}$ disjuntos, entonces $\mathrm{P}\left(\biguplus_{i=1}^n A_i\right) = \sum_{i=1}^n \mathrm{P}(A_i)$.
3. Si A, $B \in \mathcal{F}$, entonces $\mathrm{P}(B - A) = \mathrm{P}(B) - \mathrm{P}(A \cap B)$.
4. A, $B \in \mathcal{F}$ y $A \subset B$. $\mathrm{P}(B - A) = \mathrm{P}(B) - \mathrm{P}(A)$ y $\mathrm{P}(A) \leq \mathrm{P}(B)$.
5. $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$.
6. **Principio de inclusión-exclusión:**
$$\mathrm{P}\left(\cup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathrm{P}(A_i) - \sum_{i<j} \mathrm{P}(A_i \cap A_j) + \cdots + (-1)^{n+1} \mathrm{P}\left(\cap_{i=1}^n A_i\right).$$
7. $\{A_i\}_{i=1}^n \subset \mathcal{F}$, entonces $\mathrm{P}(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathrm{P}(A_i)$.
8. $\{A_i\}_{i=1}^\infty \subset \mathcal{F}$ y $A_n \uparrow A$, entonces $\mathrm{P}(A_n) \uparrow \mathrm{P}(A)$.
9. $\{A_i\}_{i=1}^\infty \subset \mathcal{F}$ y $A_n \downarrow A$, entonces $\mathrm{P}(A_n) \downarrow \mathrm{P}(A)$.
10. $\{A_i\}_{i=1}^\infty \subset \mathcal{F}$, $\mathrm{P}(\cup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty \mathrm{P}(A_i)$.

Ejemplo: En una escuela el 50 % del alumnado ha aprobado inglés, el 20 % francés y el 5 % los dos idiomas. ¿Cuál es la probabilidad de seleccionar un alumno al azar que haya aprobado alguna de las dos asignaturas?

Ejemplo: Sean A y B dos sucesos de $(\Omega, \mathcal{F}, \mathrm{P})$ tales que $\mathrm{P}(A) = \mathrm{P}(B) = 1/2$ y $\mathrm{P}(A \cup B) = 2/3$. Calcular:

(a) $\mathrm{P}(A \cap B)$.

(b) $\mathrm{P}(A^c \cap B^c)$.

(c) $\mathrm{P}(A \cap B^c)$.

(d) $\mathrm{P}(A^c \cap B)$.

Ejemplo: Sean A y B dos sucesos de un espacio de probabilidad $(\Omega, \mathcal{F}, \mathrm{P})$ tales que $\mathrm{P}(A) = 0{,}3$, $\mathrm{P}(B) = 0{,}4$ y $\mathrm{P}(A \cap B) = 0{,}1$. Señalar qué afirmaciones son falsas.

(a) $\mathrm{P}(A \cup B) = 0{,}6$.

(b) $\mathrm{P}(A \cap B^c) = 0{,}2$.

(c) $\mathrm{P}(A \cap (B \cup B^c)) = 0{,}4$.

(d) $\mathrm{P}(A^c \cap B) = 0{,}3$.

(e) $\mathrm{P}((A \cap B)^c) = 0{,}9$.

Modelo clásico o de Laplace: Dado $\Omega = \{\omega_1, \ldots, \omega_n\}$ espacio muestral finito, la aplicación:

$$\mathrm{P}(A) = \frac{\operatorname{card}(A)}{n}, \quad A \subseteq \Omega,$$

es una medida de probabilidad sobre $(\Omega, \mathcal{P}(\Omega))$.

En este ejemplo, cada elemento ω_i ($i = 1, \ldots, n$) tiene la misma probabilidad ($\mathrm{P}(\{\omega_i\}) = 1/n$, $i = 1, \ldots, n$). Esto se conoce como **equiprobabilidad**.

Ejemplo: ϵ : lanzar un dado al aire ($\Omega = \{1, \ldots, 6\}$).
$\mathrm{P}(A) = \operatorname{card}(A)/6$ ($A \subseteq \Omega$) es una probabilidad sobre Ω.

$$\mathrm{P}(\{2,4,6\}) = 3/6 = 1/2, \qquad \mathrm{P}(\{6\}) = 1/6.$$

Ejemplo: ϵ : extraer una carta de una baraja española ($\Omega = \{Xi : X \in \{O, B, C, E\}, i \in \{1,2,3,4,5,6,7,S,C,R\}\}$).
$\mathrm{P}(A) = \operatorname{card}(A)/40$ ($A \subseteq \Omega$) es una probabilidad sobre Ω.

$$\mathrm{P}(\text{«sacar figura»}) = 12/40 = 3/10, \qquad \mathrm{P}(\{O1\}) = 1/40.$$

Modelo finito: Sea $\Omega = \{\omega_1, \dots, \omega_n\}$ un espacio muestral finito y $\{p_1, \dots, p_n\}$ números con $p_i \geq 0$ $(i = 1, \dots, n)$ y $p_1 + \cdots + p_n = 1$. La aplicación:

$$\mathrm{P}(\{\omega_i\}) = p_i \quad (i = 1, \dots, n), \quad \mathrm{P}(A) = \sum_{\omega_i \in A} \mathrm{P}(\{\omega_i\}), \quad A \subseteq \Omega,$$

es una medida de probabilidad sobre $(\Omega, \mathcal{P}(\Omega))$.

Nota: El modelo de Laplace es un caso particular del modelo finito en el que todas las masas de probabilidad son iguales.

Ejemplo: ϵ : lanzar un dado al aire cargado o trucado $(\Omega = \{1, \dots, 6\})$ de forma que la probabilidad de los sucesos elementales es: $\mathrm{P}(\{6\}) = 1/2$ y $\mathrm{P}(\{i\}) = 1/10$, $i = 1, \dots, 5$.

$$\mathrm{P}(\{2,4,6\}) = 2/10 + 1/2 = 7/10, \qquad \mathrm{P}(\{1,3\}) = 1/5.$$

Ejemplo: ϵ : puntuación al lanzar dos dados $(\Omega = \{2, \dots, 12\})$.

$$\mathrm{P}(\{2\}) = 1/36, \quad \mathrm{P}(\{7\}) = 1/6, \quad \mathrm{P}(\{2,7\}) = 1/36 + 1/6 = 7/36.$$

Espacios discretos: $\Omega = \{\omega_1, \omega_2, \dots\}$ contable, $\mathcal{F} = \mathcal{P}(\Omega)$ y $\mathrm{P}(\{w_i\}) \in [0,1]$ tal que $\sum_i \mathrm{P}(\{\omega_i\}) = 1$. $(\Omega, \mathcal{F}, \mathrm{P})$ es un espacio de probabilidad.

$$\mathrm{P}(A) = \sum_{\omega_i \in A} \mathrm{P}(\{\omega_i\}), \quad A \subseteq \Omega.$$

Ejemplo: ϵ : Lanzar una moneda sucesivamente y observar cuantas caras sacamos antes de que salga la primera cruz.

En este ejemplo $\Omega = \{0, 1, 2, \dots\}$ y $\mathrm{P}(\{n\}) = 1/2^{n+1}$ $(n \geq 0)$.

$$\mathrm{P}(\{2\}) = 1/2^3, \quad \mathrm{P}(\{7\}) = 1/2^8, \quad \mathrm{P}(\{0,1,2,3\}) = 15/16.$$

Problema: El caballero de Meré (inicio de la probabilidad)

Tal caballero jugaba a dos juegos:

Juego A: Lanzar un dado 4 veces. Apuesta: sacar al menos un 6.

Juego B: Lanzar 2 dados 24 veces. Apuesta: sacar al menos un doble 6.

Observó (empíricamente) que con el juego A ganaba más de la mitad de las veces y, sin embargo, con el juego B perdía más de la mitad de las veces.

No entendía la razón de sus observaciones ya que argumentaba:

En A: $4 \cdot (1/6) = 2/3$ (4 tiradas por prob. de sacar 6)
En B: $24 \cdot (1/36) = 2/3$ (24 tiradas por prob. de sacar doble 6)

Preguntó a Pascal: ¿cómo es posible la diferencia de ganancias?

Indicación: Obviamente, Pascal sabía que no se pueden sumar probabilidades de sucesos que no sean mutuamente exclusivos.

Ejemplo: ϵ: lanzar un dado equilibrado. El modelo de probabilidad asociado es: $(\Omega = \{1, \ldots, 6\}, \mathcal{F} = \mathcal{P}(\Omega), \mathrm{P})$ (modelo de Laplace).

En este ejemplo, $\mathrm{P}(\{6\}) = 1/6$.

Supongamos ahora que poseemos cierta información adicional, por ejemplo, sabemos que ha salido un número par. Es decir, se ha producido el suceso $P = \{2, 4, 6\}$.

Es claro que ahora la probabilidad de sacar 6 ha cambiado. Ahora será una nueva probabilidad, P^*, y

- $\mathrm{P}^*(\{1\}) = 0$.
- $\mathrm{P}^*(\{6\}) = \dfrac{\text{Casos favorables}}{\text{Casos posibles}} = \dfrac{1}{3} = \dfrac{1/6}{1/2}$.

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad y A, $B \in \mathcal{F}$ con $\mathrm{P}(B) > 0$. Se llama **probabilidad de A condicionada a B** a:

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

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad y $B \in \mathcal{F}$ fijo con $\mathrm{P}(B) > 0$. La aplicación:

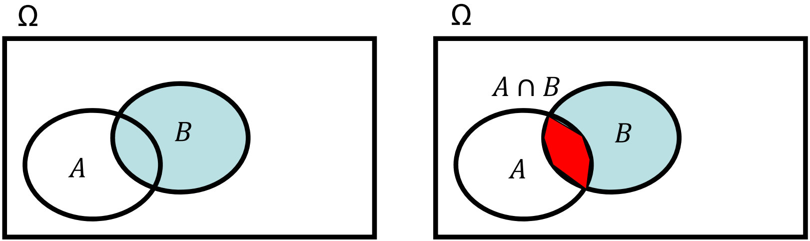
$$\begin{aligned} \mathrm{P}(\cdot|B) : \mathcal{F} &\longrightarrow [0,1] \\ A &\longmapsto \mathrm{P}(A|B) \end{aligned}$$

es una medida de probabilidad sobre $(\Omega, \mathcal{F})$. Por tanto, $\mathrm{P}(\cdot|B)$ verifica los axiomas y propiedades de una probabilidad.

La nueva información disponible (se ha dado B) ha modificado la medida de probabilidad sobre $(\Omega, \mathcal{F})$. Hemos pasado de $(\Omega, \mathcal{F}, \mathrm{P}(\cdot))$ a $(\Omega, \mathcal{F}, \mathrm{P}(\cdot|B))$. De esta manera incorporamos esta información conocida al modelo de probabilidad.

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad y A, $B \in \mathcal{F}$ con $\mathrm{P}(B) > 0$. Se llama **probabilidad de A condicionada a B** a:

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



Idea intuitiva: La $\mathrm{P}(A|B)$ es la medida de la parte de A que está en B normalizada para que la total tenga medida 1.

Ejemplo: Se lanzan dos dados legales. $(\Omega, \mathcal{F}, \mathrm{P})$ modelo de Laplace.

- ¿Cuál es la probabilidad de que la suma de los dados sea 7?
- ¿Cuál es la probabilidad de que la suma de los dados sea 7 sabiendo que...?
  - La suma de los dados es impar.
  - La suma es mayor que 6.
  - El resultado del primer dado es impar.
  - Los dos dados tuvieron el mismo resultado.
  - Los dos dados tuvieron distinto resultado.

$S(2)$	$S(3)$	$S(4)$	$S(5)$	$S(6)$	$S(7)$	$S(8)$	$S(9)$	$S(10)$	$S(11)$	$S(12)$
$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$	$(2,6)$	$(3,6)$	$(4,6)$	$(5,6)$	$(6,6)$
	$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$	$(3,5)$	$(4,5)$	$(5,5)$	$(6,5)$	
		$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$(4,4)$	$(5,4)$	$(6,4)$		
			$(4,1)$	$(4,2)$	$(4,3)$	$(5,3)$	$(6,3)$			
				$(5,1)$	$(5,2)$	$(6,2)$				
					$(6,1)$					

Fórmula del producto

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ espacio de probabilidad y $A_1, \ldots, A_n \in \mathcal{F}$ tales que $\mathrm{P}(A_1 \cap \cdots \cap A_n) > 0$. Se tiene:

- $\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1)\mathrm{P}(A_2|A_1)$.
- $\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_2|A_1)\mathrm{P}(A_3|A_1 \cap A_2)$.

En general:

- $\mathrm{P}(A_1 \cap \cdots \cap A_n) = \mathrm{P}(A_1)\mathrm{P}(A_2|A_1) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1})$.

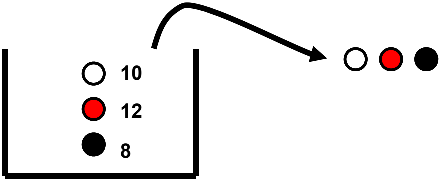
Observación: La probabilidad condicional nos permite (en algunas ocasiones) calcular la probabilidad de una intersección.

Fórmula del producto

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ espacio de probabilidad y $A_1, \ldots, A_n \in \mathcal{F}$ tales que $\mathrm{P}(A_1 \cap \cdots \cap A_n) > 0$. Se tiene:

$$\mathrm{P}(A_1 \cap \cdots \cap A_n) = \mathrm{P}(A_1)\mathrm{P}(A_2|A_1) \cdots \mathrm{P}(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

Ejemplo: Una urna contiene 10 bolas blancas, 12 rojas y 8 negras. ¿Cuál es la probabilidad de que extraigamos primero una bola blanca, luego una roja y luego una negra con y sin reemplazamiento?

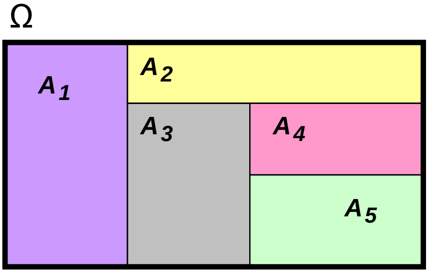


$$¿\mathrm{P}(B_1 \cap R_2 \cap N_3)?$$

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ espacio de probabilidad. Una colección contable (finita o numerable) de sucesos $\{A_1, A_2, \dots\} \subset \mathcal{F}$ se dice que es un **sistema completo de sucesos (S.C.S.)** o una **partición de** Ω si:

(a) $\mathrm{P}(A_i) > 0$, $i \geq 1$.

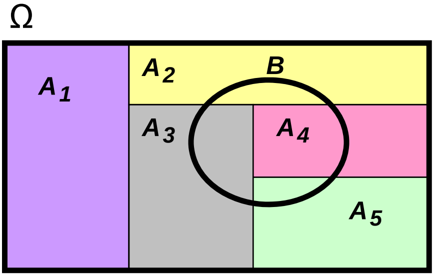
(b) $\Omega = \biguplus_i A_i$ (unión disjunta).



Fórmula de la probabilidad total

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad y $\{A_i\}_i \subset \mathcal{F}$ un S.C.S. Entonces, para cualquier $B \in \mathcal{F}$,

$$\mathrm{P}(B) = \sum_i \mathrm{P}(A_i)\mathrm{P}(B|A_i).$$



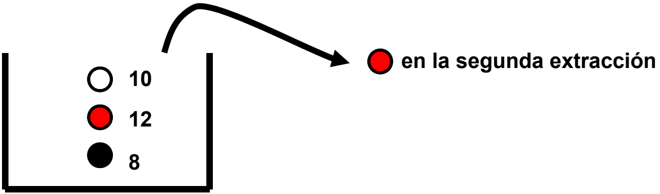
Nota: Mediante la Fórmula de la probabilidad total podemos calcular la probabilidad de cualquier suceso a través de las probabilidades condicionadas en un S.C.S.

Fórmula de la probabilidad total

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad y $\{A_i\}_i \subset \mathcal{F}$ un S.C.S. Entonces, para cualquier $B \in \mathcal{F}$,

$$\mathrm{P}(B) = \sum_i \mathrm{P}(A_i)\mathrm{P}(B|A_i).$$

Ejemplo: Una urna contiene 10 bolas blancas, 12 rojas y 8 negras. Hemos realizado dos extracciones sin reemplazamiento. ¿Cuál es la probabilidad de sacar una bola roja en la segunda extracción?



$$¿\mathrm{P}(R_2)?$$

Fórmula de Bayes (de la probabilidad a posteriori de las causas)

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad, $\{A_i\}_i \subset \mathcal{F}$ un S.C.S y $B \in \mathcal{F}$ con $\mathrm{P}(B) > 0$, entonces

$$\mathrm{P}(A_j|B) = \frac{\mathrm{P}(A_j)\mathrm{P}(B|A_j)}{\sum_i \mathrm{P}(A_i)\mathrm{P}(B|A_i)}, \qquad j \geq 1.$$

Las probabilidades $\mathrm{P}(A_n)$ se llaman **probabilidades a priori**.

Las probabilidades $\mathrm{P}(A_n|B)$ se llaman **probabilidades a posteriri**.

Nota: Mediante la Fórmula de Bayes podemos calcular las probabilidades de un S.C.S. condicionadas a B a través de las probabilidades condicionadas a ese S.C.S.

Nota: Si $A \in \mathcal{F}$ con $\mathrm{P}(A) \in (0,1)$, entonces $\{A, A^c\}$ es una partición de Ω. En este caso particular,

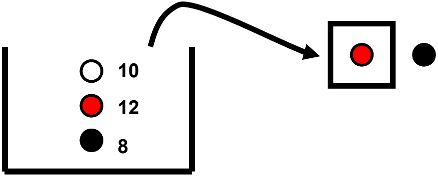
$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A)\mathrm{P}(B|A)}{\mathrm{P}(A)\mathrm{P}(B|A) + \mathrm{P}(A^c)\mathrm{P}(B|A^c)}.$$

Fórmula de Bayes

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad, $\{A_i\}_i \subset \mathcal{F}$ un S.C.S y $B \in \mathcal{F}$ con $\mathrm{P}(B) > 0$, entonces

$$\mathrm{P}(A_j|B) = \frac{\mathrm{P}(A_j)\mathrm{P}(B|A_j)}{\sum_i \mathrm{P}(A_i)\mathrm{P}(B|A_i)}, \qquad j \geq 1.$$

Ejemplo: Una urna contiene 10 bolas blancas, 12 rojas y 8 negras. Extraemos una bola (no la miramos). Extraemos una segunda bola y es negra. ¿Cuál es la probabilidad de que la bola que hemos extraído en primer lugar sea roja?



$¿\mathrm{P}(R_1|N_2)?$

Definición: $(\Omega, \mathcal{F}, \mathrm{P})$ espacio de probabilidad. Dos sucesos $A, B \in \mathcal{F}$ se dicen **independientes** si $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$.

Notación: $AB = A \cap B$. (A, B indep. sii $\mathrm{P}(AB) = \mathrm{P}(A)\mathrm{P}(B)$.)

Observaciones:

- Si $\mathrm{P}(A) = 0$, entonces A y B ind. para todo $B \in \mathcal{F}$.
- Si $\mathrm{P}(A) = 1$, entonces A y B ind. para todo $B \in \mathcal{F}$.
- Si A y B ind. con $\mathrm{P}(B) > 0$, entonces $\mathrm{P}(A|B) = \mathrm{P}(A)$.

  Nota: Si A y B son independientes, la ocurrencia de uno no altera la probabilidad de que ocurra el otro (y al revés). Esta es la idea detrás de la definición de independencia.
- Si A y B ind., entonces: A, B^c ind.; A^c, B ind.; y A^c, B^c ind.

Ejemplo: Las extracciones sucesivas de bolas con reemplazamiento son independientes. ($P(B_2|B_1) = P(B_2)$). Sin embargo, las extracciones sucesivas de bolas sin reemplazamiento *no* son independientes ($P(B_2|B_1) \neq P(B_2)$).

Definición: $(\Omega, \mathcal{F}, \mathrm{P})$ espacio de probabilidad. Tres sucesos $A, B, C \in \mathcal{F}$ se dicen **(mutuamente) independientes** si:

- $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$.
- $\mathrm{P}(A \cap C) = \mathrm{P}(A)\mathrm{P}(C)$.
- $\mathrm{P}(B \cap C) = \mathrm{P}(B)\mathrm{P}(C)$.
- $\mathrm{P}(A \cap B \cap C) = \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C)$.

Observación: $A, B, C \in \mathcal{F}$ son independientes sii $(*)$ y $(**)$.

$$(*) \qquad \left\{ \begin{array}{c} \mathrm{P}(AB) = \mathrm{P}(A)\mathrm{P}(B) \\ \mathrm{P}(AC) = \mathrm{P}(A)\mathrm{P}(C) \\ \mathrm{P}(BC) = \mathrm{P}(B)\mathrm{P}(C) \end{array} \right\} \quad \{A, B, C\} \ \textbf{ind. dos a dos}$$

$$(**) \qquad \mathrm{P}(ABC) = \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C)$$

Observación: $(*) \not\Rightarrow (**)$ y $(**) \not\Rightarrow (*)$.

Observación: $A, B, C \in \mathcal{F}$ son independientes sii $(*)$ y $(**)$.

$(*)$ $\quad \{A, B, C\}$ independientes dos a dos.
$(**)$ $\quad \mathrm{P}(ABC) = \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C)$.

Observación: $(*) \not\Rightarrow (**)$ y $(**) \not\Rightarrow (*)$.

Contraejemplo 1: $(*) \not\Rightarrow (**)$

$(\Omega = \{a, b, c, d\}, \mathcal{P}(\Omega), \mathrm{P})$ modelo de Laplace.

(Los sucesos elementales $\{a\}$, $\{b\}$, $\{c\}$ y $\{d\}$ son equiprobables.)

$A = \{a, b\}$, $B = \{b, c\}$ y $C = \{c, a\}$ son independientes dos a dos, pero *no* son (mutuamente) independientes.

Contraejemplo 2: $(**) \not\Rightarrow (*)$

$(\Omega = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8\}, \mathcal{P}(\Omega), \mathrm{P})$ modelo de Laplace.

$A = \{a_1, a_2, a_3, a_4\}$, $B = \{a_1, a_3, a_5, a_7\}$, $C = \{a_1, a_2, a_4, a_8\}$ verifican $(**)$, pero no son independientes dos a dos.

Definición: Sea $\{A_i\}_{i\in I} \subset \mathcal{F}$. Se dice que los sucesos $\{A_i\}_{i\in I}$ son **(mutuamente) independientes** si

$$\forall\{A_{i_1},\ldots,A_{i_n}\} \subset \{A_i\}_{i\in I},\ \mathrm{P}(A_{i_1}\cdots A_{i_n}) = \mathrm{P}(A_{i_1})\cdots\mathrm{P}(A_{i_n}).$$

En particular, n sucesos $\{A_1,\ldots,A_n\}$ son independientes sii:

- $\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1)\mathrm{P}(A_2)$.
- $\mathrm{P}(A_1 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_3)$.
- $\vdots \qquad \vdots$
- $\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_2)\mathrm{P}(A_3)$.
- $\mathrm{P}(A_1 \cap A_2 \cap A_4) = \mathrm{P}(A_1)\mathrm{P}(A_2)\mathrm{P}(A_4)$.
- $\vdots \qquad \vdots \qquad \vdots$
- $\mathrm{P}(A_1 \cap \cdots \cap A_n) = \mathrm{P}(A_1)\cdots\mathrm{P}(A_n)$.

Pregunta: Comprobar la independencia de algunos sucesos puede ser una tarea complicada. ¿Cuántas condiciones debemos comprobar para asegurar que $\{A_1,\ldots,A_n\}$ son independientes?

Proposición: $\{A_i\}_{i\in I}$ independientes $\Longrightarrow \{A_i^c\}_{i\in I}$ independientes.

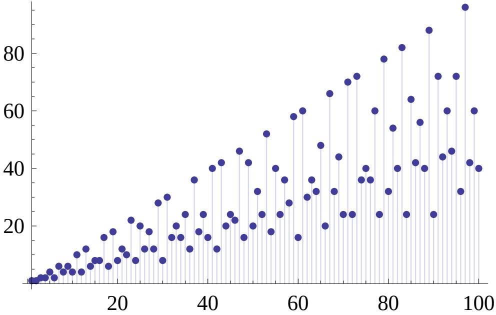
Una demostración probabilística de la fórmula de la φ de Euler

La **función φ de Euler**: $\varphi(n)$ ($n \in \mathbb{N}$) cuenta el número de naturales menores o iguales que n y coprimos con n. Es decir,

$$\varphi(n) = \text{Card}\{m \leq n : \text{m.c.d.}(m,n) = 1\}.$$

Resultado: Si $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$ con $p_1, \ldots, p_r$ primos, entonces

$$\varphi(n) = n(1 - 1/p_1) \cdots (1 - 1/p_r) = n \prod_{p|n} (1 - 1/p).$$



Ejercicio: Sean $A_1, A_2, \ , A_n$ sucesos independientes con $\mathrm{P}(A_i) = p$, para $i = 1, 2, \ldots, n$. Halla la probabilidad de que:

1. ninguno de los A_i ocurra;
2. un número par de los A_i ocurran.

Ejercicio: Con las hipótesis habituales, halla la probabilidad de que en una reunión de 25 personas haya al menos dos con la misma fecha de cumpleaños.

¿Cuál es el número mínimo de personas para que esa probabilidad sea al menos 0,5?

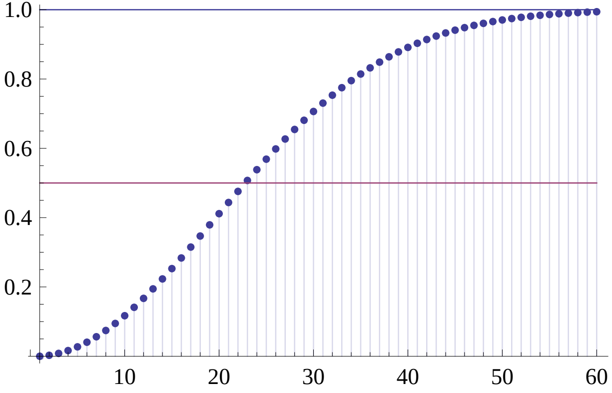
Nota: Para calcular $p = \frac{364}{365} \cdot \frac{363}{365} \cdot \cdots \cdot \frac{341}{365}$ se puede aproximar su logaritmo:

$$\begin{aligned}
\log p &= \left(\sum_{i=0}^{24} \log(365 - i) \right) - 25 \log 365 \\
&\approx 25 \frac{\log 365 + \log 341}{2} - 25 \log 365 \\
&= \frac{25}{2} (\log 341 - \log 365) = -0{,}8502.
\end{aligned}$$

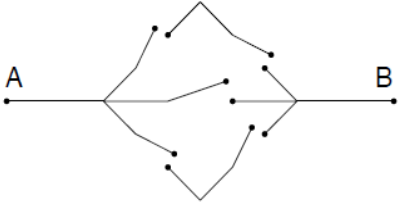
Por tanto, $1 - p \approx 0{,}57$. (El verdadero valor es $1 - p = 0{,}56869970 \cdots$.)

Ejercicio: Con las hipótesis habituales, halla la probabilidad de que en una reunión de 25 personas haya al menos dos con la misma fecha de cumpleaños.

¿Cuál es el número mínimo de personas para que esa probabilidad sea al menos 0,5?



Ejercicio: En el circuito de la figura, cada interruptor está cerrado con probabilidad p, independientemente de todos los demás. Subsana los fallos en la siguiente argumentación para calcular la probabilidad de que pueda circular la corriente entre A y B.



Sea C el suceso que indica que la corriente circula entre A y B.

1. $C = C_1 \cup C_2 \cup C_3$, donde C_1 significa que la corriente pasa por arriba, C_2 por el medio y C_3 por abajo.
2. $\mathrm{P}(C) = \mathrm{P}(C_1) + \mathrm{P}(C_2) + \mathrm{P}(C_3)$.
3. Por simetría, $\mathrm{P}(C) = 3\mathrm{P}(C_1)$.
4. Por independencia, $\mathrm{P}(C) = 3(1-p)^2$.

Ejercicio: El borracho. Un trasnochador dispone de un llavero con tres llaves totalmente indistinguibles en la oscuridad, de las cuales sólo una abre la puerta de su casa. Para dar con la llave en cuestión sigue uno de los siguientes métodos:

M_1: Prueba las llaves una tras otra teniendo cuidado de no volver a usar la misma.

M_2: Prueba una llave y si no abre agita el llavero y prueba otra vez.

Contéstese a las siguientes preguntas:

(a) ¿Cuál es la probabilidad de que abra al tercer intento si sigue el método primero? ¿y si sigue el segundo método?

(b) Se sabe además que el trasnochador usa el segundo método cuando vuelve a casa después de haber bebido en exceso (lo cual ocurre uno de cada tres días) y el primer método cuando regresa sobrio. Si se conoce que en los dos primeros intentos ha fracasado, ¿cuál es la probabilidad de que el trasnochador esté borracho?

Combinatoria: Teoría matemática que proporciona métodos para determinar el número de elementos de un conjunto finito.

Recordamos que, si $m, n \geq 0$, y $n \leq m$,

- **Factorial de** n: $n! = n \cdot (n-1) \cdots 2 \cdot 1$
  ($0! = 1$ por definición).
- **Número combinatorio** m **sobre** n:

$$\binom{m}{n} = \frac{m!}{n!(m-n)!}.$$

Problema: Sean $\{a_1, \ldots, a_m\}$ m elementos.

¿Cuántos grupos de n elementos pueden formarse con $\{a_1, \ldots, a_m\}$?

Problema análogo: Disponemos de una urna con m bolas numeradas. Extraemos sucesivamente n de ellas, es decir, tomamos una muestra de tamaño n.

¿Cuántos muestras posibles de tamaño n podemos extraer?

Problema: Disponemos de una urna con m bolas numeradas. Extraemos sucesivamente n de ellas, es decir, tomamos una muestra de tamaño n.

¿Cuántos muestras posibles de tamaño n podemos extraer?

Hay que precisar dos cuestiones:

- **Tipos de muestreo:**
  (1) **Sin repetición** (sin reemplazamiento): no podemos sacar dos veces la misma bola.
  (2) **Con repetición** (con reemplazamiento): podemos sacar más de una vez la misma bola.

- **Tipos de muestras:**
  (A) **Variaciones**: Importa el orden $((1,2) \neq (2,1))$.
  (B) **Combinaciones**: No importa el orden $((1,2) = (2,1))$.

(1) **Solución:** $\{a_1, \ldots, a_m\}$ m elementos. ¿Cuántos grupos de n elementos pueden formarse sin reemplazamiento?

(1-A) **Variaciones ordinarias:** Un grupo es distinto de otro atendiendo a los elementos y a su orden. *Variaciones de m elementos tomadas de n en n.*

$$V_m^n = \frac{m!}{(m-n)!} = m \cdot (m-1) \cdots (m-n+1).$$

(Este número combinatorio también se conoce como el *factorial descendente*, $m^{\underline{n}} = m \cdot (m-1) \cdots (m-n+1)$.)

Ejemplo: $m = 4$, $n = 2$, $\{a_1, a_2, a_3, a_4\}$.

$$\begin{array}{cccc} (a_1, a_2) & (a_2, a_1) & (a_3, a_1) & (a_4, a_1) \\ (a_1, a_3) & (a_2, a_3) & (a_3, a_2) & (a_4, a_2) \\ (a_1, a_4) & (a_2, a_4) & (a_3, a_4) & (a_4, a_3) \end{array}$$

$$V_4^2 = \frac{4!}{2} = 4 \cdot 3 = 12, \qquad ((a_1, a_2) \neq (a_2, a_1)).$$

(1) **Solución:** $\{a_1, \ldots, a_m\}$ m elementos. ¿Cuántos grupos de n elementos pueden formarse sin reemplazamiento?

(1-A) **Permutaciones:** Caso $n = m$ (Variaciones de m elementos tomados de m en m). *Permutaciones de m elementos*.

$$P_m = V_m^m = m! = m \cdot (m-1) \cdots 2 \cdot 1.$$

Ejemplo: $m = 3$, $n = 3$, $\{a_1, a_2, a_3\}$.

$$\begin{array}{ccc} (a_1, a_2, a_3) & (a_2, a_3, a_1) & (a_2, a_1, a_3) \\ (a_1, a_3, a_2) & (a_3, a_1, a_2) & (a_3, a_2, a_1) \end{array}$$

$$P_3 = 3! = 3 \cdot 2 = 6, \qquad ((a_1, a_2, a_3) \neq (a_2, a_1, a_3)).$$

(1) **Solución:** $\{a_1, \dots, a_m\}$ m elementos. ¿Cuántos grupos de n elementos pueden formarse sin reemplazamiento?

(1-B) **Combinaciones ordinarias:** Un grupo es distinto de otro atendiendo a los elementos. *Combinaciones de m elementos tomadas de n en n.*

$$C_m^n = \binom{m}{n}.$$

Ejemplo: $m = 4$, $n = 2$, $\{a_1, a_2, a_3, a_4\}$.

$$\begin{array}{lll} (a_1, a_2) & & \\ (a_1, a_3) & (a_2, a_3) & \\ (a_1, a_4) & (a_2, a_4) & (a_3, a_4) \end{array}$$

$$C_4^2 = \binom{4}{2} = \frac{4!}{2!2!} = 6, \qquad ((a_1, a_2) = (a_2, a_1)).$$

(2) **Solución:** $\{a_1, \ldots, a_m\}$ m elementos. ¿Cuántos grupos de n elementos pueden formarse con reemplazamiento?

(2-A) **Variaciones con repetición:** Un grupo es distinto de otro atendiendo a los elementos y a su orden y admitimos repetición. *Variaciones con repetición de m elementos tomadas de n en n.*

$$VR_m^n = m^n.$$

Ejemplo: $m = 3$, $n = 2$, $\{a_1, a_2, a_3\}$.

$$\begin{array}{ccc} (a_1, a_1) & (a_2, a_1) & (a_3, a_1) \\ (a_1, a_2) & (a_2, a_2) & (a_3, a_2) \\ (a_1, a_3) & (a_2, a_3) & (a_3, a_3) \end{array}$$

$$VR_3^2 = 3^2 = 9.$$

(2) **Solución:** $\{a_1, \ldots, a_m\}$ m elementos. ¿Cuántos grupos de n elementos pueden formarse con reemplazamiento?

(2-B) **Combinaciones con repetición:** Un grupo es distinto de otro atendiendo a los elementos y admitimos repetición.
Combinaciones con repetición de m elementos tomadas de n en n.

$$CR_m^n = \binom{m+n-1}{n} = \frac{(m+n-1)!}{n!(m-1)!}.$$

Ejemplo: $m = 3$, $n = 2$, $\{a_1, a_2, a_3\}$.

$$\begin{array}{lll} (a_1, a_1) & & \\ (a_1, a_2) & (a_2, a_2) & \\ (a_1, a_3) & (a_2, a_3) & (a_3, a_3) \end{array}$$

$$CR_3^2 = \binom{4}{2} = 6.$$

Resumen

	Muestras ordenadas	Muestras sin ordenar
Sin reemplazamiento	V_m^n	C_m^n
Con reemplazamiento	VR_m^n	CR_m^n

Otro problema: Supongamos que tenemos n elementos de los cuales $n_1, n_2, \dots, n_m$ son iguales $(n_1 + \cdots + n_m = n)$. Es decir,

$$\{\overbrace{a_1, \dots, a_1}^{n_1}, \overbrace{a_2, \dots, a_2}^{n_2}, \dots, \overbrace{a_m, \dots, a_m}^{n_m}\}.$$

¿Cuántos grupos de n elementos pueden formarse atendiendo al orden?

Permutaciones con repetición:

$$P_n^{n_1,\dots,n_m} = \frac{n!}{n_1!n_2!\cdots n_m!}.$$

Ejemplo: $n_1 = 3$, $n_2 = 2$ $(n = n_1 + n_2 = 5)$, $\{a_1, a_1, a_1, a_2, a_2\}$.

$$\begin{array}{lll}
(a_1, a_1, a_1, a_2, a_2) & (a_1, a_2, a_1, a_2, a_1) & (a_2, a_1, a_2, a_1, a_1) \\
(a_1, a_1, a_2, a_1, a_2) & (a_1, a_2, a_2, a_1, a_1) & (a_2, a_2, a_1, a_1, a_1) \\
(a_1, a_1, a_2, a_2, a_1) & (a_2, a_1, a_1, a_1, a_2) & \\
(a_1, a_2, a_1, a_1, a_2) & (a_2, a_1, a_1, a_2, a_1) &
\end{array}$$

$$P_5^{3,2} = \frac{5!}{3!2!} = 10.$$

Ejercicio: De todas las sucesiones de longitud n compuestas por las cifras 0, 1 y 2 se elige una al azar. Hallar la probabilidad de los sucesos siguientes:

(a) La sucesión comienza con 0.

(b) La sucesión contiene exactamente m unos.

Probabilidad I

Grado en Matemáticas

Tema 2
Variables aleatorias

Javier Cárcamo

Departamento de Matemáticas
Universidad Autónoma de Madrid
javier.carcamo@uam.es

Descripción del tema

1. Variables aleatorias.
2. Función de distribución.
3. Variables discretas.
4. Variables continuas.
5. Transformaciones de variables aleatorias.
6. Algunas distribuciones discretas notables.
7. Algunas distribuciones continuas notables.

Objetivos principales

- Comprender la noción de variable aleatoria.
- Manejar con soltura las funciones de distribución y de densidad.
- Calcular la distribución de una transformación de una variable en los casos más sencillos.
- Identificar y saber utilizar las distribuciones más importantes.

Observación: Cuando observamos un experimento aleatorio es frecuente que nos interese una característica concreta del resultado, más que el resultado en sí mismo. Por ejemplo, la suma de las puntuaciones al lanzar dos dados o el número de caras que hemos obtenido al lanzar una moneda varias veces.

Una **variable** es cualquier aspecto relativo a un experimento. Si $(\Omega, \mathcal{F}, \mathrm{P})$ es un espacio de probabilidad, podemos pensar en una aplicación:

$$\begin{aligned} X : \Omega &\longrightarrow E \\ \omega &\longmapsto X(\omega). \end{aligned}$$

Las variables pueden ser cuantitativas o cualitativas. Nos centraremos en las **variables cuantitativas** ($E \subseteq \mathbb{R}$). Se denomina variable *aleatoria* ya que a priori no conocemos el resultado del experimento, luego no sabemos el valor que va a tomar X.

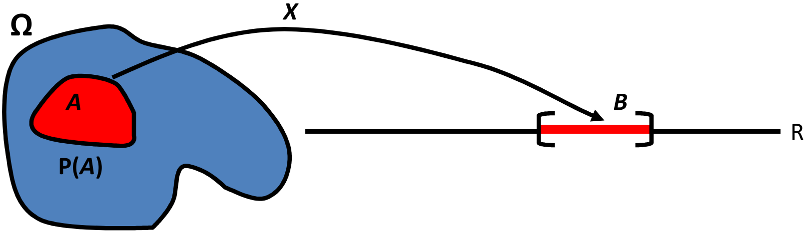
Dado $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad, se llama **variable aleatoria** (v.a.) a toda aplicación:

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \quad (\text{o } \overline{\mathbb{R}}) \\ \omega &\longmapsto X(\omega). \end{aligned}$$

Observación: Realmente, para que X sea una v.a. debe cumplir una condición de *medibilidad*: Para todo $a \in \mathbb{R}$, se pide que el conjunto $\{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}$. Cuando $\mathcal{F} = \mathcal{P}(\Omega)$, esto siempre se cumple.

Idea: Como P es medida de probabilidad sobre $(\Omega, \mathcal{F})$, sabemos "calcular probabilidades" en Ω vía P. Al tener una aplicación $X : \Omega \longrightarrow \mathbb{R}$ vamos tener una forma de "medir" en $\mathbb{R}$.

La medida de probabilidad P sobre $(\Omega, \mathcal{F})$ induce una *nueva* medida de probabilidad sobre $\mathbb{R}$, P_X, que llamaremos **distribución de probabilidad** de la variable aleatoria X.



Notación: $\{X \in B\} = X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}, \quad B \subset \mathbb{R}.$

$X : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}$ variable aleatoria. Consideramos la aplicación

$$\begin{aligned} \mathrm{P}_X : \mathcal{B} \subset \mathcal{P}(\mathbb{R}) &\longrightarrow [0,1] \\ B &\longmapsto \mathrm{P}_X(B) = \mathrm{P}(X \in B). \end{aligned}$$

P_X se llama **distribución de probabilidad** de X y es una medida de probabilidad en $\mathbb{R}$.

Nota: $\mathcal{B}$ es una σ-álgebra sobre $\mathbb{R}$ (σ-álgebra de Borel) que incluye todos los conjuntos que nos pueden interesar para calcular probabilidades.

Ejemplos elementales

- $X \equiv$ escaños de un partido en unas elecciones.
- $X \equiv$ peso de una persona (elegida aleatoriamente).
- $X \equiv$ altura de una persona (elegida aleatoriamente).
- $X \equiv$ nota de Probabilidad de un estudiante.
- $X \equiv$ horas de estudio de Probabilidad de un estudiante.
- $X \equiv$ beneficios anuales de una empresa.
- $X \equiv$ valor en bolsa de una compañía.
- $X \equiv$ tiempo de procesado de un programa informático.
- $X \equiv$ vida útil de una componente de un sistema.
- $X \equiv$ tiempo entre la llegada de un mail y su respuesta.
- $X \equiv$ número de visitas al día de una página web.
- $X \equiv$ número de errores de código de un programador.
- $X \equiv \cdots\cdots$

Ejemplo: ϵ: lanzar una moneda al aire 3 veces.

$$\Omega = \{(C, C, C), (C, C, +), \ldots, (+, +, +)\}.$$

(Ω, P): modelo de Laplace (equiprobabilidad). $\mathrm{Card}(\Omega) = 2^3 = 8$.

$$\mathrm{P}(\{(C, C, C)\}) = \mathrm{P}(\{(C, C, +)\}) = \cdots = \mathrm{P}(\{(+, +, +)\}) = 1/8.$$

Consideramos la v.a. $X : \Omega \longrightarrow \mathbb{R}$ que mide el número de caras que hemos obtenido.

$$\begin{array}{rl} X : \Omega & \longrightarrow \mathbb{R} \\ (C, C, C) & \longmapsto 3 \\ \\ (C, C, +) & \longmapsto 2 \\ (C, +, C) & \longmapsto 2 \\ (+, C, C) & \longmapsto 2 \end{array} \qquad \begin{array}{rl} X : \Omega & \longrightarrow \mathbb{R} \\ (+, +, +) & \longmapsto 0 \\ \\ (C, +, +) & \longmapsto 1 \\ (+, C, +) & \longmapsto 1 \\ (+, +, C) & \longmapsto 1 \end{array}$$

Pregunta: ¿Cuál es la distribución de probabilidad de X?

X toma los valores $\{0,1,2,3\}$ con probabilidad:

$$\begin{aligned}
\mathrm{P}_X(\{0\}) &= \mathrm{P}(X=0) = \mathrm{P}(\{\omega \in \Omega : X(\omega)=0\}) \\
&= \mathrm{P}(\{(+,+,+)\}) = 1/8.
\end{aligned}$$

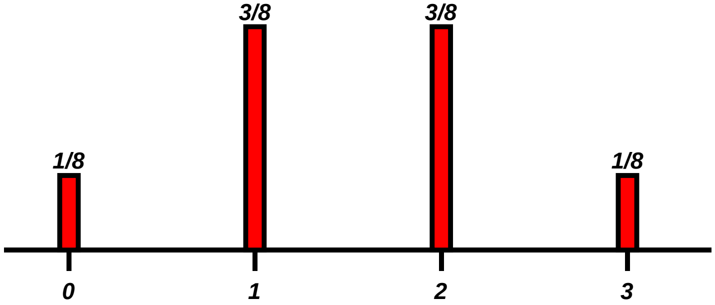
$$\begin{aligned}
\mathrm{P}_X(\{1\}) &= \mathrm{P}(X=1) = \mathrm{P}(\{\omega \in \Omega : X(\omega)=1\}) \\
&= \mathrm{P}(\{(C,+,+),(+,C,+),(+,+,C)\}) \\
&= \mathrm{P}(\{(C,+,+)\}) + \mathrm{P}(\{(+,C,+)\}) + \mathrm{P}(\{(+,+,C)\}) \\
&= 1/8+1/8+1/8 = 3/8.
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}_X(\{2\}) &= \mathrm{P}(X=2) = \mathrm{P}(\{\omega \in \Omega : X(\omega)=2\}) \\
&= \mathrm{P}(\{(C,C,+),(C,+,C),(+,C,C)\}) \\
&= \mathrm{P}(\{(C,C,+)\}) + \mathrm{P}(\{(C,+,C)\}) + \mathrm{P}(\{(+,C,C)\}) \\
&= 1/8+1/8+1/8 = 3/8.
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}_X(\{3\}) &= P(X=3) = \mathrm{P}(\{\omega \in \Omega : X(\omega)=3\}) \\
&= \mathrm{P}(\{(C,C,C)\}) = 1/8.
\end{aligned}$$

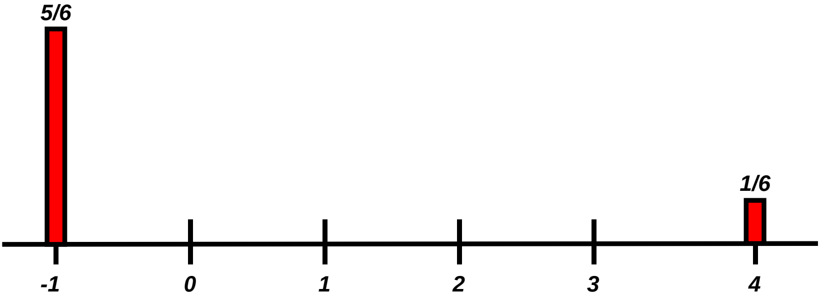
$X \equiv$ número de caras al lanzar 3 monedas.

X toma los valores 0,1,2 y 3 con probabilidades $1/8$, $3/8$, $3/8$ y $1/8$, resp.



Ejemplo: *Craps* es un juego de dados en el que se pueden hacer una gran diversidad de apuestas en el lanzamiento de dos dados (una o más veces). Una de las más sencillas es la llamada *Any seven* en la que se apuesta en un solo lanzamiento a que se obtienen un total de 7 puntos. La apuesta paga 4 a 1.

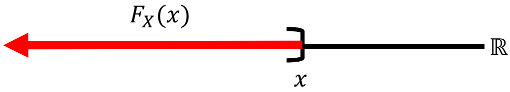
La variable G (ganancias) toma los valores -1, 4, con probabilidades $5/6$, $1/6$, respectivamente.



Sea $X : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}$ variable aleatoria. Se llama **función de distribución** de X a la función F_X definida mediante:

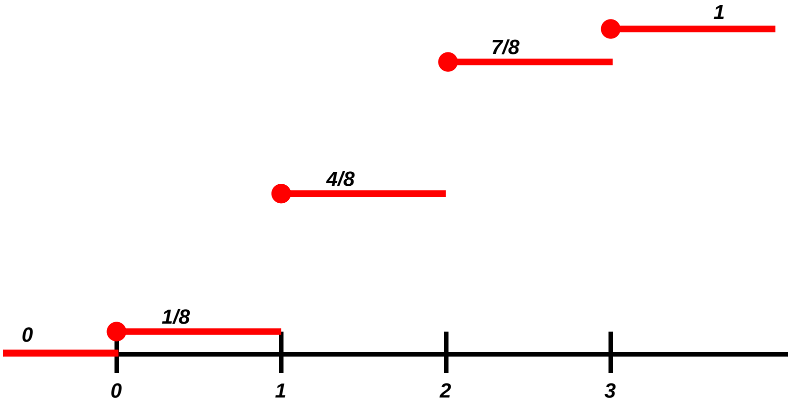
$$\begin{aligned} F_X : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto F_X(x) = \mathrm{P}(X \leq x) = \mathrm{P}_X((-\infty, x]). \end{aligned}$$

Idea: Para cualquier $x \in \mathbb{R}$, $F_X(x)$ mide la masa de probabilidad que la variable X concentra a la izquierda del punto x.



Notación: Si no hay duda, denotaremos simplemente por F (en vez de F_X) la función de distribución de la variable aleatoria X.

Ejemplo: $X \equiv$ número de caras al lanzar tres monedas.



Observación: La función de distribución de X es constante a trozos y "salta" en los puntos donde X concentra masa de probabilidad.

Propiedades básicas de la función de distribución

Sea X variable aleatoria y F su función de distribución.

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

Teorema de unicidad

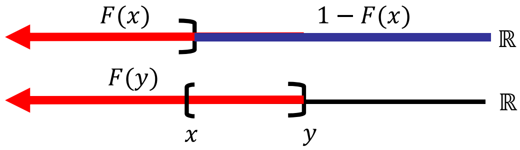
Sea $F : \mathbb{R} \longrightarrow \mathbb{R}$ verificando 1, 2 y 3 de arriba, entonces, existe $X : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}$ variable aleatoria tal que $F = F_X$.

Además, la variable X es única en distribución (es decir, la f.d. caracteriza la distribución de probabilidad de una v.a. de forma biunívoca).

Cálculo de probabilidades con la función de distribución

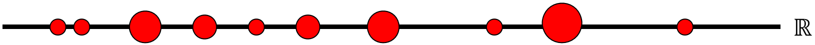
Sea $x, y \in \mathbb{R}$ con $x < y$. Se tiene:

1. $\mathrm{P}(X \leq x) = F(x)$.
2. $\mathrm{P}(X > x) = 1 - F(x)$.
3. $\mathrm{P}(X < x) = F(x^-)$.
4. $\mathrm{P}(X \geq x) = 1 - F(x^-)$.
5. $\mathrm{P}(x < X \leq y) = F(y) - F(x)$.
6. $\mathrm{P}(x \leq X \leq y) = F(y) - F(x^-)$.
7. $\mathrm{P}(x \leq X < y) = F(y^-) - F(x^-)$.
8. $\mathrm{P}(x < X < y) = F(y^-) - F(x)$.
9. $\mathrm{P}(X = x) = F(x) - F(x^-)$.



3. Variables discretas

Una variable X se dice que tiene **distribución discreta** si existe un conjunto $S \subset \mathbb{R}$ contable (finito o infinito numerable) tal que $\mathrm{P}(X \in S) = 1$ (o, alternativamente, $\mathrm{P}(X \in S^c) = 0$). Es decir, X concentra su masa en S. El conjunto S se llama **soporte de la distribución** (de X).



Nota: X variable discreta con soporte $S \subset \mathbb{R}$, entonces

$$\mathrm{P}(X \in B) = \sum_{s \in S \cap B} \mathrm{P}(X = s), \quad B \subset \mathbb{R}.$$

Ejemplos:

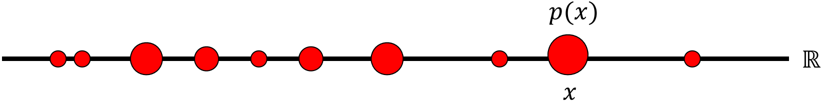
- $X \equiv$ número de caras al lanzar 3 monedas: $S = \{0, 1, 2, 3\}$.
- $X \equiv$ puntuación de un dado: $S = \{1, 2, \ldots, 6\}$.
- $X \equiv$ número de personas esperando en una cola: $S = \{0, 1, 2, \ldots\} = \mathbb{N} \cup \{0\}$.

Sea X es una variable discreta con soporte S. La función:

$$\begin{aligned} p_X : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto p_X(x) = \mathrm{P}(X = x) \end{aligned}$$

se denomina **función (de masa de) probabilidad** de X. Es decir, $p_X(x)$ es la masa de probabilidad que X concentra en el punto x.

Notación: Cuando no haya confusión denotaremos por p (en lugar de p_X) la función de probabilidad de la variable X.

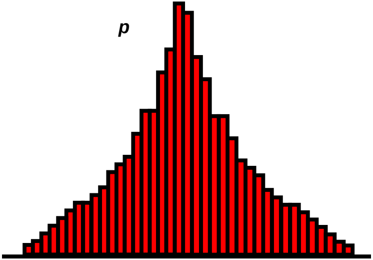


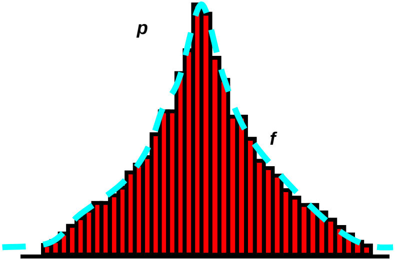
Si X es variable discreta con función de probabilidad p, tenemos:

- $p(x) = 0$, si $x \in S^c$ y $\sum_{x \in S} p(x) = 1$.
- $\mathrm{P}(X \in A) = \sum_{x \in A \cap S} p(x)$ (suma finita o serie abs. conv.).

¿Como es la función de distribución de una variable discreta?

Idea intuitiva: La función de probabilidad de una variable discreta que toma un número muy grande de valores muy próximos entre sí suele ser similar a una función real positiva.



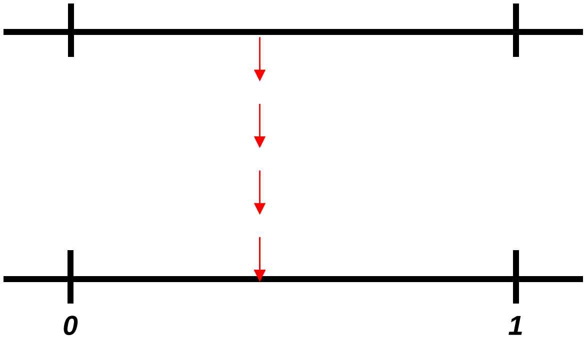


Además, hay muchas variables que pueden tomar un número infinito no numerable de valores. Por ejemplo:

- $X \equiv$ peso de una persona.
- $X \equiv$ tiempo de duración de un suceso.
- $X \equiv$ tiempo de espera en una cola.

Ejemplo: $X \equiv$ elegir un punto "al azar" en el intervalo $[0, 1]$.

Podemos pensar en dejar caer un alfiler que se mueve sobre una barra de longitud 1.



Pregunta: Intuitivamente, ¿cuál será la probabilidad de que el alfiler caiga exactamente en el punto $1/2$? ¿y en el punto x, donde x es un punto cualquiera de $[0, 1]$?

Una función $f : \mathbb{R} \longrightarrow \mathbb{R}$ se llama **función de densidad (de probabilidad)** sobre $\mathbb{R}$ si cumple:

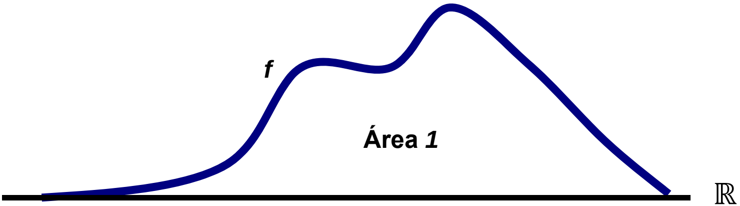
(a) $f(x) \geq 0$, para todo $x \in \mathbb{R}$.

(b) f es integrable (Riemman).

(c) $\displaystyle\int_{-\infty}^{+\infty} f(t)\, dt = 1$.

Nota: En la práctica, la condición (b) de arriba se puede sustituir por la más exigente:

(b′) f tiene un número finito de discontinuidades.



Una variable aleatoria X con función de distribución F se dice **(absolutamente) continua** si existe una función de densidad f tal que

$$F(x) = \int_{-\infty}^{x} f(t)\, dt, \qquad x \in \mathbb{R}. \qquad (*)$$

Observaciónes:

- Si X es continua, entonces $\mathrm{P}(X = x) = 0$, para todo $x \in \mathbb{R}$.
- Si X es una v.a. continua, entonces F es una función continua.
- Cada variable aleatoria continua X tiene asociada una función de densidad.

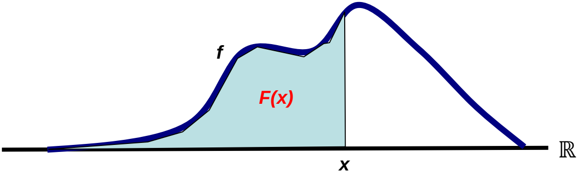
$$X \text{ v.a. continua} \Longleftrightarrow f \text{ densidad.}$$

- De $(*)$, y usando el teorema fundamental del cálculo, se tiene que en cada punto de continuidad x de f

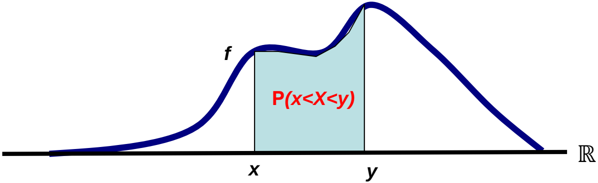
$$\frac{d}{dx} F(x) = f(x).$$

Cálculo de probabilidades con variables continuas

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



$$\mathrm{P}(x < X \leq y) = \mathrm{P}(x \leq X < y) = \mathrm{P}(x \leq X \leq y)$$
$$= F(y) - F(x) = \int_{x}^{y} f(t)\,dt$$



Ejemplo: Seleccionar un punto "al azar" en el intervalo $[0,1]$. La densidad de la variable que nos da el punto seleccionado es

$$f(x) = \begin{cases} 1, & \text{si } x \in [0,1], \\ 0, & \text{si } x \notin [0,1]. \end{cases}$$

Ejercicio: Consideremos la función

$$f(x) = \begin{cases} \frac{3}{2}x^2, & \text{si } x \in [-1,1], \\ 0, & \text{si } x \notin [-1,1]. \end{cases}$$

(a) Muéstrese que f es una función de densidad.

(b) Calcúlese la función de distribución F de la variable X que define la densidad f.

(c) Calcúlense las siguientes probabilidades:

$$\mathrm{P}(-1 \le X \le 0), \quad \mathrm{P}(X \le 1/2) \quad \text{y} \quad \mathrm{P}(-1 \le X \le 1).$$

Tras definir el concepto de variable aleatoria, hemos estudiado dos grandes grupos:

- Variables aleatorias discretas.
- Variables aleatorias continuas.

Sin embargo, hay variables que no pertenecen a ninguno de estos dos grupos.

Ejemplo: Basta considerar una función de distribución que tenga a la vez discontinuidades de salto y tramos continuos estrictamente crecientes:

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{4} + \frac{1}{2}x, & 0 \leq x < 1, \\ 1, & 1 \leq x. \end{cases}$$

En este ejemplo, la distribución es una mezcla de una variable continua y otra discreta, pero también existe otro gran grupo de distribuciones de probabilidad (*distribuciones continuas singulares*) que no estudiaremos en este curso.

Dada una v.a. $X : \Omega \longrightarrow \mathbb{R}$ y una función (continua) $g : \mathbb{R} \longrightarrow \mathbb{R}$, la aplicación:

$$\begin{aligned} Y = g(X) : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto Y(\omega) = g(X(\omega)), \end{aligned}$$

es una v.a. que se denomina **transformada de X a través de g**.

Ejemplo: X^2, e^X, $\log X$ $(X > 0)$, $X^3 + 1$, $\sin(X)$ son v.a.

Problema: Calcular la distribución de probabilidad de Y a partir de la de X.

Método de las funciones de distribución:

Expresar la f.d. de Y, F_Y, mediante la f.d. de X, F_X:

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(g(X) \leq y) = \mathrm{P}(X \in A_y),$$

donde $A_y = \{x \in \mathbb{R} : g(x) \leq y\}$.

Ejemplo (transformaciones lineales): La v.a. $Y = aX + b$, con $a, b \in \mathbb{R}$, se llama **transformada lineal de** X.

$$F_Y(y) = \begin{cases} F_X\left(\frac{y-b}{a}\right), & \text{si } a > 0, \\ 1 - F_X\left(\left(\frac{y-b}{a}\right)^{-}\right), & \text{si } a < 0. \end{cases}$$

Caso discreto: Si X es discreta con f.p. p_X e $Y = g(X)$, se tiene:

- $p_Y(y) = \mathrm{P}(Y = y) = \displaystyle\sum_{x:g(x)=y} p_X(x)$ f.p. de Y.
- $F_Y(y) = \mathrm{P}(Y \leq y) = \displaystyle\sum_{x:g(x)\leq y} p_X(x)$ f.d. de Y.

Caso continuo: Si X es v.a. continua con f.d. f_X e $Y = g(X)$, con g continua y monótona (creciente o decreciente) se tiene:

$$F_Y(y) = \begin{cases} F_X\left(g^{-1}(y)\right) & \text{si } g \text{ creciente} \\ 1 - F_X\left(g^{-1}(y)\right) & \text{si } g \text{ decreciente.} \end{cases}$$

En cualquier caso, Y es continua con densidad:

$$f_Y(y) = f_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|.$$

Ejercicio: Sea $X \sim \mathrm{U}(0,1)$, es decir, X tiene densidad

$$f(x) = \begin{cases} 1, & \text{si } 0 \leq x \leq 1, \\ 0, & \text{en otro caso.} \end{cases}$$

Calcular la distribución de

- $Y = \log \frac{1}{X}$.
- $Y = a + (b-a)X$, con $a < b$.

Ejercicio: Si X tiene densidad

$$f(x) = \begin{cases} \frac{3}{2}x^2, & \text{si } x \in [-1,1], \\ 0, & \text{si } x \notin [-1,1]. \end{cases}$$

- Calcular la densidad de $Y = e^X$.
- Calcular la densidad de $Y = X^2$.

1. **Distribución degenerada en un punto a**

$$\mathrm{P}(X = a) = 1.$$

2. **Distribución de Bernoulli de parámetro $p \in (0, 1)$**

$$\mathrm{P}(X = 0) = q = 1 - p, \quad \mathrm{P}(X = 1) = p.$$

Aplicaciones: Con esta distribución se modelizan fenómenos que sólo pueden presentar dos resultados posibles (dicotómicos). El valor 1 se identifica con el éxito y 0 con el fracaso. Es la base de la distribución binomial y de otras muchas distribuciones discretas importantes.
Notación: $X \sim \mathrm{B}(1; p)$.

3. **Distribución uniforme en n puntos $a_1, \ldots, a_n$**

$$\mathrm{P}(X = a_i) = \frac{1}{n}, \quad i = 1, \ldots, n.$$

Notación: $X \sim \mathrm{U}(\{a_1, \ldots, a_n\})$.

4. **Distribución binomial de parámetros** n, p

$$\mathrm{P}(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \ldots, n.$$

Aplicaciones: Sea $n \in \mathbb{N}$ y $p \in (0, 1)$. Consideremos un experimento ϵ en el que prestamos atención a un suceso A (éxito) y A^c (fracaso), con $\mathrm{P}(A) = p$. Repetimos n veces el experimento en pruebas independientes, es decir, bajo las mismas condiciones. Estas repeticiones se denominan **pruebas (independientes) de Bernoulli**.

La variable aleatoria

$$X \equiv \text{ número de éxitos en los } n \text{ intentos}$$

se dice que tiene distribución **binomial** (de parámetros n y p) y toma los valores $0, 1, \ldots, n$.

Notación: $X \sim \mathrm{B}(n; p)$.

4. **Distribución binomial de parámetros n, p**

$$\mathrm{P}(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \ldots, n.$$

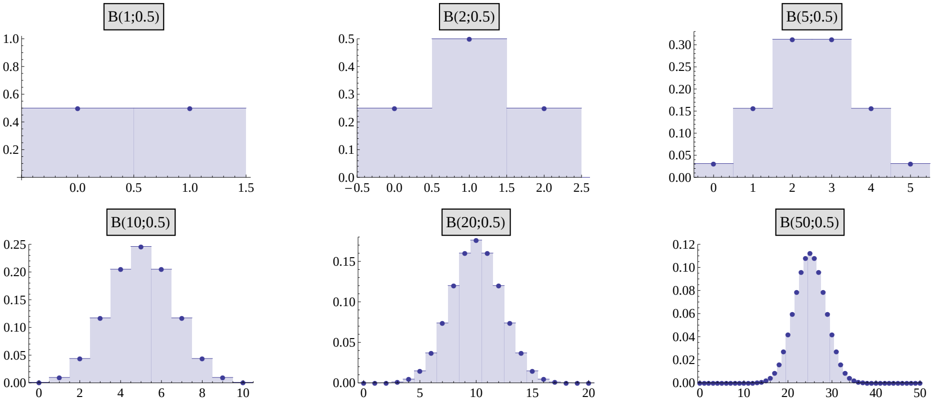
Suma de variables de Bernoulli independientes:
Consideremos $X_1, \ldots, X_n$ v.a. $B(1, p)$ independientes, donde X_i toma el valor 1 si en la i-ésima prueba de Bernoulli ocurre el suceso A (éxito). Tenemos,

$$X = X_1 + \cdots + X_n, \quad X_i \sim \mathrm{B}(1; p) \quad (i = 1, \ldots, n).$$

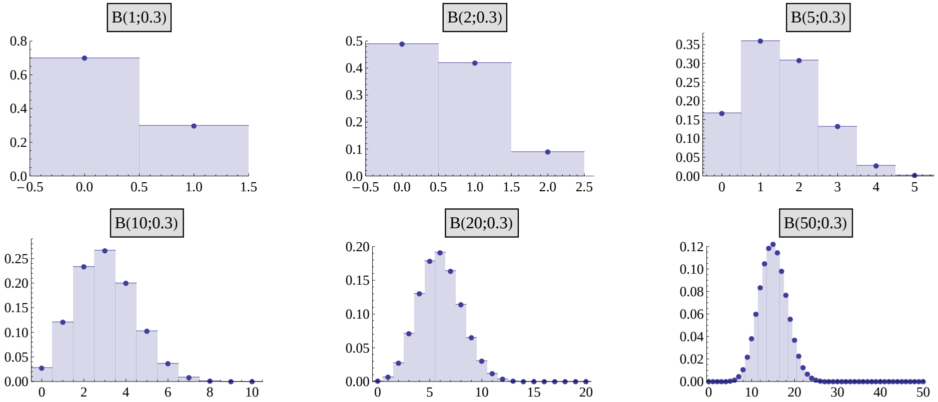
Luego, la v.a. con distribución binomial de parámetros n y p se puede expresar como suma de n v.a.s (independientes) de Bernoulli de parámetro p.

Suma de binomiales independientes:
Si $X_1, \ldots, X_k$ v.a. (independientes) con $X_i \sim \mathrm{B}(n_i; p)$ $(1 \leq i \leq n)$, la v.a. $X = X_1 + \cdots + X_k \sim \mathrm{B}(n; p)$, con $n = n_1 + \cdots + n_k$.

Función de probabilidad de B$(n; 0{,}5)$



Función de probabilidad de B$(n; 0{,}3)$



Ejercicio (uso de la tabla de probabilidades binomiales): El dueño de una fotocopiadora afirma que el 20 % de todos los errores son provocados por los clientes (que no siguen las instrucciones de operación correctamente). Si esto es cierto:

(a) ¿Cuál es la probabilidad de que entre 20 errores cometidos, más de 4 hayan sido ocasionados por los clientes que no siguen las instrucciones?

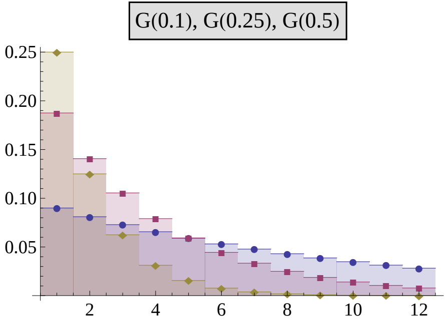
(b) ¿Cuál es la probabilidad de que entre 20 errores cometidos, menos de 15 hayan sido provocados por la máquina?

5 **Distribución geométrica de parámetro** $p \in (0,1)$

$$\mathrm{P}(X = k) = pq^k, \quad k = 0, 1, \ldots$$

Aplicaciones: Esta distribución aparece al considerar una variable aleatoria que representa el número de unidades de tiempo que es necesario esperar hasta la ocurrencia, por primera vez, de un determinado acontecimiento aleatorio.

Notación: $X \sim \mathrm{G}(p)$.



5. **Distribución geométrica de parámetro** $p \in (0,1)$

Nota: En algunas ocasiones interesan conocer la distribución de la variable Y que cuenta el número de intentos hasta el primer éxito. Los intentos se suponen independientes unos de otros y la probabilidad de éxito es p. En este caso, la variable de interés es $Y = X + 1$, donde $X \sim \mathrm{G}(p)$, es decir, Y verifica

$$\mathrm{P}(Y = k) = pq^{k-1}, \quad k = 1, 2, \ldots$$

Dependiendo de la literatura consultada, en algunas ocasiones la distribución de la variable Y también se conoce como geométrica de parámetro p. Por tanto, en los ejemplos prácticos conviene saber cuál de las dos distribuciones nos interesa.

6. **Distribución binomial negativa (parámetros $t \in \mathbb{N}$, $p \in (0,1)$)**

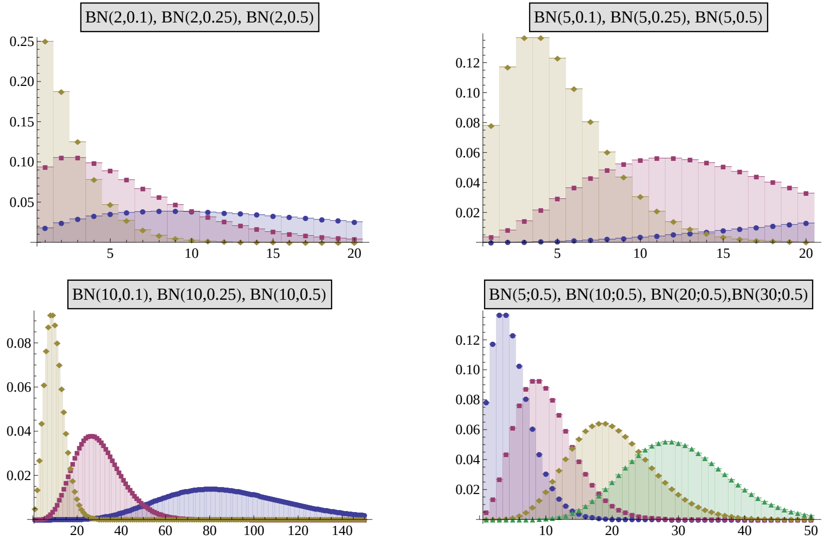
$$\mathrm{P}(X = k) = \binom{t+k-1}{k} p^t q^k, \quad k = 0, 1, \ldots$$

Aplicaciones: Esta distribución es un modelo discreto de tiempo de espera: En una sucesión de experimentos independientes de Bernoulli con probabilidad de éxito p, la distribución número de fracasos antes de obtener t éxitos es binomial negativa de parámetros t, p.

Notación: $X \sim \mathrm{BN}(t; p)$.

Observación: Si $t = 1$, es geométrica de parámetro p ($\mathrm{BN}(1; p) \sim \mathrm{G}(p)$).

Función de probabilidad de BN$(t; p)$



7. **Distribución de Poisson de parámetro** $\lambda > 0$

$$\mathrm{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \ldots$$

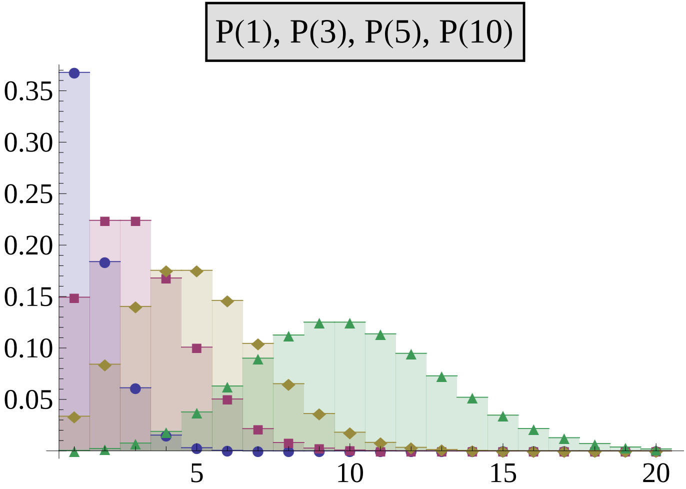
Aplicaciones: Se utiliza como modelo probabilístico para el estudio de fenómenos en los que ocurren determinados sucesos por unidad de tiempo, espacio, volumen, área, etc.

- El número de plaquetas en un ml. de sangre.
- El número de mutaciones en un fragmento de ADN después de una cierta cantidad de radiación.
- Número de personas que llegan en un intervalo de tiempo a una cola.

Notación: $X \sim \mathrm{P}(\lambda)$.

Pregunta: ¿Por qué la distribución de Poisson es la adecuada para modelizar este tipo de fenómenos?

Función de probabilidad de P(λ)



Aproximación de Poisson a la binomial:

$$\binom{n}{k} p^k (1-p)^{n-k} \approx e^{-\lambda}\frac{\lambda^k}{k!}, \quad n \to \infty,\ np \to \lambda.$$

Se dice que la binomial **converge en distribución** a una v.a. $\mathrm{P}(\lambda)$.

Condiciones para la aproximación binomial a Poisson:

Dada $X \sim \mathrm{B}(n; p)$. Para calcular

$$\mathrm{P}(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \ldots, n,$$

muchas veces es útil utilizar la aproximación anterior, es decir,

$$\mathrm{P}(X = k) \approx \mathrm{P}(Y = k), \quad \text{donde } Y \sim \mathrm{P}(np).$$

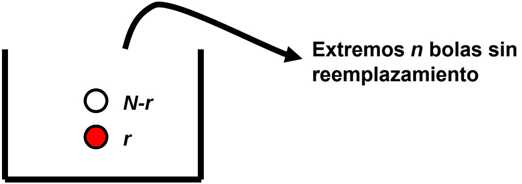
Para que el error cometido en la aproximación no sea muy grande se suele pedir: $n \geq 30$; $\pi \leq 0,1$; $n\pi < 18$, con $\pi = \min\{p, 1-p\}$.

Ejercicio: La probabilidad de que un petrolero tenga un accidente y genere una marea negra es 0.0001. Calcular la probabilidad de que entre 1000 petroleros haya al menos un accidente.

8 **Distribución hipergeométrica ($N, n, r \in \mathbb{N}$, con $n, r \leq N$)**

$$\mathrm{P}(X = k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}, \quad \text{máx}\{0, n - N + r\} \leq k \leq \text{mín}\{n, r\}.$$

Aplicaciones: Consideremos la extracción (sin reemplazamiento) de n bolas de una urna que contiene N bolas, de las cuales r son rojas y $N - r$ blancas.

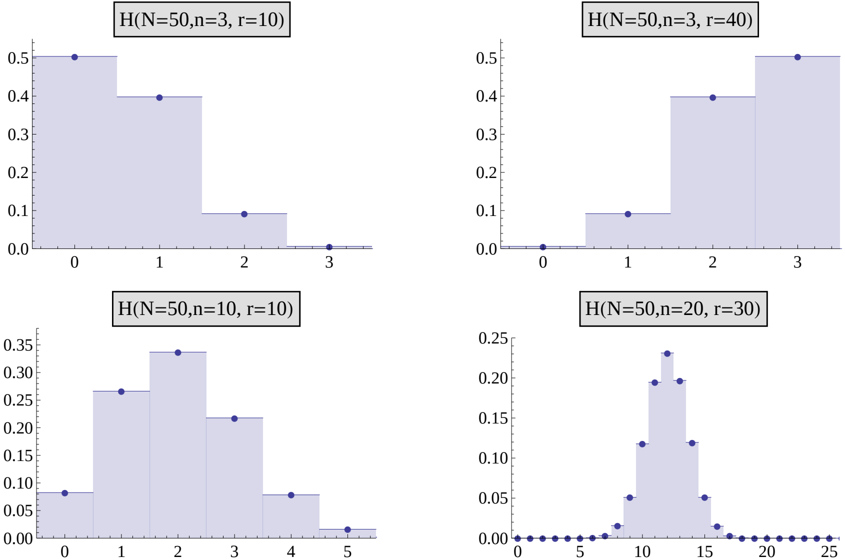


La variable X que cuenta el número de bolas rojas extraídas tiene distribución hipergeométrica de parámetros N, n y r.

Se utiliza habitualmente en Control de calidad.

Notación: $X \sim \mathrm{H}(N; n; r)$.

Función de probabilidad de $\mathbf{H}(N; n; r)$



1. **Distribución uniforme en el intervalo** (a, b)

$$f(x) = \frac{1}{b-a} 1_{(a,b)}(x).$$

Aplicaciones: Se utiliza en los métodos de generación de números (pseudo-)aleatorios. Supongamos que queremos generar observaciones aleatorias de una variable con función de distribución F (continua). Primeramente se generan números de una variable U uniforme estándar (i.e. con $a = 0$, $b = 1$) y se transforman con F^{-1}. Esta aplicación se basa en la siguiente propiedad.

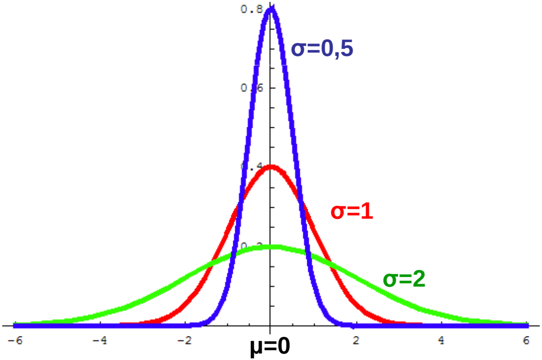
Propiedad: generación de números aleatorios

Sea X variable aleatoria con función de distribución F continua. La variable $U = F(X)$ tiene distribución uniforme estándar.

Notación: $X \sim \mathrm{U}(a, b)$.

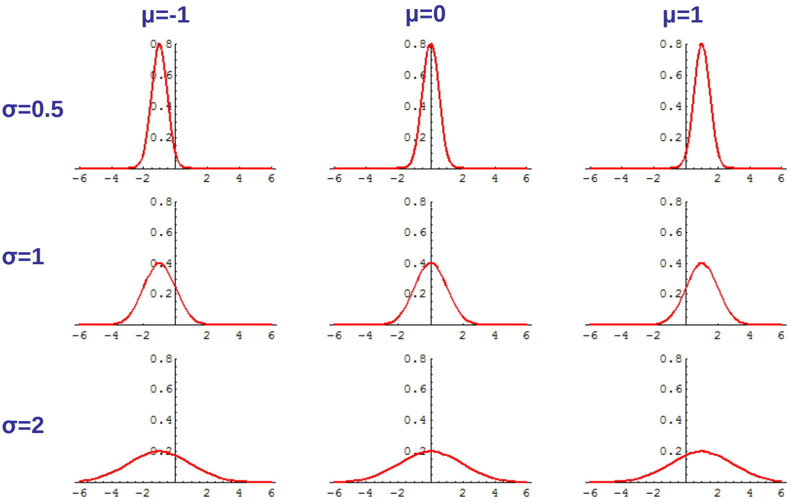
2 **Distribución normal de parámetros** $\mu \in \mathbb{R}, \sigma > 0$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad x \in \mathbb{R}.$$



Notación: $X \sim \mathrm{N}(\mu; \sigma)$.

Ejemplos de densidades normales



La distribución normal es la ley en la cual todo el mundo cree firmemente, los matemáticos porque creen que es un hecho comprobado experimentalmente y los experimentadores, porque creen que se trata de un teorema matemático.
Gabriel Lippman (1845-1921), Premio Nobel de Física (1906).

Importancia de la distribución normal (o de Gauss):

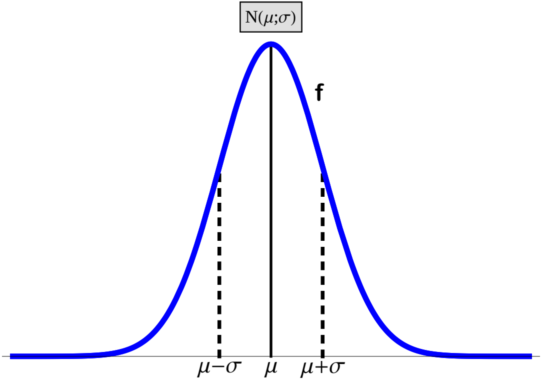
(1) Modeliza muchos fenómenos aleatorios muy usuales de la naturaleza que se pueden considerar como la suma de muchos pequeños efectos independientes como: Peso o altura de una persona o animal (datos biométricos en general); Magnitudes físicas; Ingesta de alimentos; Errores de medición; Datos meteorológicos; etc.

(2) En Estadística aparece como distribución límite de muchos estadísticos que se usan para la inferencia.

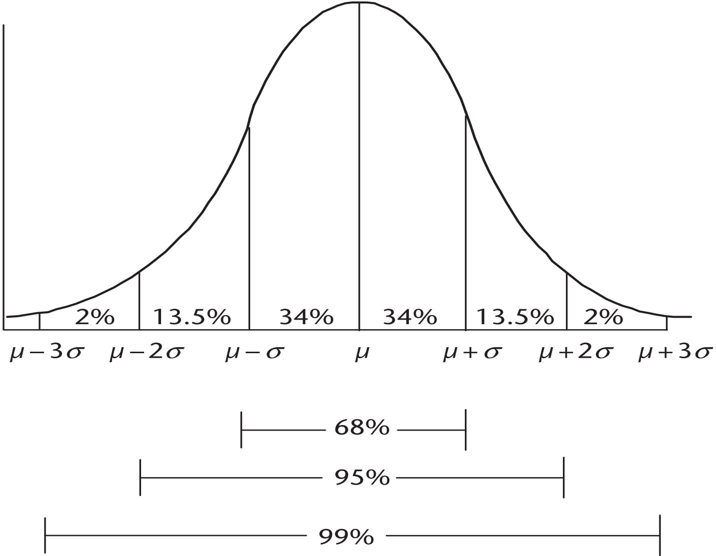
(3) (1) y (2) son debidos (principalmente) a que se verifica el Teorema Central del Límite (TCL).

Algunas propiedades de la densidad normal

- f es simétrica respecto de la recta $x = \mu$ $(f(\mu - x) = f(\mu + x))$.
- Si $\mu = 0$ ($X \sim \mathrm{N}(0;\sigma)$), f es par ($f(-x) = f(x)$).
- f alcanza su máximo absoluto en μ y $f(\mu) = 1/(\sigma\sqrt{2\pi})$.
- Los puntos $\mu \pm \sigma$ son puntos de inflexión de f.



Regla 68-95-99



Regla 68-95-99

Ejemplo: Sea X la v.a. que representa la cantidad diaria de kcal que toma una persona elegida al azar en una población. Se sabe que la población es normal con media $\mu = 2500$ kcal y desviación típica $\sigma = 100$ kcal. Usando las propiedades anteriores da respuestas aproximadas a las preguntas siguientes:

(1) ¿Cuál es la probabilidad de que X esté entre 2300 y 2700 kcal?

(2) ¿Cuál es la probabilidad de que X sea mayor que 2700 kcal?

(3) ¿Cuál es la probabilidad de que X sea mayor que 2500 kcal?

(4) ¿Cuál es la probabilidad de que X sea mayor que 2300 kcal?

Tipificación de variables normales

Llamamos **variable normal tipificada** a la v.a. $Z \sim \mathrm{N}(0;1)$. Es decir, Z tiene densidad

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

Su función de distribución se suele denotar mediante Φ,

$$\Phi(z) := \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt, \quad z \in \mathbb{R}.$$

$$\text{Si } X \sim \mathrm{N}(\mu;\sigma) \Longrightarrow Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0;1).$$

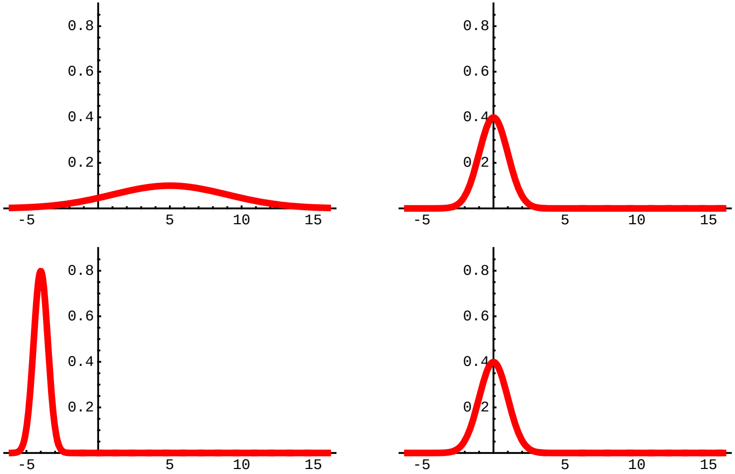
Esta transformación se denomina **tipificación** o **estandarización**.

Ejemplos:

- Si $X \sim \mathrm{N}(5;4)$, entonces $Z = \frac{X-5}{4} \sim \mathrm{N}(0;1)$.
- Si $X \sim \mathrm{N}(-4;0{,}5)$, entonces $Z = \frac{X+4}{0{,}5} \sim \mathrm{N}(0;1)$.

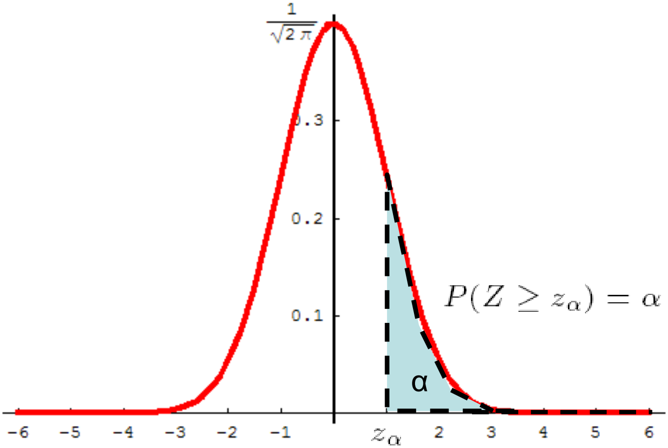
Tipificación de variables normales

- Si $X \sim N(5; 4)$, entonces $Z = \frac{X-5}{4} \sim N(0; 1)$.
- Si $X \sim N(-4; 0{,}5)$, entonces $Z = \frac{X+4}{0{,}5} \sim N(0; 1)$.



Tabulación de variables normales

En las tablas están los valores α y z_α de forma que $\mathrm{P}(Z \geq z_\alpha) = \alpha$.



Tabulación de variables normales

Ejercicio: Sea $Z \sim \mathrm{N}(0;1)$. Calcular:

1. $\mathrm{P}(Z \geq 1,12)$, $\mathrm{P}(Z \leq -0,5)$, $\mathrm{P}(Z \geq -1)$, $\mathrm{P}(Z \leq 1,45)$.
2. z_0 tal que $\mathrm{P}(Z < z_0) = 0,25$.
3. z_0 tal que $\mathrm{P}(Z \leq z_0) = 0,95$.
4. z_0 tal que $\mathrm{P}(Z \geq z_0) = 0,99$.
5. $X \sim \mathrm{N}(30;3)$, $\mathrm{P}(X \leq 38)$.
6. $X \sim \mathrm{N}(30;4)$, $\mathrm{P}(X \geq 25)$.
7. $X \sim \mathrm{N}(20;6)$, $\mathrm{P}(10 \leq X \leq 25)$.
8. $X \sim \mathrm{N}(30;4)$, a tal que $\mathrm{P}(X \leq a) = 0,47$.
9. $X \sim \mathrm{N}(30;4)$, a tal que $\mathrm{P}(X \geq a) = 0,3$.

Ejercicio: Se supone que el gasto anual por hogar en servicios médicos y sanitarios es una variable aleatoria normal con $\mu = 392{,}75$ euros y $\sigma = 33{,}10$ euros. Se elige un hogar al azar:

(a) ¿Cuál es la probabilidad de que el gasto sea mayor que 451,80 euros?

(b) ¿Cuál es la probabilidad de que el gasto no se desvíe de μ en más de 33,10 euros?

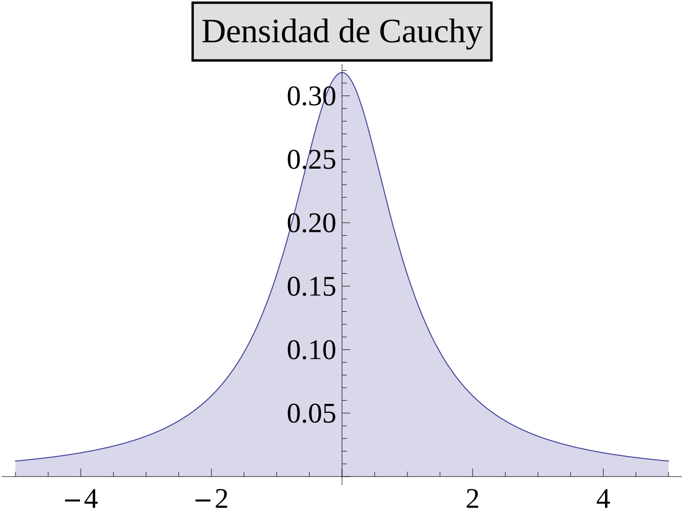
Ejercicio: La cantidad de agua destilada que entrega cierta máquina tiene una distribución normal con parámetros $\mu = 64$ onzas y $\sigma = 0{,}78$ onzas. ¿Qué tamaño de recipiente *c* asegura que el sobreflujo ocurre sólo 5 % del tiempo?

3. **Distribución de Cauchy**

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

Aplicaciones: En el estudio de emisiones de partículas. Si Z es un ángulo aleatorio distribuido uniformemente entre $-\pi/2$ y $\pi/2$, $\mathrm{tg}(Z)$ tiene distribución de Cauchy. El cociente de dos v.a. normales estándar independientes tiene también distribución de Cauchy.

Esta distribución es muy útil para muchos contraejemplos ya que *no* es integrable.


Densidad de Cauchy
0.30
0.25
0.20
0.15
0.10
0.05
−4
−2
2
4

4. **Distribución exponencial de parámetro** $\lambda > 0$

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

Aplicaciones: La distribución exponencial se utiliza para modelar el tiempo transcurrido hasta la ocurrencia por primera vez de un suceso. Muy utilizada en modelos de fiabilidad.

- Tiempo entre dos llamadas consecutivas en una centralita.
- Tiempo entre la llegada de dos pacientes a un servicio de urgencias.
- Tiempos de vida de artículos o piezas (tiempo que tarda en averiarse un electrodoméstico).
- Tiempo que tarda una cierta cantidad de una sustancia radiactiva en reducir su masa a la mitad (datación de fósiles o cualquier materia orgánica mediante la técnica del C^{14}).
- Distancia entre mutaciones en un fragmento de ADN.
- Tiempos de espera en colas.

Notación: $X \sim \text{Exp}(\lambda)$.

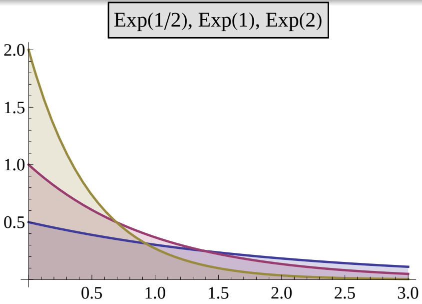
Una variable aleatoria $X \geq 0$ se dice que **no tiene memoria** si

$$\mathrm{P}(X > s + t | X > s) = \mathrm{P}(X > t), \quad \text{para todo } s, t > 0.$$

Propiedad: falta de memoria de la exponencial

Sea $X \geq 0$ una v.a. absolutamente continua. Son equivalentes:

1. X tiene distribución exponencial.
2. X no tiene memoria.



Ejercicio: El tiempo de procesado de un programa de ordenador determinado tiene una distribución exponencial de parámetro $\lambda = 1/2$. Calcúlese la probabilidad de esperar más de 3 segundos para procesar el programa.

5. **Distribución gamma de parámetros** $\alpha, \beta > 0$

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

$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x}\, dx$, $p > 0$, es la función **gamma de Euler**.

Propiedades: $\Gamma(p+1) = p\Gamma(p)$, $p > 0 \Rightarrow \Gamma(n) = (n-1)!$, $n \in \mathbb{N}$. También se puede ver que $\Gamma(1/2) = \sqrt{\pi}$.

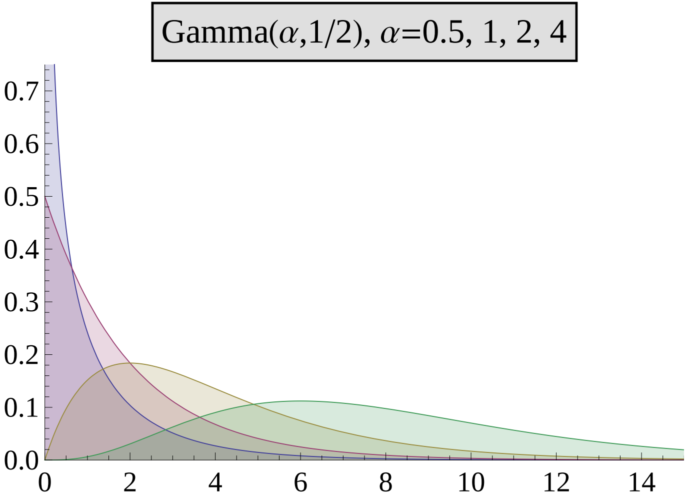
Aplicaciones: Cuando $\alpha \in \mathbb{N}$, se llama distribución de Erlang y se usa en problemas de fiabilidad (tiempo de espera hasta α fallos), cantidad de lluvia caída, cuantía de las reclamaciones a las compañías de seguros, modelos de supervivencia,....

Cuando $\alpha = n/2$ ($n \in \mathbb{N}$) y $\beta = 1/2$, se llama χ^2 con n grados de libertad y desempeña un importante papel en Estadística.

Si $\alpha = 1$ y $\beta = \lambda$, es una distribución $\text{Exp}(\lambda)$.

Notación: $X \sim \text{Gamma}(\alpha; \beta)$.

Función de densidad de Gamma$(\alpha;\beta)$



6 **Distribución beta de parámetros** $\alpha, \beta > 0$

$$f(x) = \frac{1}{\beta(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in (0,1),$$

donde

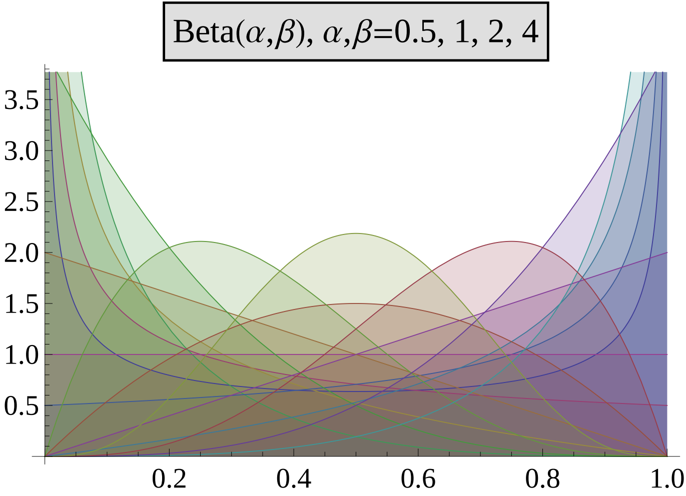
$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad a, b > 0,$$

es la función **beta de Euler**.

Aplicaciones: Dependiendo de los valores de los parámetros la densidad beta adopta formas muy variadas. Esta distribución (o sus versiones "reescaladas" en otros intervalos diferentes a $[0,1]$) proporciona un modelo muy flexible para describir variables aleatorias reales de soporte compacto.

Notación: $X \sim \text{Beta}(\alpha; \beta)$.

Función de densidad de Beta(α, β)



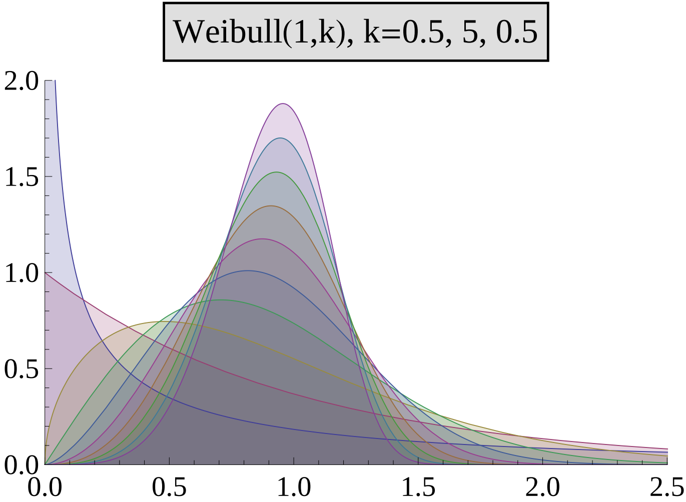
7. **Distribución Weibull de parámetros** $\theta, k > 0$

$$f(x) = \frac{k}{\theta}\left(\frac{x}{\theta}\right)^{k-1} e^{-(x/\theta)^k}, \quad x > 0.$$

Aplicaciones: Esta distribución suele modelizar tiempos de supervivencia en problemas de fiabilidad en ingeniería, velocidad del viento, periodos de incubación de algunas enfermedades, etc.

Notación: $X \sim \text{Weibull}(\theta; k)$.
Observación: $\text{Weibull}(1/\lambda; 1) \sim \text{Exp}(\lambda)$.

Función de densidad de Weibull$(\theta; k)$



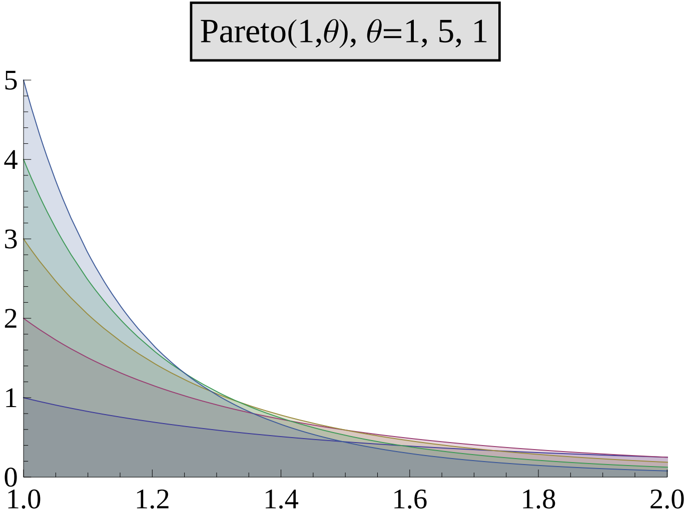
8. **Distribución de Pareto de parámetros $a > 0$ y $\theta > 0$**

$$f(x) = \frac{\theta a^{\theta}}{x^{\theta+1}}, \quad x > a.$$

Aplicaciones: Distribución de ingresos, de reservas de petróleo, de área quemadas en bosques, de tamaños de ficheros enviados por e-mail, de tamaños de partículas,...

Notación: $X \sim \text{Pareto}(a; \theta)$.

Función de densidad de Pareto$(a;\theta)$



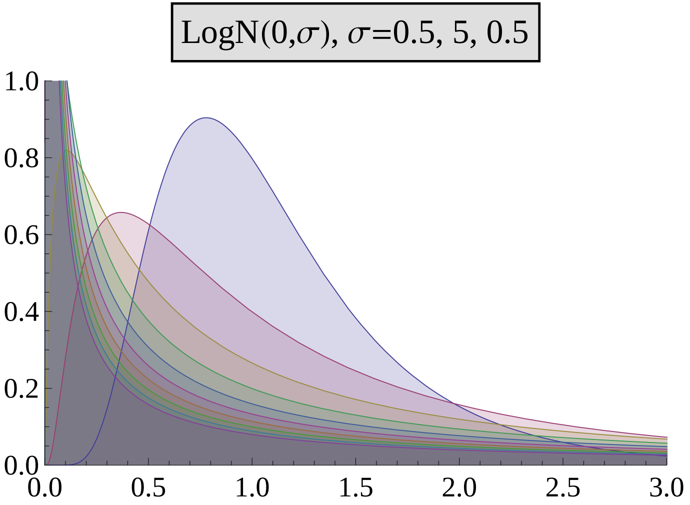
9. **Distribución lognormal de parámetros** $\mu \in \mathbb{R}$ **y** $\sigma > 0$

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right], \quad x > 0.$$

Aplicaciones: Si $X \sim \mathrm{LogN}(\mu;\sigma)$, entonces $\log X \sim \mathrm{N}(\mu;\sigma)$. Se usa en geología (tamaño de rocas sedimentarias) y en general en aquellos casos en los que una variable puede considerarse producto de muchos factores de pequeño efecto individual.

Notación: $X \sim \mathrm{LogN}(\mu;\sigma)$.

Función de densidad de LogN$(\mu;\sigma)$



Probabilidad I

Grado en Matemáticas

Tema 3
Vectores aleatorios

Javier Cárcamo

Departamento de Matemáticas
Universidad Autónoma de Madrid
javier.carcamo@uam.es

Descripción del tema

1. Vectores aleatorias.
2. Función de distribución conjunta.
3. Vectores discretos.
4. Vectores continuos.
5. Ejemplos de distribuciones discretas multidimensionales.
6. Ejemplos de distribuciones continuas multidimensionales.
7. Distribuciones condicionadas.
8. Variables aleatorias independientes.
9. Transformaciones de vectores aleatorios.

Objetivos principales

- Comprender la noción de vector aleatorio.
- Saber calcular distribuciones marginales y condicionadas.
- Calcular la distribución de una transformación de un vector en los casos más sencillos.

En el tema anterior, dado un experimento aleatorio ϵ, prestamos atención a una sóla característica de ϵ considerando una variable $X : \Omega \longrightarrow \mathbb{R}$.

Sin embargo, nos puede interesar estudiar simultáneamente varias características del experimento. Es decir, estudiar varias variables a la vez: $X : \Omega \longrightarrow \mathbb{R}$, $Y : \Omega \longrightarrow \mathbb{R}$, $Z : \Omega \longrightarrow \mathbb{R}$, ...

Estudiar las variable simultáneamente proporciona más información que su estudio por separado ya que permite analizar las interrelaciones entre las características involucradas.

Por ejemplo, al seleccionar una persona se puede observar su peso y altura. Por tanto, podemos analizar dos variables por separado o estudiar el comportamiento de las dos variables a la vez. Al estudiar estas variables de manera conjunta podemos observar la relación existente entre ambas. (En general, es razonable esperar que a mayor altura de una persona, su peso será mayor. Esto no lo podríamos apreciar estudiando las variables "peso" y "altura"de forma independiente.)

Otra situación en la que es importante el estudio de muchas características simultáneamente es la realización de una encuesta (por ejemplo electoral).

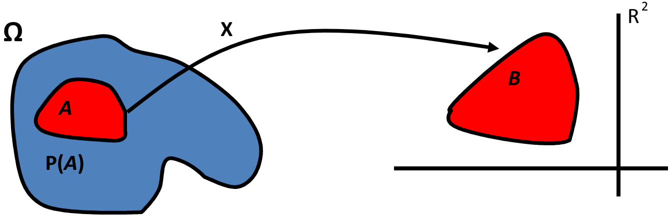
Dado $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad, se llama **vector aleatorio** (d-dimensional) a toda aplicación:

$$\begin{aligned} \mathbf{X} = (X_1, \ldots, X_d) : \Omega &\longrightarrow \mathbb{R}^d \quad (\text{o } \overline{\mathbb{R}}^d) \\ \omega &\longmapsto \mathbf{X}(\omega) = (X_1(\omega), \ldots, X_d(\omega)). \end{aligned}$$

Observación: Realmente, para que $\mathbf{X}$ sea un v.a. debe cumplir una condición de *medibilidad*: Para todo $a_1, \ldots, a_d \in \mathbb{R}^d$, se pide que el conjunto $\{\omega \in \Omega : X_1(\omega) \leq a_1, \ldots, X_d(\omega) \leq a_d\} \in \mathcal{F}$. Cuando $\mathcal{F} = \mathcal{P}(\Omega)$, esto siempre se cumple.

Idea: Como P es medida de probabilidad sobre $(\Omega, \mathcal{F})$, sabemos "calcular probabilidades" en Ω vía P. Al tener una aplicación $\mathbf{X} : \Omega \longrightarrow \mathbb{R}^d$ vamos tener una forma de "medir" en $\mathbb{R}^d$.

La medida de probabilidad P sobre $(\Omega, \mathcal{F})$ induce una *nueva* medida de probabilidad sobre $\mathbb{R}^d$, $\mathrm{P}_{\mathbf{X}}$, que llamaremos **distribución de probabilidad** del vector aleatorio $\mathbf{X}$.



Notación: $\{\mathbf{X} \in B\} = \mathbf{X}^{-1}(B) = \{\omega \in \Omega : \mathbf{X}(\omega) \in B\}, \quad B \subset \mathbb{R}^d.$

$\mathbf{X} = (X_1, \ldots, X_d) : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^d$ v.a. Consideramos

$$\begin{aligned} \mathrm{P}_{\mathbf{X}} : \mathcal{B}(\mathbb{R}^d) \subset \mathcal{P}(\mathbb{R}^d) &\longrightarrow [0,1] \\ B &\longmapsto \mathrm{P}_{\mathbf{X}}(B) = \mathrm{P}(\mathbf{X} \in B). \end{aligned}$$

$\mathrm{P}_{\mathbf{X}}$ se llama **distribución de probabilidad** de $\mathbf{X}$ o **distribución conjunta de las variables** $X_1, \ldots, X_d$ y es una medida de probabilidad en $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Nota: $\mathcal{B}(\mathbb{R}^d)$ es una σ-álgebra sobre $\mathbb{R}^d$ (σ-álgebra de Borel) que incluye todos los conjuntos que nos pueden interesar para calcular probabilidades.

Si $\mathbf{X} = (X_1, \ldots, X_d) : \Omega \longrightarrow \mathbb{R}^d$ es un vector aleatorio, las proyecciones,

$$\begin{aligned} X_i : \Omega &\longrightarrow \mathbb{R} \quad (\text{o } \overline{\mathbb{R}}) \\ \omega &\longmapsto X_i(\omega), \end{aligned}$$

son variables aleatorias. Las variables $X_1, \ldots, X_d$ se llaman **variables marginales** del vector aleatorio $\mathbf{X}$. Las d distribuciones de probabilidad (sobre $\mathbb{R}$) asociadas, $\mathrm{P}_{X_1}, \ldots, \mathrm{P}_{X_d}$, se denominan **distribuciones marginales** de $\mathbf{X}$.

Nota importante: En general, si conocemos $\mathrm{P}_{\mathbf{X}}$ podemos calcular las distribuciones marginales $\mathrm{P}_{X_1}, \ldots, \mathrm{P}_{X_d}$. Sin embargo, el recíproco *no* es cierto. Es decir, (salvo en algunos casos especiales) aunque sepamos la distribución de $X_1, \ldots, X_d$ no podemos calcular la distribución conjunta del vector $\mathbf{X} = (X_1, \ldots, X_d)$.

Ejemplo: ϵ: lanzar una moneda al aire 3 veces.

$$\Omega = \{(C, C, C), (C, C, +), \ldots, (+, +, +)\}.$$

$(\Omega, \mathcal{F}, \mathrm{P})$: modelo de Laplace (equiprobabilidad). Card$(\Omega) = 8$.

$$\mathrm{P}(\{(C, C, C)\}) = \mathrm{P}(\{(C, C, +)\}) = \cdots = \mathrm{P}(\{(+, +, +)\}) = 1/8.$$

Consideramos $\mathbf{X} = (X, Y) : \Omega \longrightarrow \mathbb{R}^2$, de forma que:

- $X \equiv$ número de caras obtenidas.
- $Y \equiv$ máximo número de caras seguidas obtenidas.

Problema: Calcular la distribución de probabilidad conjunta de (X, Y) y las distribuciones marginales de las variables X e Y.

Primero tenemos que ver qué valores puede tomar el vector $\mathbf{X}$.

Ejemplo: ϵ: lanzar una moneda al aire 3 veces.

$$\Omega = \{(C,C,C),(C,C,+),\ldots,(+,+,+)\}.$$

- $X \equiv$ número de caras obtenidas al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas obtenidas.

Valores que toma el vector (X, Y)

$$\begin{array}{rcl} (X,Y):\Omega & \longrightarrow & \mathbb{R}\times\mathbb{R} \\ (C,C,C) & \longmapsto & (3,3) \\ (C,C,+) & \longmapsto & (2,2) \\ (+,C,C) & \longmapsto & (2,2) \\ (C,+,C) & \longmapsto & (2,1) \end{array} \qquad \begin{array}{rcl} (X,Y):\Omega & \longrightarrow & \mathbb{R}\times\mathbb{R} \\ (+,+,+) & \longmapsto & (0,0) \\ (C,+,+) & \longmapsto & (1,1) \\ (+,C,+) & \longmapsto & (1,1) \\ (+,+,C) & \longmapsto & (1,1) \end{array}$$

(X, Y) toma los valores $\{(0,0),(1,1),(2,1),(2,2),(3,3)\}$. Para calcular la distribución de probabilidad de (X, Y) tenemos que calcular la probabilidad de que (X, Y) tome estos 5 valores.

- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

$$\begin{aligned}
\mathrm{P}_{(X,Y)}(\{(0,0)\}) &= \mathrm{P}(X=0, Y=0) = \mathrm{P}(\{\omega \in \Omega : X(\omega)=0, Y(\omega)=0\}) \\
&= \mathrm{P}(\{(+,+,+)\}) = 1/8.
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}_{(X,Y)}(\{(1,1)\}) &= \mathrm{P}(X=1, Y=1) = \mathrm{P}(\{\omega \in \Omega : X(\omega)=1, Y(\omega)=1\}) \\
&= \mathrm{P}(\{(C,+,+),(+,C,+),(+,+,C)\}) = 3/8.
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}_{(X,Y)}(\{(2,1)\}) &= \mathrm{P}(X=2, Y=1) = \mathrm{P}(\{\omega \in \Omega : X(\omega)=2, Y(\omega)=1\}) \\
&= \mathrm{P}(\{(C,+,C)\}) = 1/8.
\end{aligned}$$

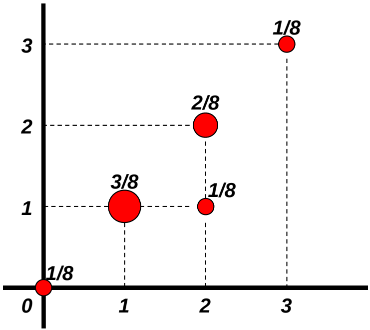
$$\begin{aligned}
\mathrm{P}_{(X,Y)}(\{(2,2)\}) &= \mathrm{P}(X=2, Y=2) = \mathrm{P}(\{\omega \in \Omega : X(\omega)=2, Y(\omega)=2\}) \\
&= \mathrm{P}(\{(C,C,+),(+,C,C)\}) = 2/8.
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}_{(X,Y)}(\{(3,3)\}) &= \mathrm{P}(X=3, Y=3) = \mathrm{P}(\{\omega \in \Omega : X(\omega)=3, Y(\omega)=3\}) \\
&= \mathrm{P}(\{(C,C,C)\}) = 1/8.
\end{aligned}$$

El vector (X,Y) toma los valores $(0,0),(1,1),(2,1),(2,2),(3,3)$ con probabilidades respectivas $1/8$, $3/8$, $1/8$, $2/8$ y $1/8$.

- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

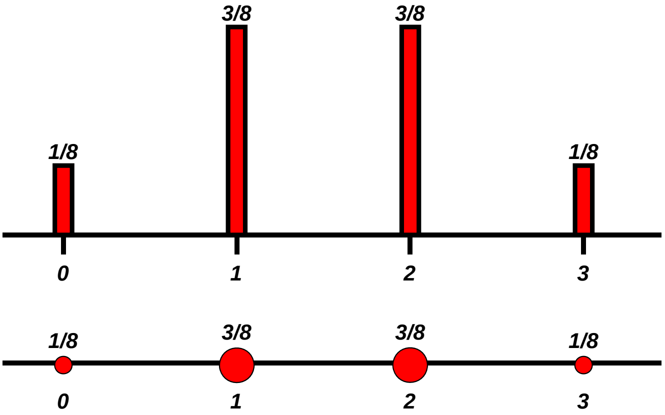
El vector (X, Y) toma los valores $(0,0), (1,1), (2,1), (2,2), (3,3)$ con probabilidades respectivas $1/8$, $3/8$, $1/8$, $2/8$ y $1/8$.



X \ Y	0	1	2	3
0	1/8	0	0	0
1	0	3/8	0	0
2	0	1/8	2/8	0
3	0	0	0	1/8

- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

Distribución marginal de X



- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

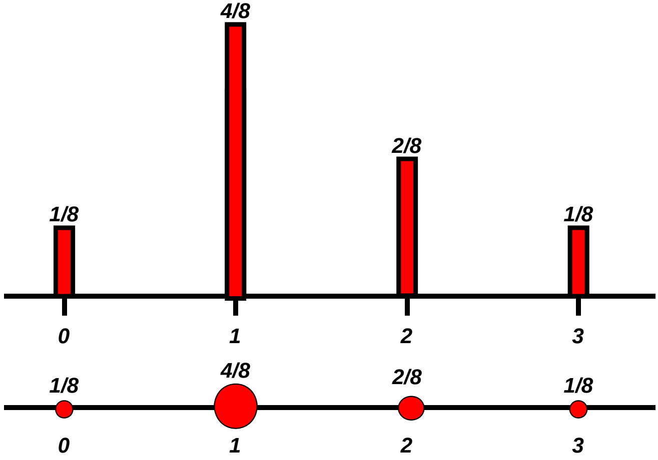
Distribución marginal de Y

$$
\begin{array}{rcl}
Y : \Omega & \longrightarrow & \mathbb{R} \\
(C, C, C) & \longmapsto & 3 \\
(C, C, +) & \longmapsto & 2 \\
(+, C, C) & \longmapsto & 2 \\
(C, +, C) & \longmapsto & 1
\end{array}
\qquad
\begin{array}{rcl}
Y : \Omega & \longrightarrow & \mathbb{R} \\
(+, +, +) & \longmapsto & 0 \\
(C, +, +) & \longmapsto & 1 \\
(+, C, +) & \longmapsto & 1 \\
(+, +, C) & \longmapsto & 1
\end{array}
$$

La variable aleatoria Y toma los valores 0,1,2 y 3 con probabilidades respectivas $1/8$, $4/8$, $2/8$ y $1/8$.

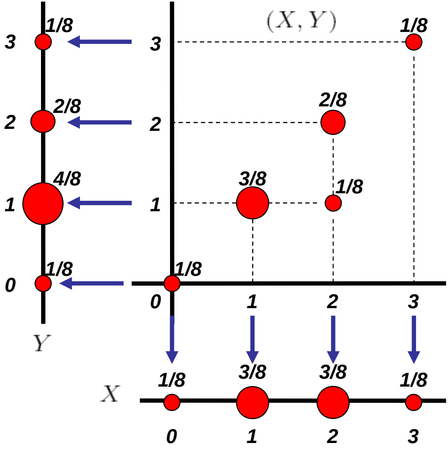
- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

Distribución marginal de Y



- $X \equiv$ número de caras al lanzar 3 monedas.
- $Y \equiv$ máximo número de caras seguidas al lanzar 3 monedas.

Distribución conjunta y marginales de (X, Y)

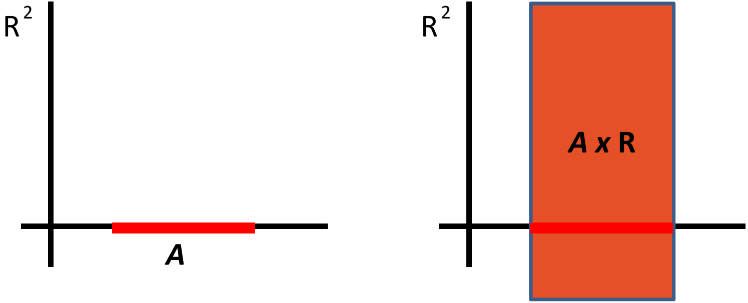


Y \ X	0	1	2	3	X
0	1/8	0	0	0	1/8
1	0	3/8	0	0	4/8
2	0	1/8	2/8	0	2/8
3	0	0	0	1/8	1/8
Y	1/8	3/8	3/8	1/8	1

Observación: $A \subset \mathbb{R}$, $\{X_1 \in A\} = \{\mathbf{X} \in A \times \mathbb{R} \times \cdots \times \mathbb{R}\}$.

Luego, $\mathrm{P}(X_1 \in A) = \mathrm{P}(\mathbf{X} \in A \times \mathbb{R} \times \cdots \times \mathbb{R})$.

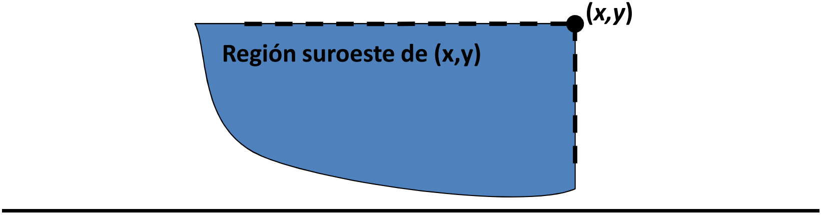
Idea física: $n = 2$, $\mathrm{P}(X \in A) = \mathrm{P}((X, Y) \in A \times \mathbb{R})$



Relación entre la distribución conjunta y las marginales: La masa que la variable aleatoria X concentra en A es la masa que el vector aleatorio (X, Y) concentra en $A \times \mathbb{R}$.

Dado $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, la **región suroeste** generada por $\mathbf{x}$ es

$$S_{\mathbf{x}} = (-\infty, x_1] \times \cdots \times (-\infty, x_d].$$



Sea $\mathbf{X} = (X_1, \ldots, X_d) : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^d$ vector aleatorio. Se llama **función de distribución** de $\mathbf{X}$ o **función de distribución conjunta** de las X_i-s a la función $F_{\mathbf{X}} : \mathbb{R}^d \longrightarrow \mathbb{R}$ tal que

$$F_{\mathbf{X}}(\mathbf{x}) = \mathrm{P}(\mathbf{X} \in S_{\mathbf{x}}) = \mathrm{P}(X_1 \leq x_1, \ldots, X_d \leq x_d).$$

Sea $\mathbf{X} = (X_1, \ldots, X_k) : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^d$ vector aleatorio. Se llama **función de distribución** de $\mathbf{X}$ o **función de distribución conjunta** de las X_i-s a la función $F_{\mathbf{X}} : \mathbb{R}^d \longrightarrow \mathbb{R}$ tal que

$$F_{\mathbf{X}}(\mathbf{x}) = \mathrm{P}(\mathbf{X} \in S_{\mathbf{x}}) = \mathrm{P}(X_1 \leq x_1, \ldots, X_d \leq x_d).$$

Teorema: Propiedades básicas de la función de distribución

Sea $\mathbf{X}$ vector aleatorio y F su función de distribución.

1. F es no decreciente (en $\mathbb{R}^d$).
2. F es continua por la derecha en cada variable.
3. $\lim_{x_i \to -\infty} F(x_1, \ldots, x_d) = 0$ $(i = 1, \ldots, d)$;
   $\lim_{x_1, \ldots, x_d \to +\infty} F(x_1, \ldots, x_d) = 1.$

Pregunta: ¿Qué significa que una función sea monótona (no decreciente) en $\mathbb{R}^d$?

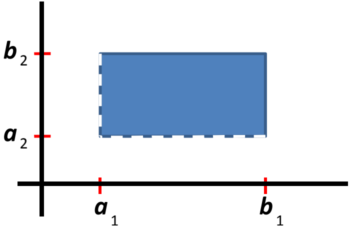
Funciones monótonas (no decrecientes) en $\mathbb{R}^d$

Notación: Si $G : \mathbb{R}^k \longrightarrow \mathbb{R}$ y $u \leq v$,

$\Delta_i(u, v) = G(x_1, \ldots, x_{i-1}, v, x_{i+1}, \ldots, x_k) - G(x_1, \ldots, x_{i-1}, u, x_{i+1}, \ldots, x_k)$.

En $d = 1$, para $a \leq b$, $\Delta(a, b)F(x) = F(b) - F(a) \geq 0$.

En $d = 2$, para $a_1 \leq b_1$ y $a_2 \leq b_2$, se pide que
$\Delta_1(a_1, b_1)\Delta_2(a_2, b_2)F(x, y) = \Delta_1(a_1, b_1)(F(x, b_2) - F(x, a_2)) =$
$F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \geq 0$.



En general, se pide que, para $a_i \leq b_i$ $(i = 1, \ldots, d)$

$$\Delta_1(a_1, b_1)\Delta_2(a_2, b_2) \cdots \Delta_d(a_d, b_d)F(x_1, \cdots, x_k) \geq 0.$$

Sea $\mathbf{X} = (X_1, \ldots, X_k) : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^d$ vector aleatorio.

$$F_{\mathbf{X}}(\mathbf{x}) = \mathrm{P}(\mathbf{X} \in S_{\mathbf{x}}) = \mathrm{P}(X_1 \leq x_1, \ldots, X_d \leq x_d).$$

Teorema: Propiedades básicas de la función de distribución

Sea $\mathbf{X}$ vector aleatorio y F su función de distribución.

1. F es no decreciente (en $\mathbb{R}^d$).
2. F es continua por la derecha en cada variable.
3. $\lim_{x_i \to -\infty} F(x_1, \ldots, x_d) = 0$ $(i = 1, \ldots, d)$;
   $\lim_{x_1, \ldots, x_d \to +\infty} F(x_1, \ldots, x_d) = 1$.

Teorema de unicidad

Sea $F : \mathbb{R}^d \longrightarrow \mathbb{R}$ verificando 1, 2 y 3 de arriba, entonces, existe $\mathbf{X} : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^d$ vector aleatorio tal que $F = F_{\mathbf{X}}$.

Además, el vector $\mathbf{X}$ es único en distribución (es decir, la f.d. caracteriza la distribución de probabilidad de un v.a. de forma biunívoca).

Sea $\mathbf{X} = (X_1, \ldots, X_k) : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^d$ vector aleatorio. Se llama **función de distribución** de **X** o **función de distribución conjunta** de las X_i-s a la función $F_{\mathbf{X}} : \mathbb{R}^d \longrightarrow \mathbb{R}$ tal que

$$F_{\mathbf{X}}(\mathbf{x}) = \mathrm{P}(\mathbf{X} \in S_{\mathbf{x}}) = \mathrm{P}(X_1 \leq x_1, \ldots, X_d \leq x_d).$$

Cada una de las variables marginales, $X_1, \ldots, X_d$, tiene una función de distribución (unidimensional):

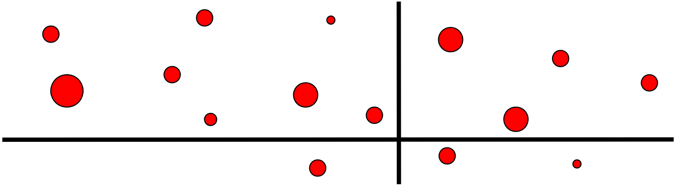
$$F_{X_i}(x_i) = \mathrm{P}(X_i \leq x_i), \qquad i = 1, \ldots, d.$$

Las funciones $F_1, \ldots, F_d$ se llaman las **funciones de distribución marginales** de las variables $X_1, \ldots, X_d$.

Pregunta: ¿Cuál es la relación entre $F_{\mathbf{X}}$ y $F_{X_1}, \ldots, F_{X_d}$?

$$F_{X_i}(x_i) = \lim_{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_d \to \infty} F_{\mathbf{X}}(x_1, \ldots, x_d).$$

Un vector aleatorio $\mathbf{X} : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^d$ se dice que tiene **distribución discreta** si existe un conjunto $S \subset \mathbb{R}^d$ contable (finito o numerable) tal que $\mathrm{P}(\mathbf{X} \in S) = 1$ $(\mathrm{P}(\mathbf{X} \in S^c) = 0)$. Es decir, $\mathbf{X}$ concentra su masa en S. El conjunto S se llama **soporte** de la distribución (de $\mathbf{X}$).



Nota: $\mathbf{X}$ vector discreto con soporte $S \subset \mathbb{R}^d$, entonces

$$\mathrm{P}(\mathbf{X} \in B) = \sum_{s \in S \cap B} \mathrm{P}(\mathbf{X} = s), \quad B \subset \mathbb{R}^d.$$

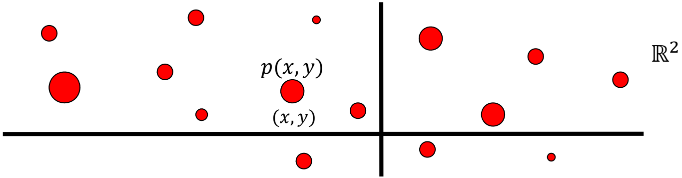
En otras palabras, sólo nos interesa conocer cómo es la distribución de $\mathbf{X}$ en un conjunto finito o infinito numerable S.

Sea $\mathbf{X} : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^d$ un vector con distribución discreta y soporte S. La función:

$$\begin{aligned} p_{\mathbf{X}} : \mathbb{R}^d &\longrightarrow \mathbb{R} \\ \mathbf{x} &\longmapsto p_{\mathbf{X}}(\mathbf{x}) = \mathrm{P}(\mathbf{X} = \mathbf{x}) \end{aligned}$$

se denomina **función (de masa de) probabilidad conjunta** de $\mathbf{X}$.

Notación: Cuando no haya confusión denotaremos por p (en lugar de $p_{\mathbf{X}}$) la función de probabilidad del vector $\mathbf{X}$.



Si $\mathbf{X}$ es un vector discreto con función de probabilidad p, tenemos:

- $p(\mathbf{x}) = 0$, si $\mathbf{x} \in S^c$ y $\sum_{\mathbf{x} \in S} p(\mathbf{x}) = 1$.
- $\mathrm{P}(\mathbf{X} \in B) = \sum_{\mathbf{x} \in B \cap S} p(\mathbf{x})$ (suma finita o serie abs. conv.).

Un vector aleatorio $\mathbf{X} : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^d$ se dice **vector aleatorio discreto** si existe un conjunto $S \subset \mathbb{R}^d$ contable tal que $\mathrm{P}(\mathbf{X} \in S) = 1$ $(\mathrm{P}(\mathbf{X} \in S^c) = 0)$. El conjunto S se llama **soporte** de la distribución (de $\mathbf{X}$).

Teorema: Vectores discretos

Sea $\mathbf{X} = (X_1, \dots, X_d) : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^d$ vector aleatorio.

$$\mathbf{X} \text{ vector discreto} \iff X_1, \dots, X_d \text{ variables discretas.}$$

Si $\mathbf{X} = (X_1, \dots, X_d)$ vector discreto, las funciones de probabilidad de $X_1, \dots, X_d$, $p_{X_1}, \dots, p_{X_d}$, se denominan **funciones (de masa) de probabilidad marginales**.

Pregunta: ¿Qué relación hay entre $p_{\mathbf{X}}$ y $p_{X_1}, \dots, p_{X_d}$?

$$p_{X_i}(x_i) = \sum_{j_1, \dots, j_{i-1}, j_{i+1}, \dots, j_d} p_{\mathbf{X}}((x_{j_1}, \dots, x_{j_{i-1}}, x_i, x_{j_{i+1}}, \dots, x_{j_d})).$$

Una función $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ se llama **función de densidad (de probabilidad) conjunta** sobre $\mathbb{R}^d$ si cumple:

(a) $f(\mathbf{x}) \geq 0$, para todo $\mathbf{x} \in \mathbb{R}^d$.

(b) f es integrable (Riemman).

(c) $\displaystyle\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(t_1, \ldots, t_d)\, dt_1 \cdots dt_d = 1.$

Un vector aleatorio $\mathbf{X} = (X_1, \ldots, X_d)$ con función de distribución F se dice que es **(absolutamente) continuo** si existe una función de densidad sobre $\mathbb{R}^d$, f, tal que para todo $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$

$$F(x_1, \ldots, x_d) = \int_{S_{\mathbf{x}}} f = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(t_1, \ldots, t_d)\, dt_d \cdots dt_1. \quad (*)$$

De $(*)$, y usando el teorema fundamental del cálculo, se tiene que en cada punto de continuidad $(x_1, \ldots, x_d)$ de f

$$\frac{\partial^d F(x_1, \ldots, x_d)}{\partial x_1 \cdots \partial x_d} = f(x_1, \ldots, x_d).$$

Cálculo de probabilidades con vectores continuos

Si $\mathbf{X} = (X_1, \ldots, X_d)$ es un vector continuo con densidad f y $A \subset \mathbb{R}^d$ un conjunto "regular" para la integración, entonces

$$\mathrm{P}(\mathbf{X} \in A) = \underset{A}{\int \cdots \int} f(x_1, \ldots, x_d) dx_1 \cdots dx_d.$$

Teorema: Vectores continuos

Si $\mathbf{X} = (X_1, \ldots, X_d)$ es un vector aleatorio continuo, entonces $X_1, \ldots, X_d$ son variables continuas. Además, si $\mathbf{X}$ tiene densidad f, la densidad de la variable marginal X_i $(i = 1, \ldots, d)$ es

$$f_i(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_d)\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_d.$$

f_i se llama **función de densidad marginal** de X_i $(i = 1, \ldots, d)$.

Atención: El recíproco del resultado anterior *no* es cierto.

Teorema: Vectores continuos

Si $\mathbf{X} = (X_1, \ldots, X_d)$ es un vector aleatorio continuo, entonces $X_1, \ldots, X_d$ son variables continuas. Además, si $\mathbf{X}$ tiene densidad f, la densidad de la variable marginal X_i $(i = 1, \ldots, d)$ es

$$f_i(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_d)\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_d.$$

f_i se llama **función de densidad marginal** de X_i $(i = 1, \ldots, d)$.

Ejercicio: Si $\mathbf{X} = (X, Y)$ vector aleatorio bidimensional con densidad conjunta $f(x, y)$. Las densidades de X e Y son:

$f_1(x) =$

$f_2(y) =$

Observación: Puede ocurrir que $X_1, \ldots, X_d$ variables aleatorias con densidad (es decir, cada una de ellas es absolutamente continua) y que el vector $\mathbf{X} = (X_1, \ldots, X_d)$ *no* tenga densidad.

Ejemplo:

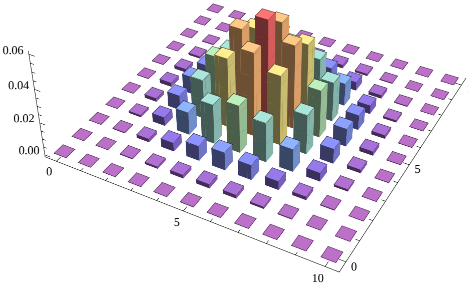
1. **Producto de binomiales** $(n_1, n_2 \in \mathbb{N},\ p_1, p_2 \in (0,1))$

$$\mathrm{P}(X = k, Y = l) = \binom{n_1}{k} p_1^k (1-p_1)^{n_1-k} \binom{n_2}{l} p_2^l (1-p_2)^{n_2-l}.$$

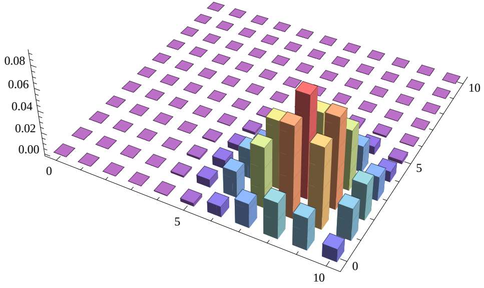
El soporte es $S = \{0, 1, \ldots, n_1\} \times \{0, 1, \ldots, n_2\}$.

Las marginales son $X \sim \mathrm{B}(n_1, p_1)$ e $Y \sim \mathrm{B}(n_2, p_2)$.

$\mathrm{B}(10, 0{,}5) \otimes \mathrm{B}(10, 0{,}5)$



$\mathrm{B}(10, 0{,}2) \otimes \mathrm{B}(10, 0{,}8)$



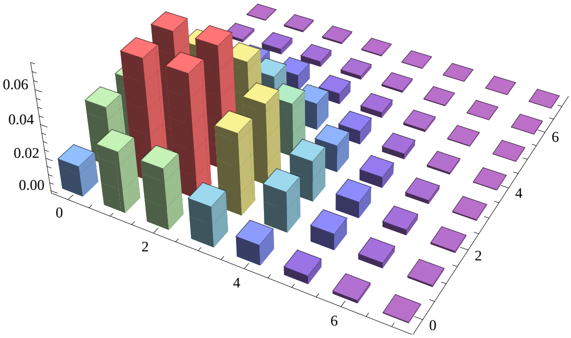
2. **Producto de Poisson** de parámetros $\lambda, \mu > 0$

$$\mathrm{P}(X = k, Y = l) = e^{-\lambda}\frac{\lambda^k}{k!}e^{-\mu}\frac{\mu^l}{l!}, \quad k, l = 0, 1, \ldots$$

El soporte es $S = \{0, 1, \ldots\} \times \{0, 1, \ldots\}$.

Las marginales son $X \sim \mathrm{P}(\lambda)$ e $Y \sim \mathrm{P}(\mu)$.

$\mathrm{P}(2) \otimes \mathrm{P}(2)$



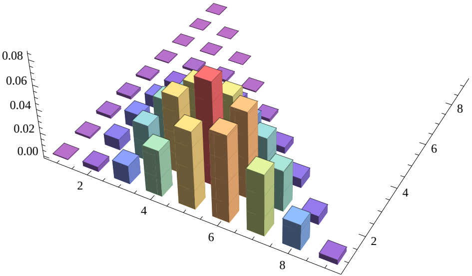
3. **Trinomial** ($n \in \mathbb{N}$, $p_1, p_2 \in (0,1)$ con $p_1 + p_2 \leq 1$)

$$\begin{aligned} \mathrm{P}(X = k, Y = l) &= \binom{n}{k}\binom{n-k}{l} p_1^k p_2^l (1 - p_1 - p_2)^{n-k-l} \\ &= \frac{n!}{k!l!(n-k-l)!} p_1^k p_2^l (1 - p_1 - p_2)^{n-k-l}. \end{aligned}$$

El soporte $S = \{(k, l) : k, l = 0, 1, \ldots, n \text{ y } k + l \leq n\}$.

Las marginales son $X \sim \mathrm{B}(n, p_1)$ e $Y \sim \mathrm{B}(n, p_2)$.

$\mathrm{B}(10, 0{,}2, 0{,}5)$



Teorema: Producto de densidades univariadas

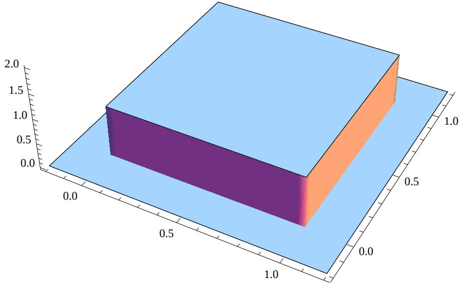
Sean f, g funciones de densidad sobre $\mathbb{R}$. La función

$$h(x,y) = f(x)g(y), \quad (x,y) \in \mathbb{R}^2$$

es una densidad en $\mathbb{R}^2$ cuyas densidades marginales son f y g.

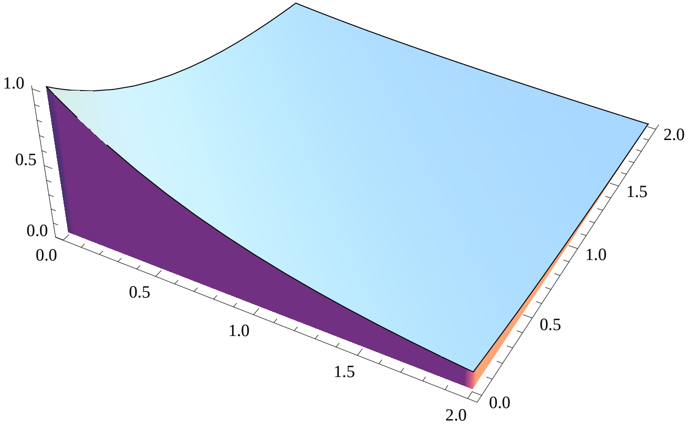
1. **Uniforme en el rectángulo** $(a,b) \times (c,d)$

$$f(x,y) = \frac{1}{(b-a)(c-d)} 1_{(a,b)\times(c,d)}(x,y).$$



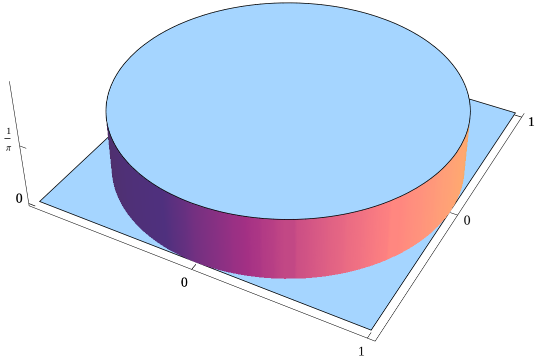
2. **Producto de exponenciales de parámetros $\lambda, \mu > 0$**

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



3 Uniforme en el círculo unidad

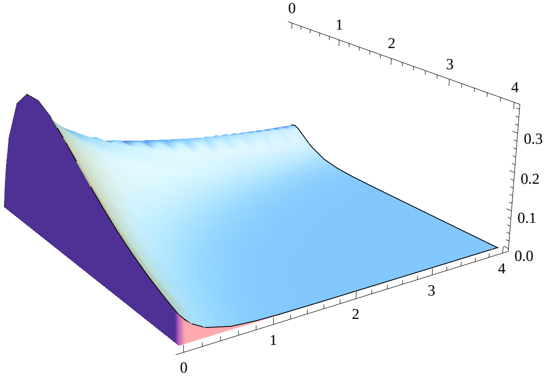
$$f(x,y) = \begin{cases} 1/\pi, & \text{si } x^2 + y^2 < 1, \\ 0, & \text{en otro caso.} \end{cases}$$



Ejercicio: Calcula las densidades marginales, $f_1(x)$ y $f_2(y)$.

4. (X, Y) con densidad:

$$f(x,y) = \begin{cases} xe^{-x(y+1)}, & \text{si } x, y > 0, \\ 0, & \text{en otro caso.} \end{cases}$$



Ejercicio: Calcula las densidades marginales, $f_1(x)$ y $f_2(y)$.

5. **Normal multivariante:** El vector aleatorio $\mathbf{X}$ es **normal** (d-dimensional) con parámetros μ y Σ (notación: $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$) si tiene densidad dada por:

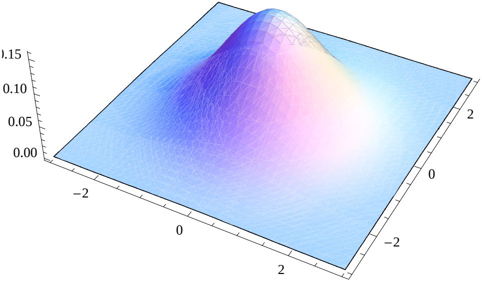
$$f(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1/2}(2\pi)^{-d/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}.$$

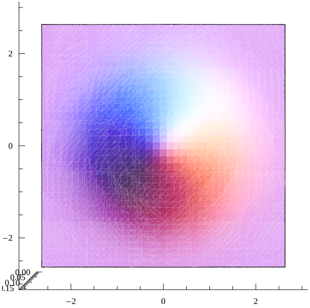
$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{p2} & \cdots & \sigma_{dd} \end{bmatrix}.$$

Nota: Se puede comprobar que las variables marginales también son normales. De hecho, $X_i \sim \mathrm{N}(\mu_i, \sigma_{ii})$.

Ejemplos de densidades normales en dimensión 2

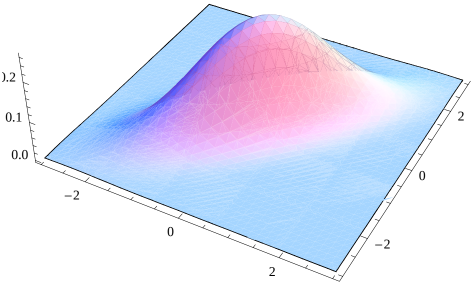
$$\text{Densidad de } \mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ con } \boldsymbol{\mu} = (0,0)' \text{ y } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

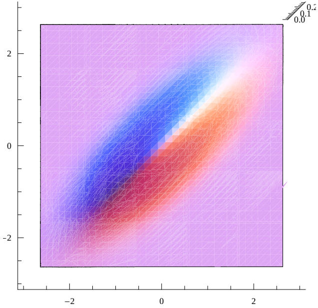




Ejemplos de densidades normales en dimensión 2

$$\text{Densidad de } \mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ con } \boldsymbol{\mu} = (0,0)' \text{ y } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0{,}8 \\ 0{,}8 & 1 \end{pmatrix}.$$





6. Ejemplos de distribuciones continuas multidimensionales

Nota: En dimensión 2, la densidad de la normal (bivariante) se suele expresar también como

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\times \exp\left[\frac{-1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)\right] \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)} Q(x,y)\right], \end{aligned}$$

donde Q es la forma cuadrática

$$Q(x,y) = \frac{1}{1-\rho^2}\left(\frac{x-\mu_1}{\sigma_1}, \frac{y-\mu_2}{\sigma_2}\right)\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}\begin{pmatrix} \frac{x-\mu_1}{\sigma_1} \\ \frac{y-\mu_2}{\sigma_2} \end{pmatrix}.$$

Los parámetros del vector son $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ y $\rho \in (-1,1)$.

CASO DISCRETO: Sea (X, Y) un vector discreto y $x_0 \in \mathbb{R}$ un valor *fijo* con $\mathrm{P}(X = x_0) > 0$. La variable Y **condicionada** a $X = x_0$, $Y|X = x_0$, es la variable con distribución de probabilidad

$$p_{Y|X=x_0}(y) = \mathrm{P}(Y = y|X = x_0) = \frac{\mathrm{P}(X = x_0, Y = y)}{\mathrm{P}(X = x_0)}, \quad y \in \mathbb{R}.$$

De forma análoga se define $X|Y = y_0$, para aquellos y_0 tales que $\mathrm{P}(Y = y_0) > 0$.

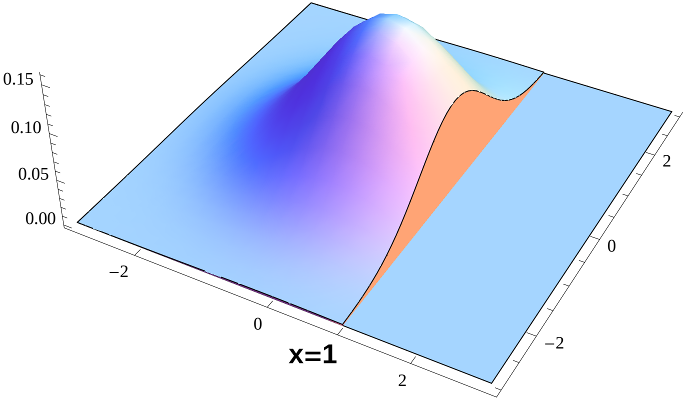
CASO CONTINUO: Sea (X, Y) un vector aleatorio continuo con densidad conjunta f y $x_0 \in \mathbb{R}$ un valor *fijo* con $f_1(x_0) > 0$. La variable Y **condicionada** a $X = x_0$, $Y|X = x_0$, es la variable con densidad de probabilidad dada por

$$f_{Y|X=x_0}(y) = \frac{f(x_0, y)}{f_1(x_0)}, \quad y \in \mathbb{R}.$$

De forma análoga se define la densidad de $X|Y = y_0$, para aquellos y_0 tales que $f_2(y_0) > 0$.

Ejemplo: $\mathbf{X} = (X, Y) \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{\mu} = (0,0)'$ y $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

La densidad de $Y|X = 1$ es también normal $\mathrm{N}(0, 1)$.



Ejercicio: (X, Y) vector aleatorio con densidad conjunta

$$f(x, y) = \begin{cases} xe^{-x(y+1)}, & \text{si } x, y > 0, \\ 0, & \text{en otro caso.} \end{cases}$$

(a) Para $a > 0$, calcula la densidad de la variable condicionada $Y|X = a$.

(b) Para $a > 0$, calcula la densidad de la variable condicionada $X|Y = a$.

(c) Identifica las distribuciones de probabilidad anteriores.

CASO DISCRETO: Dado un vector aleatorio discreto $\mathbf{X} = (X_1, \dots, X_d)$, se dice que las variables $X_1, \dots, X_d$ son **independientes** si para cualesquiera $x_1, \dots, x_d \in \mathbb{R}$,

$$\mathrm{P}(X_1 = x_1, \dots, X_d = x_d) = \mathrm{P}(X_1 = x_1) \cdots \mathrm{P}(X_d = x_d).$$

Es decir, la función de probabilidad conjunta de $\mathbf{X}$ es el producto de las funciones de probabilidad marginales.

Observación: Si $X_1, \dots, X_d$ son independientes, entonces para todo $x_1, \dots, x_d \in \mathbb{R}$, los d sucesos $\{X_1 = x_1\}, \dots, \{X_d = x_d\}$ son (mutuamente) independientes.

CASO CONTINUO: Dado un vector aleatorio continuo $\mathbf{X} = (X_1, \ldots, X_d)$ con densidad conjunta f, se dice que las variables $X_1, \ldots, X_d$ son **independientes** si para cualesquiera $x_1, \ldots, x_d \in \mathbb{R}$,

$$f(x_1, \ldots, x_d) = f_1(x_1) \cdots f_d(x_d).$$

Es decir, la función de densidad conjunta de $\mathbf{X}$ es el producto de las funciones de densidad marginales.

Teorema: Caracterización mediante funciones de distribución

Sea $\mathbf{X} = (X_1, \ldots, X_d)$ un vector (discreto o continuo) con función de distribución conjunta F y marginales $F_1, \ldots, F_d$.

$$X_1, \ldots, X_d \text{ independientes} \Longleftrightarrow F(x_1, \ldots, x_d) = F_1(x_1) \cdots F_d(x_d).$$

(La función de distribución conjunta es el producto de las funciones de distribución marginales.)

Problema general

$$\begin{tikzcd}[column sep=huge]
(\Omega, \mathcal{F}, \mathrm{P}) \arrow[r, "\mathbf{X} = (X_1, \ldots, X_d)"] \arrow[dr, dotted, "\mathbf{Y} = T(\mathbf{X})"'] & \mathbb{R}^d \arrow[d, "T"] \\
 & \mathbb{R}^k
\end{tikzcd}$$

Distribuciones derivadas: Dado un vector $\mathbf{X} = (X_1, \ldots, X_d)$ con distribución de probabilidad conocida, $\mathrm{P}_{\mathbf{X}}$, y una transformación (continua) $T = (T_1, \ldots, T_k) : \mathbb{R}^d \longrightarrow \mathbb{R}^k$, el problema consiste en encontrar la distribución de $\mathbf{Y} = T(X_1, \ldots, X_d)$ (vector k-dimensional), $\mathrm{P}_{\mathbf{Y}}$.

Nota: Teóricamente, para $B \subset \mathbb{R}^k$, $\mathrm{P}_{\mathbf{Y}}(B) = \mathrm{P}_{\mathbf{X}}(T^{-1}(B))$. Sin embargo, lo que se busca es el cálculo explícito (cuando se pueda).

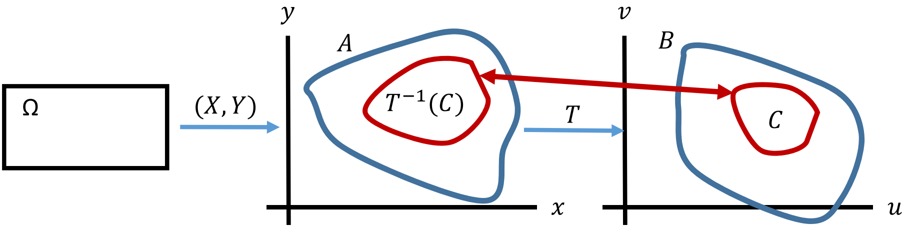
Ejemplo: Sean $X_1, \ldots, X_d$ variables independientes con funciones de distribución $F_1, \ldots, F_d$. Calcula la f.d. de $Y = \text{máx}(X_1, \ldots, X_d)$ y de $Z = \text{mín}(X_1, \ldots, X_d)$.

Teorema: Cambio de variables

Sea $(X,Y):\Omega\longrightarrow A\subset\mathbb{R}^2$ un vector continuo con densidad $h(x,y)$ y $T:A\longrightarrow B\subset\mathbb{R}^2$ difeomorfismo con T^{-1} dado por $x=x(u,v)$ e $y=y(u,v)$. El vector $(U,V)=T(X,Y)$ tiene densidad

$$g(u,v)=\begin{cases} h(x(u,v),y(u,v))|J_{T^{-1}}|, & \text{si } (u,v)\in B,\\ 0, & \text{en otro caso.}\end{cases}$$

$$J_{T^{-1}}(u,v)=\frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{vmatrix}\neq 0 \text{ (Jacobiano del cambio).}$$



Aplicación: Distribución de la suma de variables

Sea (X, Y) un vector con densidad $h(x, y)$. La densidad de la variable $U = X + Y$ es

$$g_1(u) = \int_{-\infty}^{\infty} h(u - v, v)\, dv.$$

En particular, si X, Y son independientes con densidades $f_1(x)$ y $f_2(y)$, entonces la densidad de $X + Y$ es

$$g_1(u) = \int_{-\infty}^{\infty} f_1(u - v) f_2(v)\, dv = f_1 * f_2(u).$$

($f_1 * f_2$ es la **convolución** de f_1 con f_2.)

Ejercicio: Sean $X, Y \sim \mathrm{U}(0, 1)$ independientes. Hallar la distribución de $U = X + Y$. Hacerlo de dos formas: (1) usando el resultado anterior; (2) calculando directamente la f.d. de U.

Aplicación: Suma de variables normales independientes

(a) Sean $X \sim \mathrm{N}(0,\sigma)$, $Y \sim \mathrm{N}(0,\tau)$ independientes. Mostrar que

$$X + Y \sim \mathrm{N}\left(0, \sqrt{\sigma^2+\tau^2}\right).$$

(b) Sean $X \sim \mathrm{N}(a,\sigma)$, $Y \sim \mathrm{N}(b,\tau)$ independientes. Mostrar que

$$X + Y \sim \mathrm{N}\left(a+b, \sqrt{\sigma^2+\tau^2}\right).$$

(c) Sean $X_1,\ldots,X_n$ variables independientes con $X_i \sim \mathrm{N}(\mu_i,\sigma_i)$ $(i=1,\ldots,n)$. Mostrar que

$$X_1 + \cdots + X_n \sim \mathrm{N}\left(\mu_1+\cdots+\mu_n, \sqrt{\sigma_1^2+\cdots+\sigma_n^2}\right).$$

(d) Sean $X_1,\ldots,X_n$ variables independientes con $X_i \sim \mathrm{N}(\mu_i,\sigma_i)$ $(i=1,\ldots,n)$ y $a_1,\ldots,a_n$ constantes. Mostrar que

$$X_1 + \cdots + X_n \sim \mathrm{N}\left(a_1\mu_1+\cdots+a_n\mu_n, \sqrt{a_1^2\sigma_1^2+\cdots+a_n^2\sigma_n^2}\right).$$

Aplicación: Distribución del producto/cociente de variables

Sea (X, Y) un vector con densidad $h(x, y)$. La densidad de la variable $U = XY$ es

$$g_1(u) = \int_{-\infty}^{\infty} h(u/v, v) \frac{1}{|v|} \, dv.$$

Sea (X, Y) un vector con densidad $h(x, y)$. La densidad de la variable $U = X/Y$ es

$$g_1(u) = \int_{-\infty}^{\infty} h(uv, v) |v| \, dv.$$

Ejercicio: Sean $X, Y \sim \mathrm{N}(0, 1)$ independientes. Hallar la distribución de $U = X/Y$.

Caso discreto

Ejemplo 1: Sean $X \sim \mathrm{B}(n_1, p)$, $Y \sim \mathrm{B}(n_2, p)$ independientes. Mostrar que $X + Y \sim \mathrm{B}(n_1 + n_2, p)$.

En general, si $X_i \sim \mathrm{B}(n_i, p)$ $(i = 1, \ldots, k)$ independientes, entonces $X_1 + \cdots + X_k \sim \mathrm{B}(n_1 + \cdots + n_k, p)$.

Ejemplo 2: Sean $X \sim \mathrm{P}(\lambda)$, $Y \sim \mathrm{P}(\mu)$ independientes. Mostrar que $X + Y \sim \mathrm{P}(\lambda + \mu)$.

En general, si $X_i \sim \mathrm{P}(\lambda_i)$ $(i = 1, \ldots, n)$ independientes, entonces $X_1 + \cdots + X_n \sim \mathrm{P}(\lambda_1 + \cdots + \lambda_n)$.

Probabilidad I

Grado en Matemáticas

Tema 4
Esperanza y momentos

Javier Cárcamo

Departamento de Matemáticas
Universidad Autónoma de Madrid
javier.carcamo@uam.es

Descripción del tema

1. Esperanza matemática.
2. Propiedades de la esperanza matemática.
3. La esperanza condicional.
4. Momentos de variables aleatorias.
5. La varianza.
6. Otros valores importantes.
7. Momentos de vectores aleatorios.
8. Covarianza y correlación.

Objetivos principales

- Entender y saber calcular los valores típicos que ayudan a resumir la información de una variable o vector.
- Comprender la covarianza y la correlación entre dos variables como medidas del grado de la relación (lineal) entre ellas.

Idea intuitiva: Sea Ω una población con:

n_1 personas de edad e_1.

n_2 personas de edad e_2.

$\vdots \qquad \vdots \qquad \vdots$

n_k personas de edad e_k.

$n_1 + \cdots + n_k = n \equiv$ número total de individuos.

$$\textbf{Edad media} = \frac{n_1 e_1 + \cdots + n_k e_k}{n} = e_1 \frac{n_1}{n} + \cdots + e_k \frac{n_k}{n}$$

ϵ : se elige un individuo al azar. Variable: $X \equiv$ edad del individuo.
X toma los valores $e_1, \ldots, e_k$ con probabilidades $\mathrm{P}(X = e_i) = \frac{n_i}{n}$.

$\displaystyle\sum_{i=1}^{k} e_i \,\mathrm{P}(X = e_i) \equiv$ **valor medio** o **esperanza matemática** de X.

(Nos da una idea de entorno a qué punto se distribuye la v.a.)

CASO DISCRETO: Sea X una variable aleatoria discreta, la **esperanza (matemática)** o **media** de X se define por:

$$\mathrm{E}X = \sum_x x\,\mathrm{P}(X = x) = \sum_x x\,p(x),$$

siempre que la (suma o) serie sea absolutamente convergente:

$$\sum_x |x|\,\mathrm{P}(X = x) < \infty \qquad \text{(serie abs. convergente).} \qquad (*)$$

CASO CONTINUO: Sea X una variable continua con densidad f, la **esperanza (matemática)** o **media** de X se define por:

$$\mathrm{E}X = \int_{-\infty}^{+\infty} x\,f(x)\,dx,$$

siempre que la integral sea absolutamente convergente, es decir,

$$\int_{-\infty}^{+\infty} |x|\,f(x)\,dx < \infty \qquad \text{(integral abs. convergente).} \qquad (*)$$

Si se verifica $(*)$, se dice que X es una variable **integrable**.

Observación: No todas las variables aleatorias son integrables.

Ejemplo: Sea X la variable discreta con función de probabilidad

$$p(n) = \frac{6}{\pi^2 n^2} \qquad \text{para} \quad n \in \mathbb{N}.$$

X *no* es integrable ya que la serie

$$\sum_{n=1}^{\infty} np(n) = \sum_{n=1}^{\infty} \frac{6}{\pi^2\, n} \quad \text{es divergente.}$$

Ejemplo: Si X con distribución de Cauchy, es decir con densidad

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad x \in \mathbb{R},$$

entonces X *no* es integrable ya que la integral

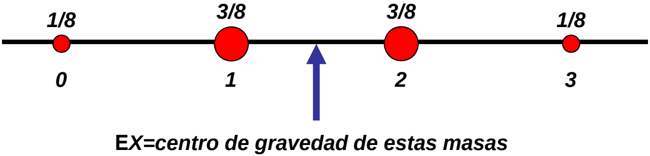
$$\int_{-\infty}^{\infty} \frac{1}{\pi}\frac{|x|}{1+x^2}\, dx \quad \text{es divergente.}$$

Interpretación física (caso discreto): Si X es discreta y sobre cada punto se pone una masa igual a la probabilidad de que X tome ese valor, la $\mathrm{E}X$ se corresponde con el **centro de masas** o **centro de gravedad** de estos puntos.

Ejemplo: $X \equiv$ número de caras al lanzar una moneda 3 veces.

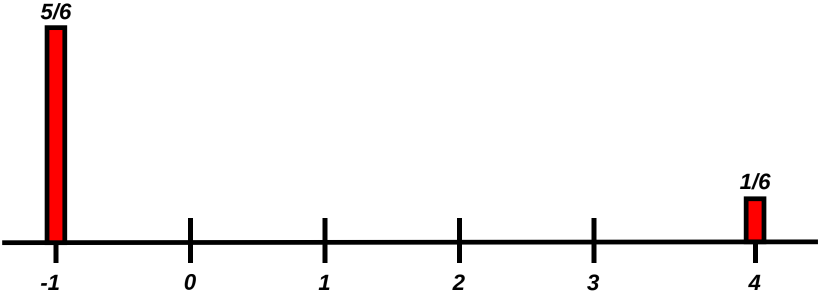
$$p(0) = p(3) = 1/8, \quad p(1) = p(2) = 3/8$$

$$\mathrm{E}X = \sum_x xp(x) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}.$$



Ejemplo: Una de las apuestas en el juego *Craps* es la llamada *Any seven* en la que se apuesta en un solo lanzamiento de dos dados a que se obtienen un total de 7 puntos. La apuesta paga 4 a 1.

La variable G (ganancias) toma los valores -1, 4, con probabilidades $5/6$, $1/6$, respectivamente.



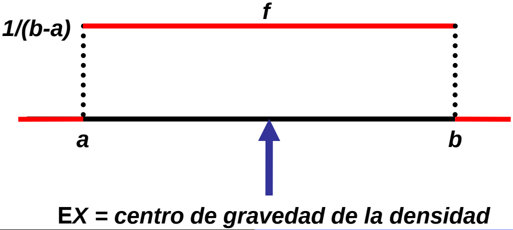
$$\mathrm{E}G = (-1) \cdot \frac{5}{6} + 4 \cdot \frac{1}{6} = -\frac{1}{6} \approx -0{,}16667.$$

Interpretación física (caso continuo): Si X es continua, la $\mathrm{E}X$ se corresponde con el **centro de gravedad** de la densidad de masa de probabilidad de X.

Ejemplo: X con distribución uniforme en el intervalo $[a, b]$.

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

$$\mathrm{E}X = \int_{-\infty}^{+\infty} x f(x)\, dx = \int_a^b \frac{x}{b-a}\, dx = \frac{a+b}{2}.$$



Ejercicio: Una persona tiene 1000 € colocados a un interés del 10 % anual. Por otra parte, se le presenta la oportunidad de incorporar sus mil euros a una empresa cuya rentabilidad anual responde a la siguiente perspectiva:

- Ganar 200 € con probabilidad 0.7.
- Perder 150 € con probabilidad 0.3.

¿Qué decisión crees que debe tomar esta persona? Razona tu respuesta.

Esperanza de funciones de variables aleatorias

Sea $g : \mathbb{R} \longrightarrow \mathbb{R}$ una función (continua) y $X : \Omega \longrightarrow \mathbb{R}$ una variable aleatoria, podemos componer las dos funciones para obtener una nueva variable aleatoria $Y = g(X)$.

$$\begin{aligned} Y = g(X) : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto g(X(\omega)) \end{aligned}$$

Podemos necesitar calcular la esperanza de $g(X)$, $\mathrm{E}g(X)$.

Dada la variable aleatoria X, nos puede interesar calcular

$$\mathrm{E}(3X + 4),\ \mathrm{E}X^2,\ \mathrm{E}X^3,\ \mathrm{E}\log(X),\ \mathrm{E}e^X, \text{ etc.}$$

Ejemplo: Si X cuenta el dinero (en euros) de una persona, nos puede interesar conocer la esperanza de la variable $Y = tX + c$, donde t es la tasa de cambio de euros a dólares y c es la comisión que nos cobrar por realizar la operación.

Esperanza de funciones de variables aleatorias

CASO DISCRETO: Sea $X : \Omega \longrightarrow \mathbb{R}$ una v.a. discreta.

$$\mathrm{E}g(X) = \sum_x g(x)\mathrm{P}(X = x),$$

siempre que la (suma o) serie sea absolutamente convergente:

$$\sum_x |g(x)|\mathrm{P}(X = x) < \infty \qquad \text{(abs. convergente).}$$

CASO CONTINUO: Sea $X : \Omega \longrightarrow \mathbb{R}$ una variable continua con función de densidad f.

$$\mathrm{E}g(X) = \int_{-\infty}^{+\infty} g(x)f(x)\,dx,$$

siempre que la integral sea absolutamente convergente, es decir,

$$\int_{-\infty}^{+\infty} |g(x)|f(x)\,dx < \infty \qquad \text{(abs. convergente).}$$

Esperanza de funciones de variables aleatorias

Observación: X es una variable integrable $\iff \mathrm{E}|X| < \infty$.

Ejercicio 1: Sea X variable que toma los valores -1, 2 y 5 con probabilidades $1/4$, $1/4$ y $1/2$. Calcúlese el valor esperado de $g(X) = X^2$, $\mathrm{E}X^2$.

Ejercicio 2: X con distribución uniforme en el intervalo $[a, b]$, es decir, X con densidad

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

Calcúlese el valor esperado de $g(X) = X^2$, $\mathrm{E}X^2$.

Esperanza de funciones de vectores aleatorios

Si $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ es una función (continua) y $(X, Y) : \Omega \longrightarrow \mathbb{R}^2$ es un vector aleatorio, nos puede interesar calcular la esperanza de la variable $Y = g(X, Y)$, $\mathrm{E}g(X, Y)$.

Dado un vector aleatorio (X, Y), nos puede interesar calcular

$$\mathrm{E}(3X + 4Y),\ \mathrm{E}(X^2 + Y^2),\ \mathrm{E}(XY),\ \mathrm{E}(X/Y^2),\ \mathrm{E}e^{X+Y}, \text{ etc.}$$

Por ejemplo, si (X, Y) miden los beneficios de dos empresas, nos puede interesar estudiar el beneficio total $X + Y$. Si (X, Y) mide el peso y la altura de una persona, quizá necesitemos estudiar el índice de masa corporal o el *índice de Quetelet*, X/Y^2.

En general, si $\mathbf{X} = (X_1, \ldots, X_d)$ es un vector aleatorio y $g : \mathbb{R}^d \longrightarrow \mathbb{R}$ es una función (continua) nos interesará calcular $\mathrm{E}g(X_1, \ldots, X_d)$. Por ejemplo:

$$\mathrm{E}(X_1 + \cdots + X_d),\ \mathrm{E}(X_1^2 + \cdots + X_d^2),\ \mathrm{E}(X_1 \cdots X_d),\ \mathrm{E}e^{X_1+\cdots+X_d}, \ldots$$

Esperanza de funciones de variables aleatorias

CASO DISCRETO: Sea $(X, Y) : \Omega \longrightarrow \mathbb{R}^2$ un vector discreto.

$$\mathrm{E}g(X, Y) = \sum_x \sum_y g(x, y) \mathrm{P}(X = x, Y = y),$$

siempre que la (suma o) serie sea absolutamente convergente:

$$\sum_x \sum_y |g(x, y)| \mathrm{P}(X = x, Y = y) < \infty \qquad \text{(abs. convergente).}$$

CASO CONTINUO: Sea $(X, Y) : \Omega \longrightarrow \mathbb{R}^2$ un vector continuo con densidad conjunta f.

$$Eg(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y)\, dx\, dy,$$

siempre que la integral sea absolutamente convergente, es decir,

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |g(x, y)| f(x, y)\, dx\, dy < \infty \qquad \text{(abs. convergente).}$$

Esperanza de funciones de variables aleatorias

Ejercicio 1: (X, Y) vector discreto con distribución uniforme en los puntos $(0,0)$, $(1,1)$, $(2,1)$, $(1,2)$ y $(2,2)$. Calcular $\mathrm{E}X$, $\mathrm{E}Y$ y el valor esperado de $g(X, Y) = XY$, es decir, $\mathrm{E}(XY)$.

Ejercicio 2: (X, Y) con distribución uniforme en el cuadrado unidad, $[0,1]^2$, es decir, (X, Y) con densidad conjunta

$$f(x,y) = \begin{cases} 1 & \text{si } 0 \leq x, y \leq 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Calcular $\mathrm{E}X$, $\mathrm{E}Y$ y el valor esperado de $g(X, Y) = XY^2$, es decir, $\mathrm{E}(XY^2)$.

1. Para $c \in \mathbb{R}$ constante, $\mathrm{E}(c) = c$.

2. $\mathrm{E}(cX) = c\mathrm{E}X$, si $c \in \mathbb{R}$ constante.

3. **Linealidad:** Si X, Y son integrables y $a, b \in \mathbb{R}$ constantes, entonces $aX + bY$ es integrable y $\mathrm{E}(aX + bY) = a\mathrm{E}X + b\mathrm{E}Y$.

4. **Positividad:** Si $X \geq 0$, entonces $\mathrm{E}X \geq 0$.

5. **Monotonicidad:** Si $X \leq Y$, entonces $\mathrm{E}X \leq \mathrm{E}Y$.

   En particular, $|\mathrm{E}X| \leq \mathrm{E}|X|$.

6. **Independencia:** Si X e Y son independientes e integrables, entonces XY es integrable y $\mathrm{E}(XY) = \mathrm{E}X\mathrm{E}Y$.

Observación 1: Si $g_1, g_2 : \mathbb{R}^2 \longrightarrow \mathbb{R}$ funciones (continuas) y X e Y variables, entonces $g_1(X,Y)$ y $g_2(X,Y)$ son variables.

3' $\mathrm{E}(ag_1(X,Y) + bg_2(X,Y)) = a\mathrm{E}g_1(X,Y) + b\mathrm{E}g_2(X,Y)$, $a, b \in \mathbb{R}$ constantes.

Observación 2: Si $g_1, g_2 : \mathbb{R} \longrightarrow \mathbb{R}$ funciones y X e Y independientes, $g_1(X)$ y $g_2(Y)$ son también independientes.

6' Si X e Y independientes, $\mathrm{E}(g_1(X)g_2(Y)) = \mathrm{E}g_1(X)\mathrm{E}g_2(Y)$.

Ejemplos: Sean X e Y variables

- $\mathrm{E}(2X^2 + 3Y^4) = 2\mathrm{E}X^2 + 3\mathrm{E}Y^4$.
- $\mathrm{E}(X+Y)^2 = \mathrm{E}X^2 + 2\mathrm{E}(XY) + \mathrm{E}Y^2$.
- $\mathrm{E}(\mathrm{sen}(XY) + 3e^X - 2Y^2) = \mathrm{E}\,\mathrm{sen}(XY) + 3\mathrm{E}e^X - 2\mathrm{E}Y^2$.

Si además X e Y son independientes:

- $\mathrm{E}(XY)^2 = \mathrm{E}X^2\mathrm{E}Y^2$.
- $\mathrm{E}(Xe^Y) = \mathrm{E}X\mathrm{E}e^Y$.
- $\mathrm{E}e^{X+Y} = \mathrm{E}e^X\mathrm{E}e^Y$.

Ejercicio: Expresa en función de las densidades las siguientes esperanzas.

Sea X variable aleatoria con densidad $f(x)$.

$\mathrm{E}X^3 =$

$\mathrm{E}e^{tX} =$

Sea (X, Y) vector aleatorio con densidad $f(x, y)$.

$\mathrm{E}\left(\frac{X^2}{X^4+Y^4}\right) =$

Sea X, Y v.a. independientes con densidades $f_1(x)$ y $f_2(y)$.

$\mathrm{E}e^{X+Y} =$

CASO DISCRETO: Sea (X, Y) un vector discreto y $x \in \mathbb{R}$ con $\mathrm{P}(X = x) > 0$. La **esperanza condicional** de la variable Y dado $X = x$ es

$$\mathrm{E}(Y|X = x) = \sum_y y\mathrm{P}(Y = y|X = x) = \sum_y y\frac{\mathrm{P}(X = x, Y = y)}{\mathrm{P}(X = x)}.$$

De forma análoga se define $\mathrm{E}(X|Y = y)$, si $\mathrm{P}(Y = y) > 0$.

CASO CONTINUO: Sea (X, Y) un vector aleatorio continuo con densidad conjunta f y densidades marginales f_1 y f_2. Si $x \in \mathbb{R}$ con $f_1(x) > 0$, la **esperanza condicional** de Y dado $X = x$ es

$$\mathrm{E}(Y|X = x) = \int_{-\infty}^{\infty} y\, f_{Y|X=x}(y)\, dy = \int_{-\infty}^{\infty} y\frac{f(x, y)}{f_1(x)}\, dy.$$

De forma análoga se define $\mathrm{E}(X|Y = y)$, si $f_2(y) > 0$.

La **variable esperanza condicional** se define mediante:

$$T(\omega) = \mathrm{E}(Y|X = X(\omega)), \qquad \omega \in \Omega.$$

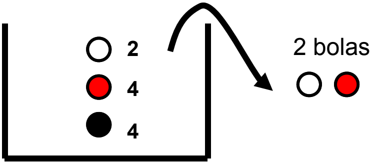
La variable esperanza condicional, $T = \mathrm{E}(Y|X)$ es una función de la variable X.

CASO DISCRETO: Sea (X, Y) un vector discreto. La variable esperanza condicional $\mathrm{E}(Y|X)$ toma los valores $\mathrm{E}(Y|X = x)$ con probabilidad $\mathrm{P}(X = x) > 0$. (Si $\mathrm{P}(X = x) = 0$, podemos definir $\mathrm{E}(Y|X = x) = 0$.)

CASO CONTINUO: Sea (X, Y) un vector aleatorio continuo con densidad conjunta f y densidades marginales f_1 y f_2. La variable esperanza condicional $\mathrm{E}(Y|X)$ toma los valores $\mathrm{E}(Y|X = x)$ con densidad de probabilidad $f_1(x)$.

Ejercicio 1: Se extraen dos bolas sin reemplazamiento de una urna con 10 bolas (2 blancas, 4 negras y 4 rojas). Consideramos

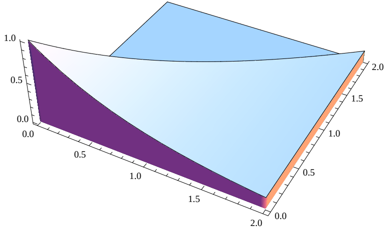
$$X \equiv \text{ número de bolas blancas en la extracción.}$$
$$Y \equiv \text{ número de bolas rojas en la extracción.}$$



(a) Calcula la función de probabilidad conjunta del vector (X, Y).
(b) Calcula las funciones de probabilidad de X e Y, $\mathrm{E}X$ y $\mathrm{E}Y$.
(c) Calcula los valores $\mathrm{E}(Y|X=0)$, $\mathrm{E}(Y|X=1)$ y $\mathrm{E}(Y|X=2)$.
(d) Calcula la distribución de $T = \mathrm{E}(Y|X)$ y $\mathrm{E}(\mathrm{E}(Y|X))$.
(e) Repite los cálculos con $\mathrm{E}(X|Y)$.

Ejercicio 2: Sea (X, Y) un vector con densidad conjunta

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



(a) Calcula e identifica las densidades marginales $f_1(x)$ y $f_2(y)$, $\mathrm{E}X$ y $\mathrm{E}Y$.

(b) Calcula e identifica la distribución de $Y|X = x$ (para $x > 0$).

(c) Calcula $\mathrm{E}(Y|X = x)$ (para $x > 0$).

(d) ¿Cuál es la distribución de $T = \mathrm{E}(Y|X)$? Calcula $\mathrm{E}(\mathrm{E}(Y|X))$.

Teorema: Regla de la doble esperanza

Si Y es una variable integrable, entonces $\mathrm{E}(Y|X)$ es una variable finita. Además, se verifica la **regla de la doble esperanza**:

$$\mathrm{E}(\mathrm{E}(Y|X)) = \mathrm{E}Y.$$

Aplicación: Si X variable discreta con valores $\{x_1, x_2, \ldots\}$, tenemos

$$\mathrm{E}(Y) = \sum_i \mathrm{P}(X = x_i)\mathrm{E}(Y|X = x_i).$$

Esta fórmula se conoce como **regla de la esperanza total** (o **teorema de la partición**).

Nota: Puede ocurrir que X *no* sea integrable y que exista la variable $\mathrm{E}(X|Y)$.

Ejemplo: Sea (X, Y) un vector con densidad conjunta

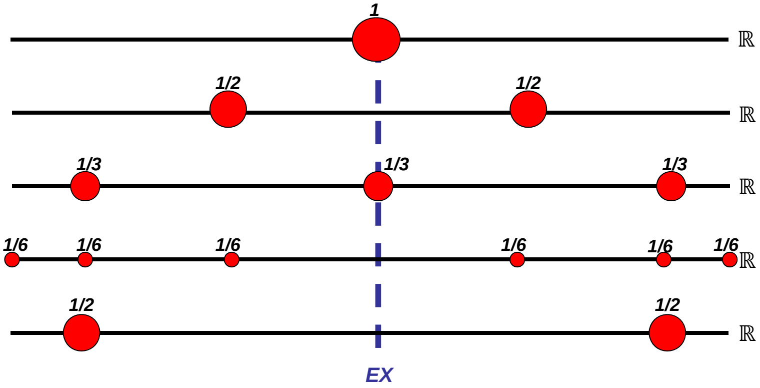
$$f(x, y) = \begin{cases} x\, e^{-x(1+y)}, & \text{si } x, y \geq 0, \\ 0, & \text{en otro caso.} \end{cases}$$

(a) Calcula las densidades marginales $f_1(x)$ y $f_2(y)$ (hecho en el tema anterior), ¿Es la variable Y integrable?

(b) Calcula e identifica la distribución de $Y|X = x$ (para $x > 0$).

(c) Calcula $\mathrm{E}(Y|X = x)$ (para $x > 0$) y la variable $\mathrm{E}(Y|X)$.

Idea intuitiva: Existen muchas variables aleatorias igual media.



Objetivo: Necesitamos más valores característicos que nos ayuden a distinguir entre estas variables. Nos interesa tener parámetros para conocer el grado de dispersión de la variable o diferentes valores que nos ayuden a conocer la forma de la distribución de probabilidad.

Para definir estos valores necesitamos de los *momentos* de una variable.

Momentos

Sea X una variable aleatoria y $n \in \mathbb{N}$. Se llama **momento de orden** n de X a

$$\alpha_n = \mathrm{E}X^n.$$

Notación: $\alpha_1 = \mu = \mathrm{E}X$.

CASO DISCRETO: Si X es v.a. discreta:

$$\alpha_n = \sum_x x^n \mathrm{P}(X = x).$$

CASO CONTINUO: Si X es v.a. continua con densidad f:

$$\alpha_n = \int_{\mathbb{R}} x^n f(x)\, dx.$$

Ejercicio 1: Sea $X \sim \mathrm{B}(1; p)$ (Bernoulli). Calcula α_n.

Ejercicio 2: Sea $X \sim \mathrm{U}(a, b)$ (Uniforme). Calcula α_n.

Momentos centrados

Sea X una variable aleatoria, $n \in \mathbb{N}$ y $c \in \mathbb{R}$.

Se llama **momento de orden n de X alrededor de c** a

$$\mu_{n,c} = \mathrm{E}(X - c)^n.$$

Se llama **el momento centrado de orden n de X** a

$$\mu_n = \mathrm{E}(X - \mathrm{E}X)^n = \mathrm{E}(X - \mu)^n.$$

CASO DISCRETO: Si X es v.a. discreta:

$$\mu_n = \sum_x (x - \mu)^n \mathrm{P}(X = x).$$

CASO CONTINUO: Si X es v.a. continua con densidad f:

$$\mu_n = \int_{\mathbb{R}} (x - \mu)^n f(x)\, dx.$$

Sea X una variable, se llama **varianza** de X al segundo momento centrado de X, es decir,

$$\mathrm{Var}(X) = \mu_2 = \mathrm{E}(X - \mathrm{E}X)^2 = \mathrm{E}(X - \mu)^2.$$

Notación: $\mathrm{Var}(X) = \sigma^2$.

CASO DISCRETO: Si X es v.a. discreta:

$$\sigma^2 = \sum_x (x - \mu)^2 \mathrm{P}(X = x).$$

CASO CONTINUO: Si X es v.a. continua con densidad f:

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

Idea: La varianza de una variable es el promedio de lo que dista la variable de su valor esperado. (Se eleva al cuadrado para que no haya compensación entre los signos y su computo sea más sencillo.) La Var$X$ es una medida de la dispersión de la variable respecto a la media.

Observación: Si la variable X se expresa en una determinada unidad de medida, la $\text{Var}X$ está dada en esas unidades al cuadrado. Para subsanar este problema se define la desviación típica.

Sea X una variable aleatoria, se llama **desviación típica** de X a la raíz cuadrada de la varianza de X, es decir,

$$\sigma = \sqrt{\text{Var}X} = \sqrt{\sigma^2}.$$

Ejercicio 1: Sea X v.a. con distribución de Bernoulli de parámetro p. Calcula la desviación típica de X.

Ejercicio 2: Sea X v.a. con distribución uniforme en el intervalo $[a, b]$. Calcula la desviación típica de X.

Propiedades de la varianza

Sean X e Y variables y a, $b \in \mathbb{R}$ constantes. Se tiene:

1. $\text{Var}X \geq 0$.
2. $\text{Var}X = 0$ si y sólo si X es una constante.
3. $\text{Var}X = \mathrm{E}X^2 - (\mathrm{E}X)^2$.
4. Si X e Y ind., entonces $\text{Var}(X+Y) = \text{Var}X + \text{Var}Y$.
5. $\text{Var}(aX+b) = a^2\text{Var}X$.
6. $\text{Var}X = \text{mín}_{c\in\mathbb{R}}\, \alpha_{2,c}$.

   (La Var$X$ es el momento mínimo entre los de segundo orden).

Observación: La propiedad 4 se puede sustituir por la más general:

4' Si X e Y (no nec. ind.) con $\mathrm{E}(XY) = \mathrm{E}X\mathrm{E}Y$, entonces $\text{Var}(X+Y) = \text{Var}X + \text{Var}Y$.

Nota: La esperanza μ y la varianza σ^2 son dos de los valores más representativos de una distribución de probabilidad.

Valores típicos de las distribuciones más notables

Ejercicio: Calcula la esperanza y varianza de las variables:

Discretas

1. Bernoulli: $\mathrm{B}(1; p)$.
2. Binomial: $\mathrm{B}(n; p)$.
3. Geométrica: $\mathrm{G}(p)$.
4. Binom. neg.: $\mathrm{BN}(t; p)$.
5. Poisson $\mathrm{P}(\lambda)$.

Continuas

6. Uniforme: $\mathrm{U}(a, b)$.
7. Normal: $\mathrm{N}(\mu; \sigma)$.
8. Cauchy.
9. Exponencial: $\mathrm{Exp}(\lambda)$.
10. Gamma: $\mathrm{Gamma}(\alpha; \beta)$.
11. Beta: $\mathrm{Beta}(\alpha; \beta)$.
12. Weibull: $\mathrm{Weibull}(\theta; k)$.
13. Pareto: $\mathrm{Pareto}(a; \theta)$.
14. Lognormal: $\mathrm{LogN}(\mu; \sigma)$.

Tipificación de variables

Una variable X se dice **tipificada** si $\mathrm{E}X = 0$ y $\mathrm{Var}X = 1$.

Si X es una variable con media μ y varianza $\sigma^2 > 0$, se define la **variable aleatoria tipificada** de X como la variable

$$X^* = \frac{X - \mu}{\sigma}.$$

Observaciones:

(a) $\mathrm{E}X^* = 0$ y $\mathrm{Var}X^* = 1$.

(b) X^* es **adimensional**.

Aplicaciones: Para comparar valores (x e y) de dos variables X e Y, expresadas en medidas diferentes, podemos comparar los valores tipificados (x^* e y^*). De esta manera también podemos comparar la misma variable en distintas poblaciones.

Ejercicio: Calcular la variable tipificada de una v.a. con distribución uniforme en el intervalo $[a, b]$ y exponencial.

Los siguientes valores (coeficiente de asimetría y coeficiente de curtosis) nos dan información sobre la forma de la distribución.

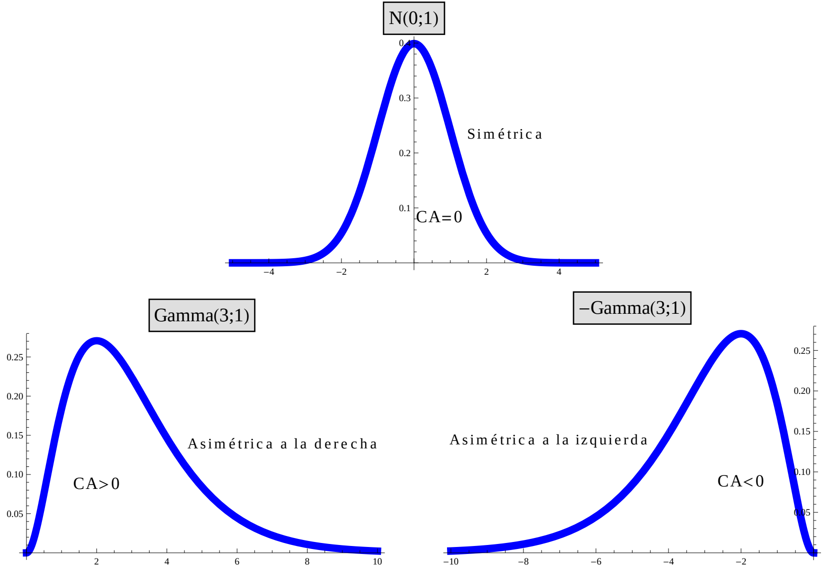
El **coeficiente de asimetría** de una variable X se define por

$$\mathsf{CA} = \frac{\mu_3}{\sigma^3} = \frac{\mathrm{E}(X - \mathrm{E}X)^3}{(\mathrm{E}(X - \mathrm{E}X)^2)^{3/2}}.$$

El coeficiente de asimetría nos dice si la distribución es simétrica alrededor de su media.

- Si CA $= 0$, la distribución es **simétrica**.
- Si CA > 0, la distribución es **asimétrica a la derecha**.
- Si CA < 0, la distribución es **asimétrica a la izquierda**.

Coeficiente de asimetría



El **coeficiente de curtosis o de apuntamiento** de una variable X:

$$\text{CAp} = \frac{\mu_4}{\sigma^4} = \frac{\mathrm{E}(X - \mathrm{E}X)^4}{\left(\mathrm{E}(X - \mathrm{E}X)^2\right)^2}.$$

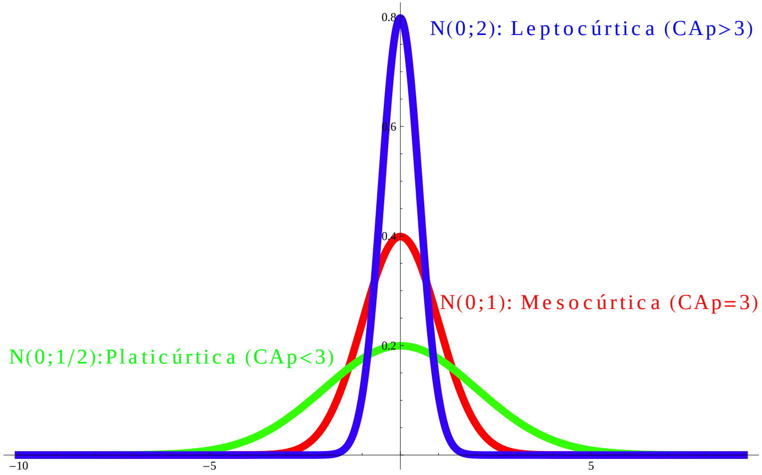
CAp mide el apuntamiento o aplastamiento de X respecto de la distribución normal:

- Si $\text{CAp} = 3$, X es **mesocúrtica**. (Como N(0;1).)
- Si $\text{CAp} < 3$, X es **platicúrtica**. (Más aplastada que N(0;1).)
- Si $\text{CAp} > 3$, X es **leptocúrtica**. (Más apuntada que N(0;1).)

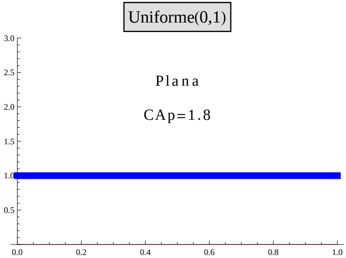
CAp también permite vislumbrar, sin necesidad de conocer todos los valores de la distribución, si la distribución tiene forma de campana o de U.

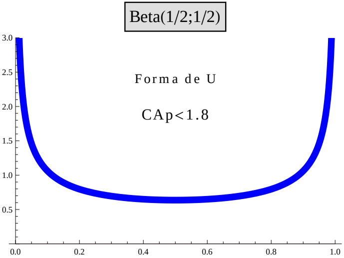
- Si $\text{CAp} = 1{,}8$, la distribución es como la uniforme.
- Si $\text{CAp} > 1{,}8$, la distribución es campaniforme.
- Si $\text{CAp} < 1{,}8$, la distribución tiene forma de U.

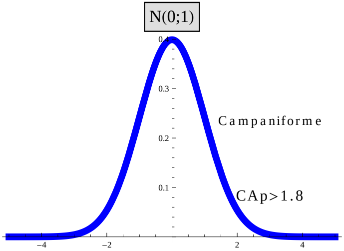
Coeficiente de curtosis o de apuntamiento



Coeficiente de curtosis o de apuntamiento







Idea: En muchas ocasiones nos interesa conocer cómo varía una variable con respecto a otra, o el grado de relación que dos variables mantienen. Para averiguar estas relaciones necesitamos definir los momentos de un vector aleatorio.

Sea (X, Y) un vector aleatorio. Para $n_1,\ n_2 \geq 0$, **los momentos de orden** $n = n_1 + n_2$ del vector (X, Y) se definen por:

$$\alpha_{n_1,n_2} = \mathrm{E}(X^{n_1} Y^{n_2}).$$

CASO DISCRETO: Si (X, Y) es vector discreto:

$$\alpha_{n_1,n_2} = \sum_x \sum_y x^{n_1} y^{n_2} \mathrm{P}(X = x, Y = y).$$

CASO CONTINUO: Si (X, Y) continuo con densidad conjunta f:

$$\alpha_{n_1,n_2} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^{n_1} y^{n_2} f(x, y)\, dx\, dy.$$

Se define el **vector de medias** de (X, Y), $\boldsymbol{\mu}$, mediante:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} \mathrm{E}X \\ \mathrm{E}Y \end{bmatrix} \qquad \text{(momentos de orden 1).}$$

Sea (X, Y) con vector de medias $(\mu_X, \mu_Y)'$. Para n_1, $n_2 \geq 0$, los **momentos centrados de orden** $n = n_1 + n_2$ de (X, Y) se definen por:

$$\mu_{n_1,n_2} = \mathrm{E}((X - \mu_X)^{n_1}(Y - \mu_Y)^{n_2}).$$

CASO DISCRETO: Si (X, Y) es vector discreto:

$$\mu_{n_1,n_2} = \sum_x \sum_y (x - \mu_X)^{n_1}(y - \mu_Y)^{n_2}\mathrm{P}(X = x, Y = y).$$

CASO CONTINUO: Si (X, Y) continuo con densidad conjunta f:

$$\mu_{n_1,n_2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - \mu_X)^{n_1}(y - \mu_Y)^{n_2} f(x, y)\, dx\, dy.$$

Sea (X, Y) un vector aleatorio. La **covarianza** de X e Y se define

$$\mathrm{Cov}(X, Y) = \mu_{1,1} = \mathrm{E}((X - \mathrm{E}X)(Y - \mathrm{E}Y)).$$

CASO DISCRETO: Si (X, Y) es vector discreto:

$$\mathrm{Cov}(X, Y) = \sum_x \sum_y (x - \mu_X)(y - \mu_Y)\mathrm{P}(X = x, Y = y).$$

CASO CONTINUO: Si (X, Y) continuo con densidad conjunta f:

$$\mathrm{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y)\, dx\, dy.$$

Observaciones:

- La $\mathrm{Cov}(X, Y)$ viene expresada en las unidades de X por las unidades de Y.
- La $\mathrm{Cov}(X, Y)$ expresa "de alguna manera que precisaremos más adelante" la variación (relativa) de una variable respecto de la otra.

Sea (X, Y) un vector aleatorio. La **covarianza** de X e Y se define

$$\text{Cov}(X, Y) = \mathrm{E}((X - \mathrm{E}X)(Y - \mathrm{E}Y)).$$

Propiedades de la covarianza

1. $\text{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}X\mathrm{E}Y$.
2. Si $Z = aX + b$ y $T = cY + d$, $\text{Cov}(Z, T) = ac\text{Cov}(X, Y)$.
3. $\text{Cov}(X, X) = \text{Var}(X)$.
4. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
5. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.

Sea (X, Y) un vector aleatorio. La **covarianza** de X e Y se define

$$\text{Cov}(X, Y) = \mathrm{E}((X - \mathrm{E}X)(Y - \mathrm{E}Y)) = \mathrm{E}(XY) - \mathrm{E}X\mathrm{E}Y.$$

Incorrelación

Dos variables X e Y se dicen **incorrelacionadas** o **incorreladas** si $\text{Cov}(X, Y) = 0$, es decir, si $\mathrm{E}(XY) = \mathrm{E}X\mathrm{E}Y$.

Observaciones:

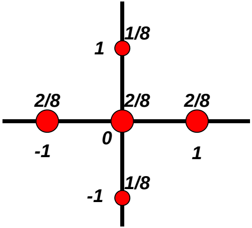
- La incorrelación quiere decir que no existe relación lineal entre las variables. Esto lo aclararemos enseguida con otro valor importante, el *coeficiente de correlación lineal de Pearson*.
- X e Y incorreladas $\Leftrightarrow \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.
- X e Y independientes $\Rightarrow$ X e Y incorreladas.
- X e Y incorreladas $\not\Rightarrow$ X e Y independientes.

Dos variables X e Y se dicen **incorrelacionadas** o **incorreladas** si $\mathrm{Cov}(X, Y) = 0$, es decir, si $\mathrm{E}(XY) = \mathrm{E}X\mathrm{E}Y$.

Contraejemplo 1: X e Y incorreladas $\not\Rightarrow$ X e Y independientes.

(X, Y) vector discreto con función de probabilidad conjunta:

X \ Y	-1	0	1
-1	0	2/8	0
0	1/8	2/8	1/8
1	0	2/8	0



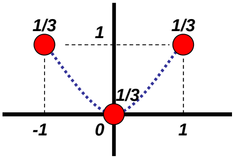
Ejercicio 1: Comprobar que X e Y son incorreladas, pero no independientes.

Dos variables X e Y se dicen **incorrelacionadas** o **incorreladas** si $\mathrm{Cov}(X,Y)=0$, es decir, si $\mathrm{E}(XY)=\mathrm{E}X\mathrm{E}Y$.

Contraejemplo 2: X e Y incorreladas $\not\Rightarrow$ X e Y independientes.

(X,Y) vector discreto con función de probabilidad conjunta:

X \ Y	0	1
-1	0	1/3
0	1/3	0
1	0	1/3



Ejercicio 2: Comprobar que X e Y son incorreladas, pero no independientes.

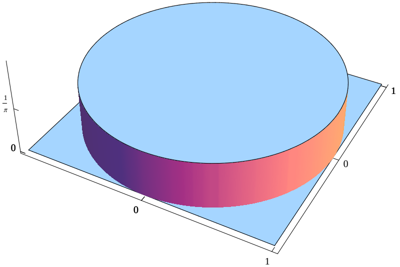
Observación: En este ejemplo existe una dependencia funcional perfecta entre X e Y ($Y=|X|$ ó $Y=X^2$, por ejemplo).

Dos variables X e Y se dicen **incorrelacionadas** o **incorreladas** si $\mathrm{Cov}(X, Y) = 0$, es decir, si $\mathrm{E}(XY) = \mathrm{E}X\mathrm{E}Y$.

Contraejemplo 3: X e Y incorreladas $\not\Rightarrow$ X e Y independientes.

(X, Y) vector discreto con función de densidad conjunta:

$$f(x, y) = 1/\pi 1_{\{x^2+y^2<1\}}(x, y)$$



Ejercicio 3: Comprobar que X e Y son incorreladas, pero no independientes.

Desigualdad de Cauchy-Schwarz

Si X e Y son variables con $\mathrm{E}X^2$, $\mathrm{E}Y^2 < \infty$, entonces

$$\mathrm{E}|XY| \leq \sqrt{\mathrm{E}(X^2)\mathrm{E}(Y^2)}.$$

Aplicación: $|\mathrm{Cov}(X,Y)| \leq \sigma_X \sigma_Y$.

Observación: La covarianza tiene el inconveniente de que depende de las unidades de medida de las variables (X e Y). Para corregir este defecto se define el *coeficiente de correlación*.

X e Y variables no degeneradas, **el coeficiente de correlación (lineal de Pearson)** entre X e Y está dado por:

$$\mathrm{Corr}(X,Y) = \rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

Observación: $\rho_{X,Y}$ es una medida *adimensional* que cuantifica el grado de asociación lineal entre las variables X e Y.

X e Y variables no degeneradas, **el coeficiente de correlación (lineal de Pearson)** entre X e Y está dado por:

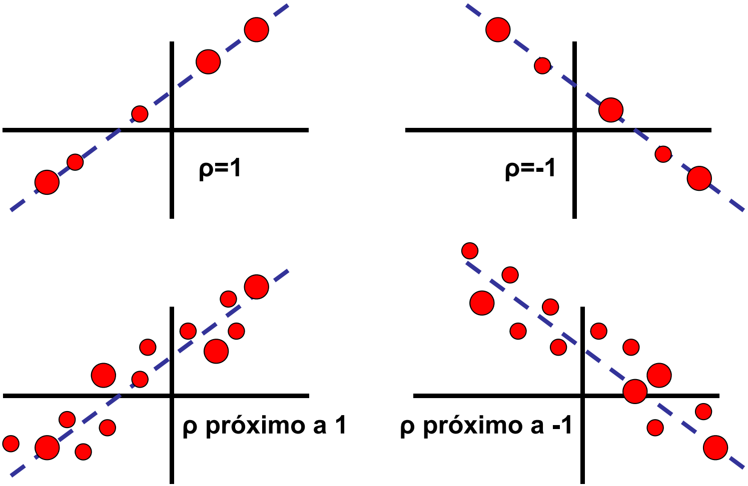
$$\mathrm{Corr}(X, Y) = \rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Propiedades del coeficiente de correlación

1. $\rho_{X,Y} \in [-1, 1]$.
2. $\rho_{X,Y} = 0 \Leftrightarrow X$ e Y incorrelados.
3. $\rho_{X,Y} = 1 \Leftrightarrow Y = aX + b$, $a > 0$, $b \in \mathbb{R}$.
   (Dependencia lineal positiva perfecta).
4. $\rho_{X,Y} = -1 \Leftrightarrow Y = aX + b$, $a < 0$, $b \in \mathbb{R}$.
   (Dependencia lineal negativa perfecta).

Observación: Cuando $\rho_{X,Y}$ está cerca de 1 ($\rho_{X,Y} \geq 0{,}9$), se dice que hay una *dependencia lineal positiva alta* entre X e Y. Análogamente, si $\rho_{X,Y}$ está cerca de -1 ($\rho_{X,Y} \leq -0{,}9$) se dice que hay una *dependencia lineal negativa alta* entre X e Y.

Correlación y dependencia lineal



Caso d dimensional

Sea $\mathbf{X} = (X_1, \ldots, X_d)$ es un vector aleatorio d dimensional. Los conceptos anteriores se pueden generalizar de manera obvia.

La **esperanza** de $\mathbf{X} = (X_1, \ldots, X_d)$ o **vector de medias** de $\mathbf{X}$ es

$$\mathrm{E}\mathbf{X} = \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix} = \begin{bmatrix} \mathrm{E}X_1 \\ \mathrm{E}X_2 \\ \vdots \\ \mathrm{E}X_d \end{bmatrix}$$

La **matriz de covarianzas** y **matriz de correlaciones** de $\mathbf{X}$:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,d} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d,1} & \sigma_{d,2} & \cdots & \sigma_{d,d} \end{bmatrix} ; \; \boldsymbol{R} = \begin{bmatrix} \rho_{1,1} & \rho_{1,2} & \cdots & \rho_{1,d} \\ \rho_{2,1} & \rho_{2,2} & \cdots & \rho_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d,1} & \rho_{d,2} & \cdots & \rho_{d,d} \end{bmatrix},$$

donde $\sigma_{i,j} = \mathrm{Cov}(X_i, X_j)$ y $\rho_{i,j} = \rho_{X_i,X_j}$.

Ejercicio: (X, Y) vector aleatorio discreto con función de probabilidad conjunta:

$X\backslash Y$	1	2
1	$1/9$	$2/9$
2	$2/9$	$4/9$

(a) Calcúlese $\mathrm{E}(X+Y)$, $\mathrm{E}(2X+3Y)$.

(b) El vector de medias, la matriz de covarianzas y la matriz de correlaciones del vector (X, Y).

(c) ¿Son X e Y independientes? ¿son incorreladas?

Probabilidad I

Grado en Matemáticas

Tema 5
Funciones características

Javier Cárcamo

Departamento de Matemáticas
Universidad Autónoma de Madrid
javier.carcamo@uam.es

Descripción del tema

1. Números complejos.
2. La exponencial compleja.
3. Variables aleatorias complejas.
4. Funciones características.
5. Momentos y derivadas de la f.c.
6. Fórmulas de inversión.
7. Identificación de funciones características.
8. Aplicaciones.

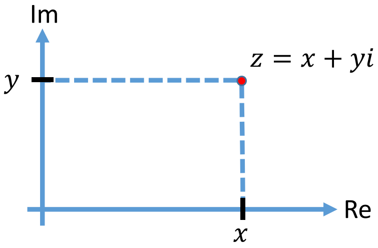
Objetivos principales

- Entender la utilidad de las funciones características en la Teoría de la probabilidad.
- Identificar y saber manejar las funciones características de las distribuciones más importantes.

Un **número complejo** es un número que se puede expresar de la forma $x + yi$, donde x e y son reales e $i = \sqrt{-1}$ es la **unidad imaginaria**. El **plano complejo** es el cuerpo

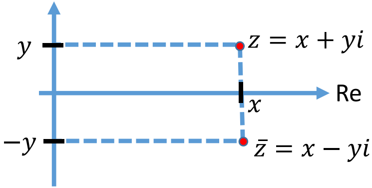
$$\mathbb{C} = \{z = x + yi : x, y \in \mathbb{R}\}.$$

Podemos representar cada número complejo $z = x + yi \in \mathbb{C}$ en el plano $\mathbb{R}^2$ identificando z con el vector (x, y).



La **parte real** del número complejo $z = x + yi$ es x $(\mathrm{Re}(z) = x)$ y la **parte imaginaria** es y $(\mathrm{Im}(z) = y)$.

El **conjugado** del número complejo $z = x + yi$ es $\bar{z} = x - yi$.

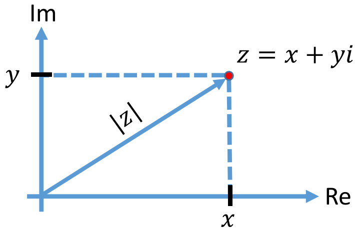


Para $z, w \in \mathbb{C}$, tenemos,

- $\bar{z}$ es una reflexión de z respecto al eje de abscisas ($\bar{\bar{z}} = z$).
- $z \in \mathbb{R} \Leftrightarrow z = \bar{z}$.
- $\mathrm{Re}(z) = (z + \bar{z})/2$.
- $\mathrm{Im}(z) = (z - \bar{z})/(2i)$.
- $\overline{z \pm w} = \bar{z} \pm \bar{w}$.
- $\overline{zw} = \bar{z}\bar{w}$.
- Si $w \neq 0$, $\overline{(z/w)} = \bar{z}/\bar{w}$.
- Si $z \neq 0$, $1/z = \bar{z}/(z\bar{z})$.

El **módulo** o **valor absoluto** del número complejo $z = x + yi$ es

$$|z| = \sqrt{x^2 + y^2} = \sqrt{\mathrm{Re}(z)^2 + \mathrm{Im}(z)^2}.$$



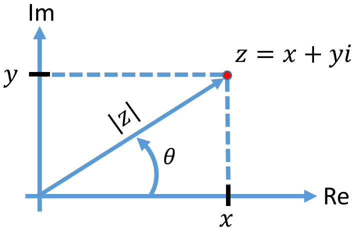
Para $z, w \in \mathbb{C}$, tenemos,

- $|z| = 0 \Leftrightarrow z = 0$.
- $|z|^2 = z\bar{z}$.
- $|z + w| \leq |z| + |w|$.
- $|zw| = |z||w|$ y si $z \neq 0$, $|1/z| = 1/|z|$.
- $|w - z| \geq |w| - |z|$.

El **argumento** del número complejo $z = x + yi$ es

$$\arg(z) = \arctan(y/x).$$

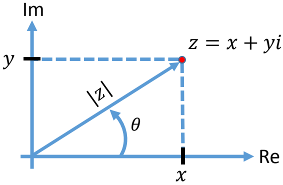
(Entendiéndose la función arctan definida en los cuatro cuadrantes.)



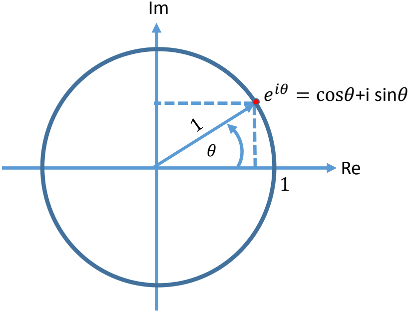
Si $\arctan(x) \in (-\pi/2, \pi/2)$ $(x \in \mathbb{R})$, $\arg(z) \in (-\pi, \pi]$ dada por:

$$\arg(z) = \begin{cases} \arctan(y/x), & \text{si } x > 0, \\ \arctan(y/x) + \pi, & \text{si } x < 0 \text{ e } y \geq 0, \\ \arctan(y/x) - \pi, & \text{si } x < 0 \text{ e } y < 0, \\ \pi/2, & \text{si } x = 0 \text{ e } y > 0, \\ -\pi/2, & \text{si } x = 0 \text{ e } y < 0, \\ \text{indefinido}, & \text{si } x = y = 0. \end{cases}$$

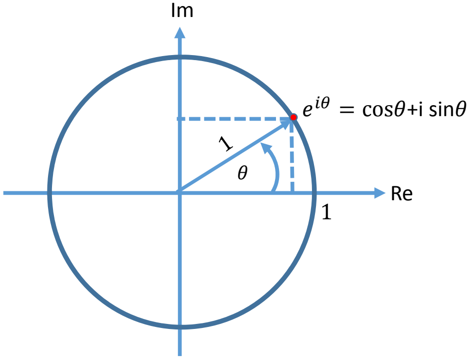
La **representación en polares** del número complejo $z = x + yi$ es $z = r\cos\theta + ir\sin\theta$, donde $r = |z|$ y $\theta = \arg(z)$.



Usando la **fórmula de Euler**, $e^{i\theta} = \cos\theta + i\sin\theta$, $z = re^{i\theta}$.



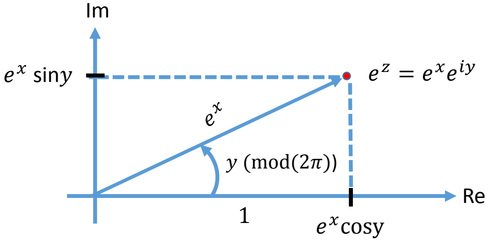
La **fórmula de Euler**, $e^{i\theta} = \cos\theta + i\sin\theta$, $\theta \in \mathbb{R}$.



- $|e^{i\theta}| = 1 \quad \text{y} \quad \overline{e^{i\theta}} = e^{-i\theta}$.
- $\cos x = \dfrac{e^{ix} + e^{-ix}}{2}$.
- $\sin x = \dfrac{e^{ix} - e^{-ix}}{2i}$.

Para $z = x + iy \in \mathbb{C}$, la función **exponencial compleja** de z,

$$e^z = e^{x+iy} := e^x e^{iy} = e^x(\cos y + i \sin y) = e^x \cos y + ie^x \sin y.$$



- $e^0 = 1$, $\quad e^{z+w} = e^z e^w \quad$ y $\quad e^z \neq 0$, para todo $z \in \mathbb{C}$.
- Para todo $z \in \mathbb{C}$, $e^z = \displaystyle\sum_{n=0}^{\infty} \frac{z^n}{n!}$.
- Para todo $z \in \mathbb{C}$, $e^z = \displaystyle\lim_{n\to\infty} \left(1 + \frac{z}{n}\right)^n$.
- Si $z_n \to z$, entonces $e^z = \displaystyle\lim_{n\to\infty} \left(1 + \frac{z_n}{n}\right)^n$.

$$Z : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{C} \text{ aplicación.}$$

Sean $X = \mathrm{Re}(Z)$, $Y = \mathrm{Im}(Z)$ parte real e imaginaria de Z, es decir $X, Y : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}$ y

$$Z = X + iY$$

La aplicación Z se dice **variable aleatoria compleja** si y solo si (X, Y) es un vector aleatorio bidimensional (es decir, si X e Y son variables aleatorias unidimensionales).

Diremos que la variable compleja Z es **integrable** si X e Y son integrables. En tal caso, la **esperanza** de Z se define

$$\mathrm{E}Z = \mathrm{E}X + i\mathrm{E}Y.$$

Propiedades de la esperanza de variables complejas

1. Si Z integrable, entonces $\mathrm{E}Z \in \mathbb{C}$.
2. Z integrable si y solo si $\mathrm{E}|Z| < \infty$ ($|Z|$ es una variable aleatoria real).
3. **Linealidad:** Si Z_1, Z_2 v.a. complejas integrables y $a, b \in \mathbb{C}$, entonces $aZ_1 + bZ_2$ es integrable y
$$\mathrm{E}(aZ_1 + bZ_2) = a\mathrm{E}Z_1 + b\mathrm{E}Z_2.$$
4. Si Z es integrable, entonces $|\mathrm{E}Z| \leq \mathrm{E}|Z|$.

Sea X una variable aleatoria. Se llama **función característica (f.c.)** de X a la función $\varphi_X : \mathbb{R} \to \mathbb{C}$ dada por

$$\varphi_X(t) = \mathrm{E}(e^{itX}), \quad t \in \mathbb{R}.$$

Notas:

1. $\varphi_X(t)$ es una función bien definida para todo $t \in \mathbb{R}$.
2. $|\varphi_X(t)| \leq 1$.
3. $\varphi_X(0) = 1$.
4. $\varphi_X(t) = \mathrm{E}(\cos(tX)) + i\mathrm{E}(\sin(tX))$.
5. Si $X =_d Y$, entonces $\varphi_X(t) = \varphi_Y(t)$, $t \in \mathbb{R}$.

Observación: Veremos más adelante que la función característica de una v.a. X, φ_X, caracteriza su distribución de probabilidad P_X. Es decir, en el punto anterior se puede escribir un "si y solo si".

Propiedades básicas

- $\varphi_X(-t) = \overline{\varphi_X(t)}$. ($\varphi_X(t) \in \mathbb{R}$ si y solo si $\varphi_X(t)$ es par.)
- **Cambio de origen y escala:** Sea $Y = aX + b$ con $a, b \in \mathbb{R}$. Se tiene $\varphi_Y(t) = e^{itb}\varphi_X(at)$. En particular, $\varphi_{-X}(t) = \overline{\varphi_X(t)}$.

Una v.a. X se dice **simétrica** si $X =_d -X$. Si X tiene densidad $f(x)$ y es simétrica, entonces $f(x)$ es par ($f(x) = f(-x)$, $x \in \mathbb{R}$).

- **Caso de v.a. simétricas:** Si X tiene densidad par, entonces $\varphi_X(t) = \mathrm{E}(\cos tX) \in \mathbb{R}$, $t \in \mathbb{R}$.

  X simétrica, entonces $\varphi_X(t)$ es real ($\Longleftrightarrow \varphi_X$ es par).
- **Independencia:** $X_1, \ldots, X_n$ v.a. independientes Se tiene:

$$\varphi_{X_1+\cdots+X_n}(t) = \varphi_{X_1}(t) \cdot \cdots \cdot \varphi_{X_n}(t).$$

  Nota: El recíproco no es cierto.

  Aplicación importante: Suma de variables i.i.d.
- **Continuidad:** φ_X es una función uniformemente continua.

1. $X = c$ c.s. (constante), $\qquad \varphi_X(t) = e^{itc}$.
2. $X \sim \mathrm{B}(1; p)$, $\qquad \varphi_X(t) = (q + pe^{it})$.
3. $X \sim \mathrm{B}(n; p)$, $\qquad \varphi_X(t) = (q + pe^{it})^n$.
4. $X \sim \mathrm{P}(\lambda)$, $\qquad \varphi_X(t) = e^{\lambda(e^{it}-1)}$.
5. $X \sim \mathrm{G}(p)$, $\qquad \varphi_X(t) = \dfrac{p}{1 - qe^{it}}$.
6. $X \sim \mathrm{BN}(r; p)$, $\qquad \varphi_X(t) = \left(\dfrac{p}{1 - qe^{it}}\right)^r$.
7. $X \sim \mathrm{U}(a, b)$, $\qquad \varphi_X(t) = \dfrac{e^{itb} - e^{ita}}{it(b-a)}$.

   En particular,

   - $X \sim \mathrm{U}(-1, 1)$, $\qquad \varphi_X(t) = \dfrac{\sin t}{t}$.
   - $X \sim \mathrm{U}(-c, c)$, $\qquad \varphi_X(t) = \dfrac{\sin tc}{tc}$.

8. X con densidad triangular $f(x) = \text{máx}\{1 - |x|, 0\}$,

$$\varphi_X(t) = 2\left(\frac{1 - \cos t}{t^2}\right).$$

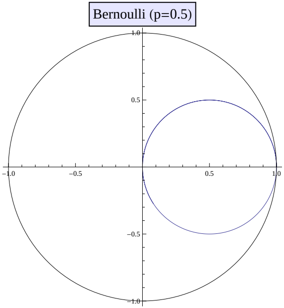
9. $X \sim \text{Exp}(a)$, $\qquad \varphi_X(t) = \left(1 - \frac{it}{a}\right)^{-1}.$

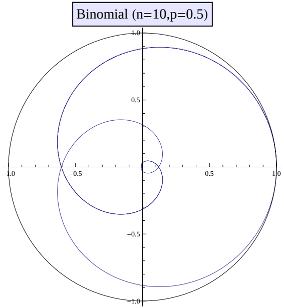
10. X Cauchy $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$, $\qquad \varphi_X(t) = e^{-|t|}.$

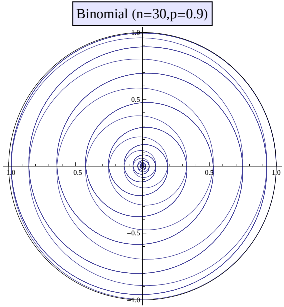
11. $X \sim \text{Gamma}(\alpha, \beta)$ $(\alpha, \beta > 0)$

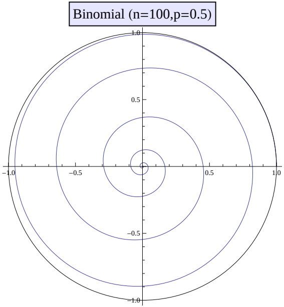
$$f(x) = \frac{x^{\alpha-1} e^{-\beta x} \beta^\alpha}{\Gamma(\alpha)}, \ (x > 0), \qquad \varphi_X(t) = \left(1 - \frac{it}{\beta}\right)^{-\alpha}.$$

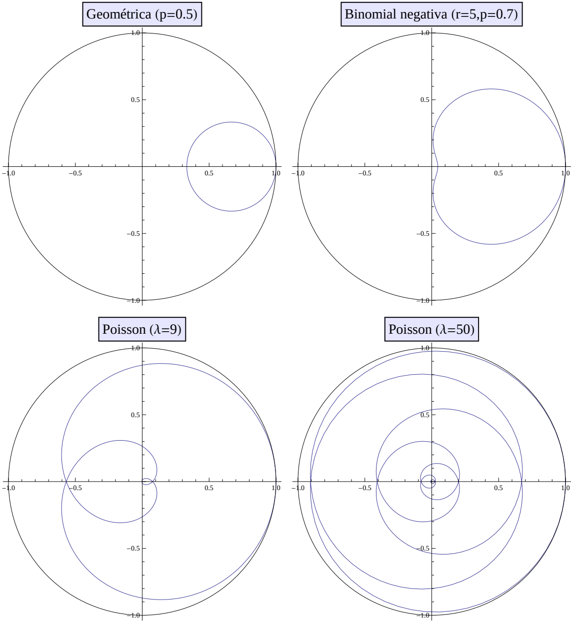
12. $X \sim \text{N}(0,1)$, $\qquad \varphi_X(t) = e^{-t^2/2}.$


Bernoulli (p=0.5)
1.0
0.5
−1.0
−0.5
0.5
1.0
−0.5
−1.0

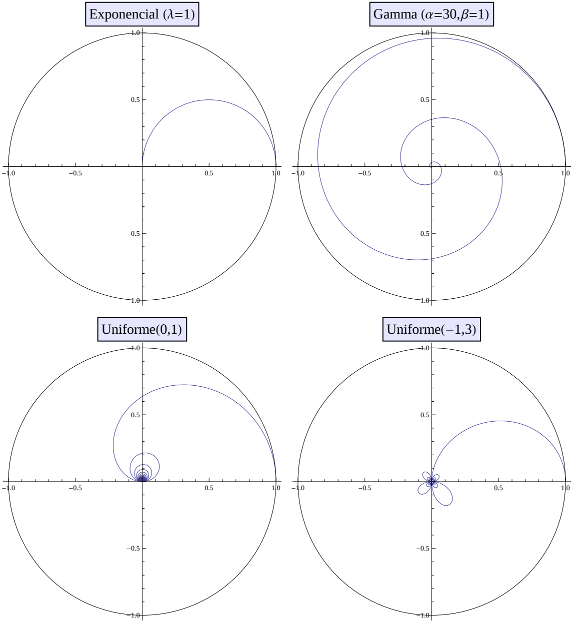

Binomial (n=10,p=0.5)
1.0
0.5
−1.0
−0.5
0.5
1.0
−0.5
−1.0


Binomial (n=30,p=0.9)
1.0
0.5
−1.0
−0.5
0.5
1.0
−0.5
−1.0


Binomial (n=100,p=0.5)
1.0
0.5
−1.0
−0.5
0.5
1.0
−0.5
−1.0


Geométrica (p=0.5)
Binomial negativa (r=5,p=0.7)
Poisson (λ=9)
Poisson (λ=50)
1.0
0.5
−0.5
−1.0

4. Funciones características. Dibujos



Idea: Si todo fuera maravilloso...

$$\varphi_X(t) = \mathrm{E} e^{itX}$$

... y la derivada de la esperanza fuera la esperanza de la derivada:

$$\begin{array}{rcl}
\varphi'_X(t) = \mathrm{E}\left[iXe^{itX}\right] & \Longrightarrow & \varphi'_X(0) = i\mathrm{E}X \\
\varphi''_X(t) = \mathrm{E}\left[(iX)^2 e^{itX}\right] & \Longrightarrow & \varphi''_X(0) = i^2\mathrm{E}X^2 \\
\varphi'''_X(t) = \mathrm{E}\left[(iX)^3 e^{itX}\right] & \Longrightarrow & \varphi'''_X(0) = i^3\mathrm{E}X^3 \\
\vdots & \vdots & \vdots \\
\varphi^{(k)}_X(t) = \mathrm{E}\left[(iX)^k e^{itX}\right] & \Longrightarrow & \varphi^{(k)}_X(0) = i^k\mathrm{E}X^k
\end{array}$$

Conclusión: En un mundo ideal, derivando sucesivamente la f.c. φ_X (y evaluando las derivadas en 0) obtendríamos (salvo las constantes i^k) los momentos de la variable aleatoria X.

Teorema: Momentos de la v.a. y derivadas de la f.c.

Supongamos que $\mathrm{E}|X|^n < \infty$ para algún $n \in \mathbb{N}$. Se tiene:

(a) φ_X es derivable hasta el orden n y

$$\varphi_X^{(k)}(t) = i^k \mathrm{E}\big[X^k e^{itX}\big], \quad k = 1, \ldots, n.$$

En particular, $\varphi_X^{(k)}(0) = i^k \mathrm{E}X^k$, $k = 1, \ldots, n$.

(b) Para $k = 1, \ldots, n$, $\varphi_X^{(k)}$ es uniformemente continua.

Observación: El recíproco de este teorema no es cierto en general. Hay ejemplos de v.a. tales que existe $\varphi'_X(0)$, pero $\mathrm{E}|X| = \infty$.

Nota (importante de cara a las aplicaciones): Supongamos que $f : (-a, a) \longrightarrow \mathbb{C}$ admite derivadas en $t = 0$ hasta el orden n. Usando el desarrollo de McLaurin, tenemos

$$f(t) = f(0) + f'(0)t + \frac{f''(0)}{2!}\, t^2 + \cdots + \frac{f^{(n)}(0)}{n!}\, t^n + R_n(t),$$

donde $R_n(t) = o(t^n)$ $(t \to 0)$, es decir, $R_n(t)/t^n \to 0$, si $t \to 0$.

Si $\mathrm{E}|X|^n < \infty$, entonces φ_X es n veces derivable y por tanto

$$\boxed{\varphi_X(t) = \sum_{k=0}^{n} \frac{i^k \mathrm{E}X^k}{k!} t^k + o(t^n), \quad t \to 0.}$$

Ejemplo: Supongamos que $X_1, X_2, \ldots,$ v.a. i.i.d. integrables de media μ. Calcular

$$\lim_{n\to\infty} \varphi_{S_n/n}(t), \quad \text{donde} \quad S_n = X_1 + \cdots + X_n.$$

Teorema: Desarrollo en serie de la f.c.

Supongamos que existe un $\rho > 0$ tal que $\mathrm{E}e^{\rho|X|} < \infty$. La función φ_X admite un desarrollo en serie de potencias

$$\varphi_X(t) = \sum_{n=0}^{\infty} \frac{i^n \mathrm{E}X^n}{n!} t^n, \quad |t| < \rho.$$

Aplicaciones:

1. $X \sim \mathrm{N}(0,1)$.
2. X con distribución de Laplace o doble exponencial.
3. $X \sim \mathrm{Gamma}(\alpha, \beta)$.

Observación: Conocida la distribución de una variable aleatoria X, podemos calcular (al menos teóricamente) su función característica, φ_X.

Las **fórmulas de inversión** tratan el problema inverso, es decir, conocida la f.c. φ_X, se trata de encontrar la distribución de probabilidad de la v.a. X.

Caso discreto

Teorema: Fórmula de inversión para variables con valores en $\mathbb{Z}$

Sea X una v.a. con soporte en $\mathbb{Z}$ y con f.c. φ_X. Se tiene:

$$\mathrm{P}(X = n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itn} \varphi_X(t)\, dt, \quad n \in \mathbb{Z}.$$

$$\mathrm{P}(X = n) = \frac{1}{2\pi} \int_{0}^{2\pi} e^{-itn} \varphi_X(t)\, dt, \quad n \in \mathbb{Z}.$$

Teorema: Fórmula de inversión de Lévy

Sea X una v.a. con f.c. φ_X y f.d. F_X. Para todo $a, b \in \mathbb{R}$ con $a < b$, se tiene

$$\frac{\mathrm{P}(X=a)+\mathrm{P}(X=b)}{2}+\mathrm{P}(a<X<b)=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\varphi_X(t)\,dt.$$

En particular, si $a, b \in C_F$ (puntos de continuidad de F_X)

$$F_X(b)-F_X(a)=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\varphi_X(t)\,dt.$$

Corolario 1: Teorema de unicidad

Sean X e Y v.a. no necesariamente definidas sobre el mismo espacio de probabilidad. Si $\varphi_X = \varphi_Y$, entonces $X =_d Y$.

Corolario 2: Variables simétricas

X es v.a. simétrica si y solo si φ_X real (si y solo si φ_X par).

Nota: Un caso importante es cuando φ_X es una función integrable, es decir, $\varphi_X \in \mathcal{L}(-\infty, \infty)$

$$\int_{-\infty}^{+\infty} |\varphi_X(t)|\, dt < \infty.$$

Teorema: Caso $\varphi_X \in \mathcal{L}(-\infty, \infty)$

Supongamos que $\varphi_X \in \mathcal{L}(-\infty, \infty)$. Se tiene:

(a) X es una variable absolutamente continua.

(b) La densidad de X es

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_X(t)\, dt.$$

Problema: Dada $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$, ¿cómo saber si φ es la f.c. de alguna variable aleatoria?

- Si $\varphi(0) \neq 1$, entonces φ no es f.c.
- Si existe $t_0 \in \mathbb{R}$ tal que $|\varphi(t_0)| > 1$, entonces φ no es f.c.

Teorema de Bochner-Herglotz

Sea $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$. φ es la f.c. de alguna v.a. si y solo si

(1) $\varphi(0) = 1$.

(2) φ es continua en $t = 0$.

(3) φ es *definida positiva*, es decir, para todo $n \in \mathbb{N}$, para todo $t_1, \ldots, t_n \in \mathbb{R}$, y, para todo $z_1, \ldots, z_n \in \mathbb{C}$, se tiene

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \varphi(t_i - t_j) z_i \overline{z}_j \geq 0.$$

Nota: El teorema anterior es de interés teórico, pero poco práctico.

Problema: Dada $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$, ¿cómo saber si φ es la f.c. de alguna variable aleatoria?

Teorema de Pólya

Sea $\varphi : \mathbb{R} \longrightarrow \mathbb{R}_+$. φ es la f.c. de alguna v.a. (simétrica) si

(1) $\varphi(0) = 1$.

(2) φ es continua en $t = 0$.

(3) φ es par.

(4) φ es convexa y decreciente en $[0, \infty)$.

Normas a seguir para averiguar si φ es una f.c.

1. ¿Es φ una de las f.c. conocidas?
2. φ puede expresarse de alguna manera mediante f.c. conocidas.

   (1) $\varphi(t) = \phi(at)$, donde ϕ es la f.c. de X y $a \in \mathbb{R}$. Entonces, φ es la f.c. de aX.

   (2) $\varphi(t) = \phi_1(t)\phi_2(t)$, donde ϕ_1 es la f.c. de X y ϕ_2 es la f.c. de Y. Entonces, φ es la f.c. de $X + Y$, con X e Y independientes.

   (3) Si $\varphi(t) = e^{itb}\phi(at)$, donde ϕ es la f.c. de X y $a, b \in \mathbb{R}$. Entonces, φ es la f.c. de $aX + b$.

   (4) $\varphi(t) = (\phi(t))^n$, donde ϕ es la f.c. de X y $n \in \mathbb{N}$. Entonces, φ es la f.c. de $X_1 + \cdots + X_n$, con $X_1, \ldots, X_n$ v.a. i.i.d. como X.
3. Aplicar el Teorema de Pólya si se puede.
4. Sospechar que φ *no* es f.c.

Las funciones características tienen muchas aplicaciones. Se utilizan para demostrar importantes resultados límite del Cálculo de probabilidades, como veremos más adelante. También se utilizan en problemas relativos a distribuciones de probabilidad.

1. **Distribuciones de sumas de v.a. independientes**

$$\left.\begin{array}{l} X, Y \text{ independientes} \\ X \sim \mathrm{B}(n; p) \\ Y \sim \mathrm{B}(m; p) \end{array}\right\} \Longrightarrow X + Y \sim \mathrm{B}(n+m; p).$$

$$\left.\begin{array}{l} X, Y \text{ independientes} \\ X \sim \mathrm{P}(\lambda) \\ Y \sim \mathrm{P}(\mu) \end{array}\right\} \Longrightarrow X + Y \sim \mathrm{P}(\lambda+\mu).$$

$$\left.\begin{array}{l} X, Y \text{ independientes} \\ X \sim \mathrm{BN}(r; p) \\ Y \sim \mathrm{BN}(s; p) \end{array}\right\} \Longrightarrow X + Y \sim \mathrm{BN}(r+s; p).$$

1. **Distribuciones de sumas de v.a. independientes**

$$\left.\begin{array}{l} X, Y \text{ independientes} \\ X \sim \text{Gamma}(\alpha_1; \beta) \\ Y \sim \text{Gamma}(\alpha_2; \beta) \end{array}\right\} \Longrightarrow X + Y \sim \text{Gamma}(\alpha_1 + \alpha_2; \beta).$$

$$\left.\begin{array}{l} X, Y \text{ independientes} \\ X \sim \text{N}(a; \sigma) \\ Y \sim \text{N}(b; \tau) \end{array}\right\} \Longrightarrow X + Y \sim \text{N}(a + b; \sqrt{\sigma + \tau}).$$

2. **Otro tipo de problemas: Identificación de distribuciones**

   (1) ¿Existen v.a. X, Y ind. y con la misma distribución tal que $X - Y \sim \text{U}(-1, 1)$?

   (2) Sean X, Y v.a. independientes con la misma distribución de media 0 y varianza 1 tales que

$$\frac{X + Y}{\sqrt{2}} =_d X =_d Y.$$

   Mostrar que la distribución común es necesariamente normal.

Probabilidad I

Grado en Matemáticas

Tema 6
Teoremas límite

Javier Cárcamo

Departamento de Matemáticas
Universidad Autónoma de Madrid
javier.carcamo@uam.es

Descripción del tema

1. La Ley de grandes números.
2. El teorema central del límite.

Objetivos principales

- Familiarizarse con los distintos modos de convergencia.
- Comprender las leyes de grandes números más sencillas.
- Entender la importancia y aplicabilidad del TCL.

Introducción

Los resultados más célebres e importantes de la Teoría de la Probabilidad son los conocidos como **leyes de los grandes números**. Tales leyes no son otra cosa que teoremas que aseguran cierta convergencia de variables aleatorias bajo unas condiciones determinadas. Ahora bien, hay varios modos de convergencia de variables aleatorias.

En este tema se analizan brevemente esos modos de convergencia y se muestran las principales leyes de los grandes números.

Ejemplo introductorio

Realizamos lanzamientos sucesivos e independientes de una moneda equilibrada.

$$X_i = \begin{cases} 1, & \text{si en el } i\text{-ésimo lanzamiento sale cara,} \\ 0, & \text{si en el } i\text{-ésimo lanzamiento sale cruz.} \end{cases}$$

$X_1, X_2, \ldots$ sucesión de v.a. independientes de Bernoulli, $\mathrm{B}(1; 1/2)$.

$$S_n = X_1 + \cdots + X_n.$$

($S_n \equiv$ número de caras en los n primeros lanzamientos)

$$\frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n}.$$

($S_n/n \equiv$ proporción de caras en los n primeros lanzamientos)

Pregunta: ¿Qué ocurre con S_n/n cuando n es grande ($n \to \infty$)?

Algunas simulaciones: Caso $X \sim \mathrm{B}(n; p = 1/2)$

Simulación: Realizamos $n = 5000$ lanzamientos de la moneda.

```
lista = RandomInteger[BinomialDistribution[1, 0.5], 5000];
Take[lista, 25]
```

{0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1}

$$\texttt{e[k\_Integer] := Extract[lista, k]; p[n\_] :=} \frac{1}{n}\sum_{i=1}^{n} \texttt{e[i];}$$

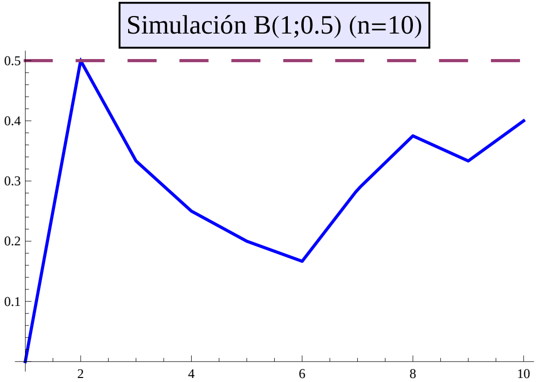
```
lista2 = {}; Do[lista2 = Append[lista2, {n, p[n]}], {n, 1, 5000}];
TableForm[Take[N[lista2], 10], TableHeadings → {None, {"n", "p[n]"}}]
```

n	p[n]
1.	0.
2.	0.5
3.	0.333333
4.	0.25
5.	0.2
6.	0.166667
7.	0.285714
8.	0.375
9.	0.333333
10.	0.4

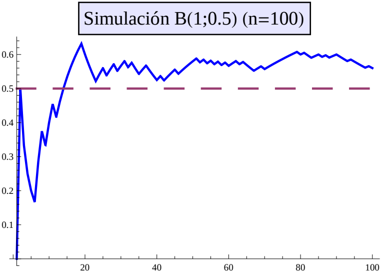
Algunas simulaciones: Caso $X \sim \mathrm{B}(n; p = 1/2)$

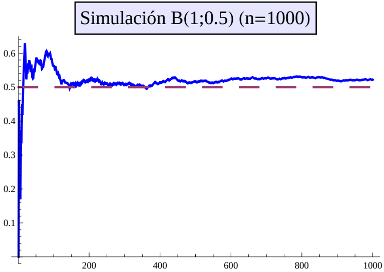
```
int = Interpolation[lista2, InterpolationOrder → 1];
Plot[{int[x], 0.5}, {x, 1, 10}, PlotRange → All,
 PlotStyle → {{Blue, Thickness[.006]}, {Thickness[.006], Dashing[{0.05, 0.05}]}}]
```

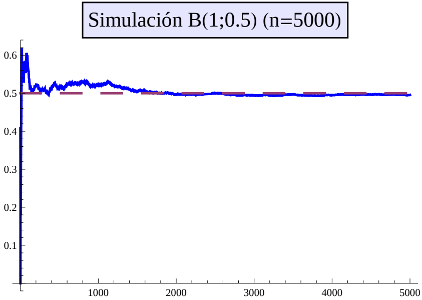
n	p[n]
1.	0.
2.	0.5
3.	0.333333
4.	0.25
5.	0.2
6.	0.166667
7.	0.285714
8.	0.375
9.	0.333333
10.	0.4

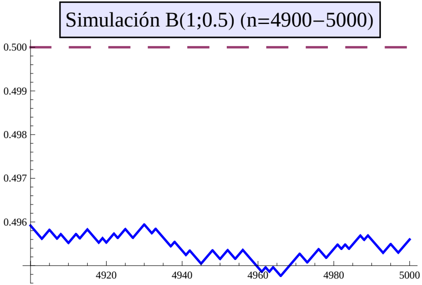


{0, 1, 0, 0, 0, 0, 1, 1, 0, 1}


Simulación B(1;0.5) (n=100)
0.6
0.5
0.4
0.3
0.2
0.1
20
40
60
80
100


Simulación B(1;0.5) (n=1000)
0.6
0.5
0.4
0.3
0.2
0.1
200
400
600
800
1000


Simulación B(1;0.5) (n=5000)
0.6
0.5
0.4
0.3
0.2
0.1
1000
2000
3000
4000
5000


Simulación B(1;0.5) (n=4900−5000)
0.500
0.499
0.498
0.497
0.496
4920
4940
4960
4980
5000

Algunas simulaciones: Caso $X \sim \mathrm{B}(n; p = 0{,}75)$

Simulación: Realizamos $n = 5000$ lanzamientos de la moneda trucada con probabilidad de cara 0,75..

```
lista3 = RandomInteger[BinomialDistribution[1, 0.75], 5000];
Take[lista3, 25]
```

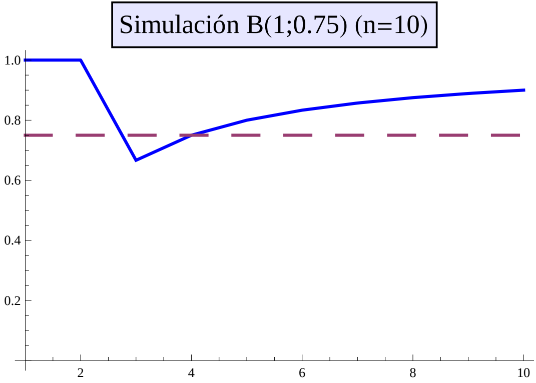
{1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0}

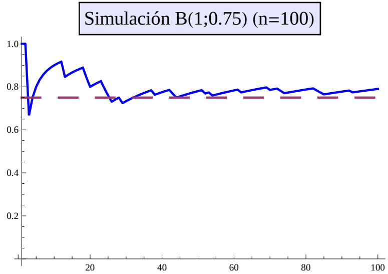
e2[k_Integer] := Extract[lista3, k]; p2[n_] := $\frac{1}{n}\sum_{i=1}^{n}$ e2[i];

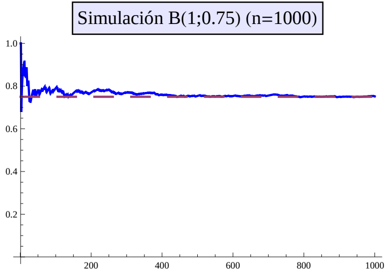
n	p2[n]
1.	1.
2.	1.
3.	0.666667
4.	0.75
5.	0.8
6.	0.833333
7.	0.857143
8.	0.875
9.	0.888889
10.	0.9

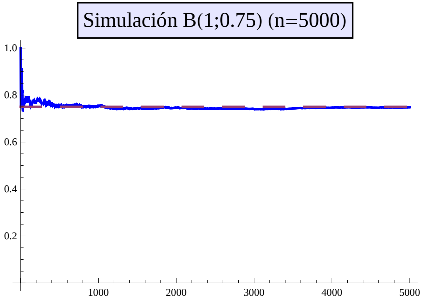
Algunas simulaciones: Caso $X \sim \mathrm{B}(n; p = 0{,}75)$

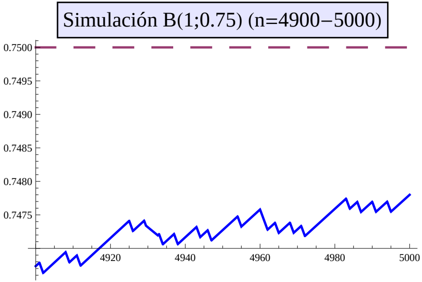
n	p2[n]
1.	1.
2.	1.
3.	0.666667
4.	0.75
5.	0.8
6.	0.833333
7.	0.857143
8.	0.875
9.	0.888889
10.	0.9




Simulación B(1;0.75) (n=100)
1.0
0.8
0.6
0.4
0.2
20
40
60
80
100


Simulación B(1;0.75) (n=1000)
1.0
0.8
0.6
0.4
0.2
200
400
600
800
1000


Simulación B(1;0.75) (n=5000)
1.0
0.8
0.6
0.4
0.2
1000
2000
3000
4000
5000


Simulación B(1;0.75) (n=4900−5000)
0.7500
0.7495
0.7490
0.7485
0.7480
0.7475
4920
4940
4960
4980
5000

Conclusiones de las simulaciones

- Si $X \sim \mathrm{B}(1; p)$, parece claro que, en algún sentido,

$$S_n/n \to p, \qquad \text{cuando } n \to \infty.$$

- $p = \mathrm{E}X$, cuando $X \sim \mathrm{B}(1; p)$.
- Como S_n/n es la media de la muestra $X_1, \ldots, X_n$ (**media muestral**), parece razonable que, en general,

$$S_n/n \to \mu, \qquad \text{cuando } n \to \infty \text{ (en algún sentido).}$$

Objetivo del tema: Estudiar la convergencia de una sucesión de variables aleatorias. (Especialmente nos interesa estudiar el comportamiento de la media muestral.)

Nota: Hay diferentes formas de convergencia de variables aleatorias.

Convergencia en media cuadrática

Sea $X, X_1, X_2, \ldots$ variables aleatorias. Decimos que X_n **converge a X en media cuadrática**, y escribimos $X_n \xrightarrow{\text{m-2}} X$, si

$$\mathrm{E}(X_n - X)^2 \to 0, \qquad n \to \infty.$$

Idea: Si $X_n \xrightarrow{\text{m-2}} X$, el valor esperado de la distancia (al cuadrado) de las variables de la sucesión a la variable límite tiende a cero, cuando n crece.

Teorema: Ley de grandes números en media cuadrática

Sean $X_1, X_2, \ldots,$ variables independientes con igual media μ y varianza σ^2. Se tiene

$$\frac{S_n}{n} \xrightarrow{\text{m-2}} \mu, \quad n \to \infty.$$

Pregunta: ¿Se pueden rebajar las condiciones del teorema anterior?

Convergencia en probabilidad

Sea $X, X_1, X_2, \ldots$ variables aleatorias. Decimos que X_n **converge a X en probabilidad**, y escribimos $X_n \xrightarrow{\mathrm{P}} X$, si

$$\text{para cada } \epsilon > 0,\ \mathrm{P}(|X_n - X| > \epsilon) \to 0, \quad n \to \infty.$$

Idea: Si $X_n \xrightarrow{\mathrm{P}} X$, a largo plazo, cada variable de la sucesión está muy concentrada entorno a la variable límite.

Teorema: Desigualdad de Markov

Sea X v.a. Para $\epsilon > 0$, $\alpha > 0$, se tiene

$$\mathrm{P}(|X| \geq \epsilon) \leq \frac{\mathrm{E}|X|^\alpha}{\epsilon^\alpha}.$$

Corolario: Desigualdad de Chebyshev

Sea X una variable integrable. Para todo $\epsilon > 0$, se tiene

$$\mathrm{P}(|X - \mathrm{E}X| \geq \epsilon) \leq \frac{\mathrm{Var}(X)}{\epsilon^2}.$$

Convergencia en probabilidad

Sea $X, X_1, X_2, \ldots$ variables aleatorias. Decimos que X_n **converge a X en probabilidad**, y escribimos $X_n \xrightarrow{\mathrm{P}} X$, si

$$\text{para cada } \epsilon > 0,\ \mathrm{P}(|X_n - X| > \epsilon) \to 0, \quad n \to \infty.$$

Idea: Si $X_n \xrightarrow{\mathrm{P}} X$, a largo plazo, cada variable de la sucesión está muy concentrada entorno a la variable límite.

Nota: Relación entre los modos de convergencia

- $X_n \xrightarrow{\text{m-2}} X \Longrightarrow X_n \xrightarrow{\mathrm{P}} X$.
- $X_n \xrightarrow{\mathrm{P}} X \not\Longrightarrow X_n \xrightarrow{\text{m-2}} X$.

Teorema: Ley débil de grandes números

Sean $X_1, X_2, \ldots,$ variables incorreladas con igual media μ y varianza σ^2. Se tiene

$$\frac{S_n}{n} \xrightarrow{\mathrm{P}} \mu, \quad n \to \infty.$$

Nota histórica: La ley débil de Bernoulli

Históricamente, la primera ley de grandes números fue obtenida por Bernoulli en el siglo XVIII. La ley se refiere a la proporción de caras que se obtienen en n lanzamientos de una p-moneda (moneda que cae de cara con probabilidad p) y al modo en que esa proporción se aproxima a p cuando el número de lanzamientos es grande. En términos más o menos coloquiales dice lo siguiente:

Haciendo un número suficientemente grande de lanzamientos, podemos conseguir que la probabilidad de que dicha proporción se diferencie de p en menos de una cantidad prefijada esté tan próxima a 1 como queramos.

Teorema: Ley débil de Bernoulli

Si $X_1, X_2, \ldots$ son independientes y tienen la misma distribución de Bernoulli de parámetro p, entonces

$$\frac{S_n}{n} \xrightarrow{\mathrm{P}} p.$$

Convergencia casi segura

Sea $X, X_1, X_2, \ldots$ variables aleatorias. Decimos que X_n **converge a** X **casi seguramente** o **con probabilidad 1**, y escribimos $X_n \xrightarrow{\text{c.s.}} X$, si

$$\mathrm{P}(X_n \to X) = \mathrm{P}(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1.$$

Idea: Si $X_n \xrightarrow{\text{c.s.}} X$, salvo en un conjunto de probabilidad nula se da la convergencia puntual de X_n a X.

Nota: Relación entre los modos de convergencia

- $X_n \xrightarrow{\text{c.s.}} X \Longrightarrow X_n \xrightarrow{\mathrm{P}} X$.
- $X_n \xrightarrow{\mathrm{P}} X \not\Longrightarrow X_n \xrightarrow{\text{c.s.}} X$.
- $X_n \xrightarrow{\text{c.s.}} X \not\Longrightarrow X_n \xrightarrow{\text{m-2}} X$.
- $X_n \xrightarrow{\text{m-2}} X \not\Longrightarrow X_n \xrightarrow{\text{c.s.}} X$.

Convergencia casi segura

Sea $X, X_1, X_2, \ldots$ variables aleatorias. Decimos que X_n **converge a** X **casi seguramente** o **con probabilidad 1**, y escribimos $X_n \xrightarrow{\text{c.s.}} X$, si

$$\mathrm{P}(X_n \to X) = \mathrm{P}(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1.$$

Idea: Si $X_n \xrightarrow{\text{c.s.}} X$, salvo en un conjunto de probabilidad nula se da la convergencia puntual de X_n a X.

Teorema: Ley fuerte de grandes números (Kolmogorov)

Sean $X_1, X_2, \ldots,$ variables independientes con igual distribución y media μ. Se tiene

$$\frac{S_n}{n} \xrightarrow{\text{c.s.}} \mu, \quad n \to \infty.$$

Nota: La Ley fuerte de grandes números es uno de los resultados más importantes de la Teoría de la probabilidad.

Resumen de las ideas principales

- Existen diferentes modos de convergencia de una sucesión de v.a.
- Los **modos de convergencia fuertes** son: convergencia casi segura (principalmente) y convergencia en media cuadrática. Los **modos de convergencia débiles** son: convergencia en probabilidad y sobre todo convergencia en distribución (siguiente punto).
- La **ley fuerte de grandes números** asegura que la media (aritmética) de una sucesión de variables independientes y con igual distribución (iid) converge con probabilidad 1 a μ (media poblacional).
- Esto resuelve el problema introductorio y confirma las simulaciones. La ley fuerte de grandes números para variables de Bernoulli afirma que si $X_1, X_2, \ldots$ son independientes y $\mathrm{B}(1; p)$, entones
$$\frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n} \longrightarrow p, \quad \text{casi seguramente.}$$
- Al realizar experimentos sucesivos e independientes, el porcentaje de veces que acaece un suceso de probabilidad p será aproximadamente el $p \times 100\,\%$. De esta forma se recupera la definición frecuentista de la probabilidad.

Convergencia en distribución

Sea $X, X_1, X_2, \ldots$ variables aleatorias y sean F y F_n las funciones de distribución de X y X_n, respectivamente. Decimos que X_n **converge a X en distribución**, y escribimos $X_n \xrightarrow{\mathcal{D}} X$, si

$$F_n(x) \to F(x), \quad \text{para todo } x \in C_F, \quad n \to \infty,$$

donde C_F es el conjunto de puntos de continuidad de F.

Idea: Si $X_n \xrightarrow{\mathcal{D}} X$, la probabilidad de $\{X_n \le a\}$ (a constante) es parecida a la probabilidad de $\{X \le a\}$ (cuando n es grande).

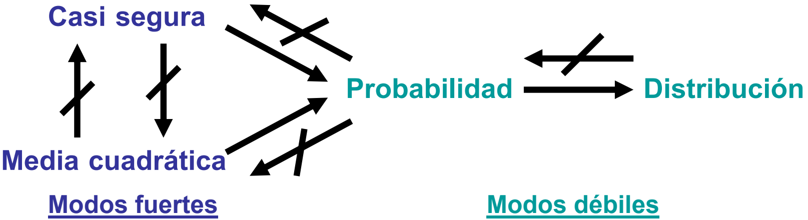
Pregunta: ¿Por qué no $\forall x \in \mathbb{R}$ en la definición anterior?

Observación: Si $X_n = 1/n$ cte. y $X = 0$ cte. Es razonable esperar que $X_n \to X$ en cualquier modo de convergencia.

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 & \text{si } x \ge 0, \end{cases} \quad \text{y} \quad F_n(x) \to \begin{cases} 0 & \text{si } x \le 0, \\ 1 & \text{si } x > 0, \end{cases} \quad \textbf{no es f.d.}$$

Nota: Las convergencias casi segura y en media cuadrática se llaman **modos de convergencia fuertes** (sobre todo la convergencia c.s.). Las convergencias en probabilidad y en distribución se denominan **modos de convergencia débiles** (sobre todo la convergencia en distribución).

Esquema de las relaciones mutuas



Nota: Los recíprocos de las anteriores implicaciones no son ciertos en general.

Idea: *Una variable que es el resultado de la suma de muchos efectos independientes entre sí sin que ninguno domine al total es aproximadamente normal.*

El Teorema Central del Límite (TCL)

Sean $X_1, X_2, \ldots,$ variables independientes y con la misma distribución de probabilidad tal que $\mathrm{E}X_i = \mu$, $\mathrm{Var}X_i = \sigma^2 > 0$ $(i \geq 1)$. Entonces, si $S_n = X_1 + \cdots + X_n$, se tiene

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} Z \sim \mathrm{N}(0;1) \quad (n \uparrow \infty).$$

Nota: El TCL es uno de los resultados clásicos más importantes de la Teoría de la probabilidad.

Observaciones importantes sobre el TCL

- Cuando n es grande, el TCL asegura que

$$X_1 + \cdots + X_n \approx \mathrm{N}\left(n\mu; \sigma\sqrt{n}\right) \quad \text{(en distribución).}$$

- En general, n se considera "grande" a partir de 30, $n \geq 30$. Sin embargo, esta regla genérica para una "buena aproximación" se debe tomar con cautela ya que depende de la distribución de partida (como veremos enseguida).
- **Importante:** No se impone ninguna *hipótesis distribucional* sobre las X_i salvo que tengan segundo momento finito. **El resultado es válido para cualquier distribución** (discreta o continua) verificando estas condiciones.
- Para calcular probabilidades de *cualquier* suma (grande) de v.a. i.i.d., es suficiente aproximar la suma por una variable normal y utilizar las tablas de probabilidad de la normal.

El Teorema Central del Límite (TCL)

Sean $X_1, X_2, \dots,$ i.i.d. tal que $\mathrm{E}X_i = \mu$, $\mathrm{Var}X_i = \sigma^2 > 0$ $(i \geq 1)$. Entonces, si $S_n = X_1 + \cdots + X_n$, se tiene

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} Z \sim \mathrm{N}(0;1) \quad (n \uparrow \infty).$$

Idea de la demostración del TCL

Para demostrar el TCL (y otras versiones más generales) se utilizan f.c. El siguiente resultado es clave.

Teorema de continuidad de Lévy-Cramer

Sean $X_1, X_2, \dots$ v.a. con f.d. $F_1, F_2, \dots$ y f.c. $\varphi_1, \varphi_2, \dots$.

Supongamos que

(a) Existe $\lim_{n\to\infty} \varphi_n(t) = \varphi(t)$.

(b) La función $\varphi(t)$ así definida es continua en $t = 0$.

Entonces, φ es la f.c. de alguna v.a. X y $X_n \xrightarrow{\mathcal{D}} X$.

Ejemplo: La demanda diaria de un producto tiene media 30 y desviación típica 6. Supuesta la independencia de la demanda de cada día respecto de los restantes:

(a) ¿Cuál es la distribución (aproximada) de la demanda en un periodo de 182 días?

(b) ¿Cuál es la probabilidad de que en 182 días, el número de unidades demandadas supere 6370 unidades?

Observaciones importantes:

- La variable aleatoria *demanda* es una variable discreta que toma los valores $0, 1, 2, \ldots$ (Muy distinta de la normal.)
- ¡¡¡¡¡Ni siquiera conocemos la distribución de probabilidad de la variable demanda!!!!! (Sólo sabemos su media y desviación.)

El **Teorema de De Moivre-Laplace** establece una aproximación normal de la distribución binomial. Es un caso particular del TCL. Históricamente, primero se mostró este resultado antes de obtener el TCL en su versión general.

El Teorema de De Moivre-Laplace

Sean $X_1, \ldots, X_n, \ldots$ v.a. independientes con la misma distribución de probabilidad de Bernoulli $\mathrm{B}(1; p)$. Entonces:

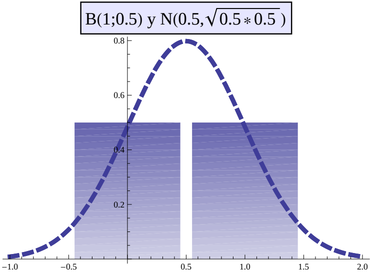
$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{D}} Z \sim \mathrm{N}(0; 1) \quad (n \uparrow \infty),$$

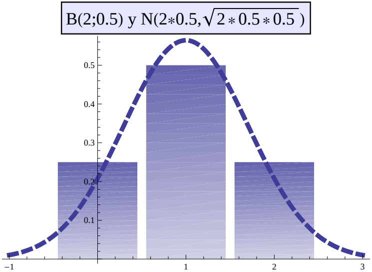
donde $S_n = X_1 + \cdots + X_n \sim \mathrm{B}(n; p)$ (Binomial).

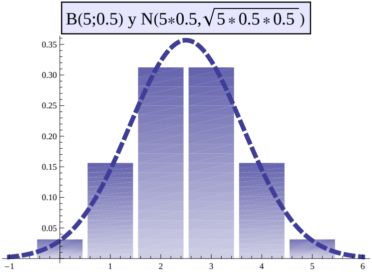
Observaciones:

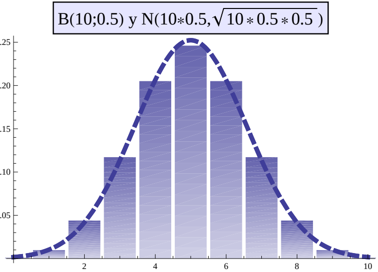
- Cuando n es grande, $\mathrm{B}(n; p) \approx \mathrm{N}\left(np; \sqrt{np(1-p)}\right)$ (en distribución).
- Cuando n es grande, podemos calcular probabilidades binomiales mediante la aproximación a la normal.

El Teorema de De Moivre-Laplace

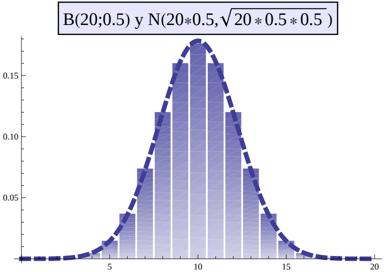


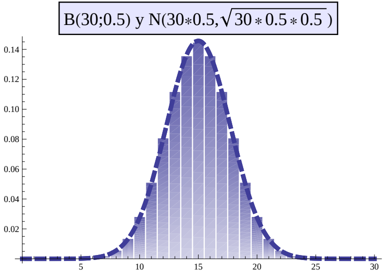


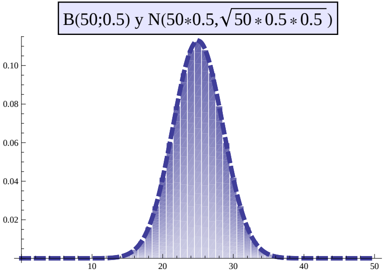


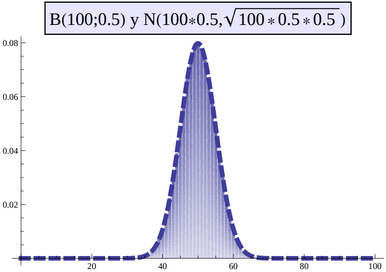


El Teorema de De Moivre-Laplace

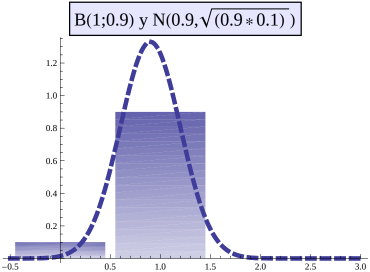


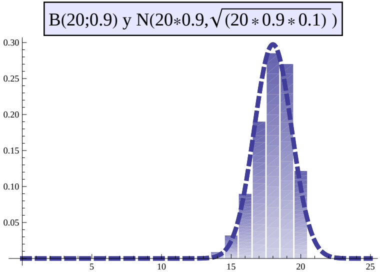


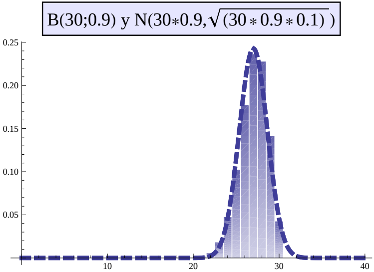


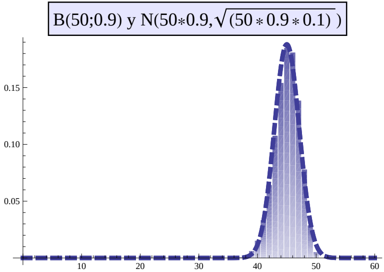


El Teorema de De Moivre-Laplace

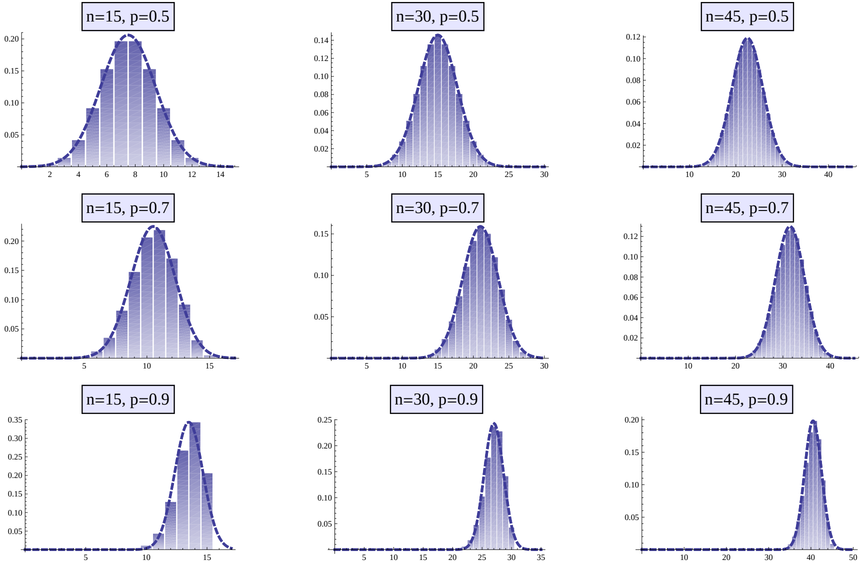








El Teorema de De Moivre-Laplace



Aplicaciones del TCL

(1) **Caso general**: En general, n se considera *grande* a partir de 30. Es decir, si $n \geq 30$ y $X_1, \ldots, X_n$ son v.a. independientes y con igual distribución con media μ y desviación típica $\sigma > 0$, entonces

$$X_1 + \cdots + X_n \approx \mathrm{N}\left(n\mu; \sigma\sqrt{n}\right).$$

(2) **Caso binomial**: Si $S_n \sim \mathrm{B}(n; p)$ y queremos calcular $\mathrm{P}(S_n \leq k)$ $(k \in \{0, \ldots, n\})$.

(2.1) Si n y p están en las tablas o la probabilidad es fácil de calcular computacionalmente, se calcula directamente.

(2.2) Si n y p *no* están en las tablas o su cálculo es delicado:

- Si $n \geq 30$, $\pi \leq 0,1$ y $n\pi < 18$ (con π : $\min\{p, 1-p\}$), entonces se utiliza la aproximación de Poisson

$$S_n \approx \mathrm{P}(np).$$

- Si $n \geq 30$ y $0,1 < p < 0,9$, se utiliza el TCL

$$S_n \approx \mathrm{N}\left(np; \sqrt{np(1-p)}\right).$$

Aplicaciones del TCL: Caso binomial

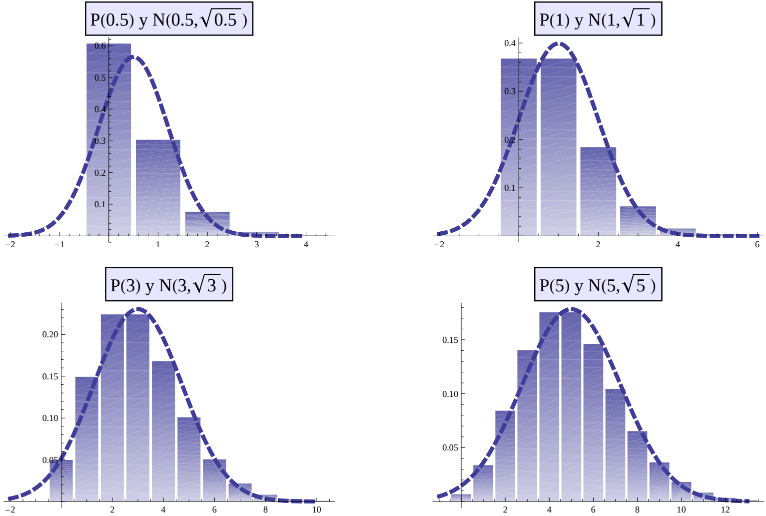
Ejemplo: Conocemos que el 70 % de los usuarios del metro utilizan bonometro.

(a) Si cogemos una muestra de 15 viajeros, ¿cuál es la probabilidad de que al menos 8 de ellos tengan este tipo de billete?

(b) Si tomamos una muestra de 10 personas en 15 estaciones, ¿cuál es la probabilidad de que sean menos de 95 los viajeros que no disponen de bonometro?

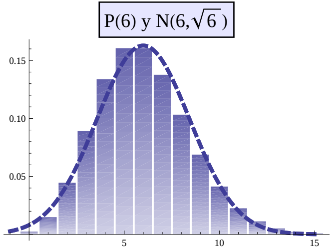
Aplicaciones del TCL: Distribución de Poisson

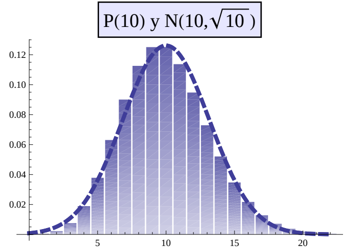
Si $Y \sim \mathrm{P}(\lambda)$ con $\lambda > 5$, entonces $Y \sim \mathrm{P}(\lambda) \approx \mathrm{N}\big(\lambda; \sqrt{\lambda}\big)$.

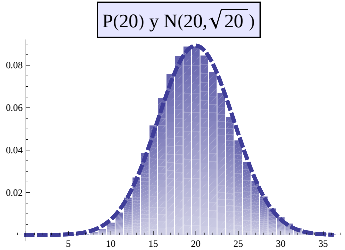


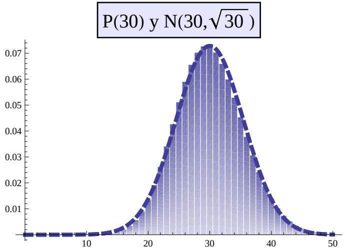
Aplicaciones del TCL: Distribución de Poisson

Si $Y \sim \mathrm{P}(\lambda)$ con $\lambda > 5$, entonces $Y \sim \mathrm{P}(\lambda) \approx \mathrm{N}\left(\lambda; \sqrt{\lambda}\right)$.





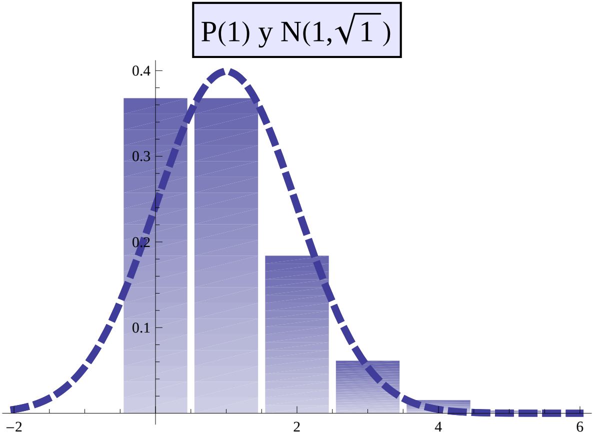




Aplicaciones del TCL: Distribución de Poisson

Pregunta: ¿Por qué $X_1 \sim \mathrm{P}(1) \nsim \mathrm{N}(1;1)$?

$$X_1 \sim Y_1 + \cdots + Y_{100}, \text{ con } Y_i \sim \mathrm{P}(1/100) \text{ ind.}$$



Aplicaciones del TCL: Distribución de Poisson

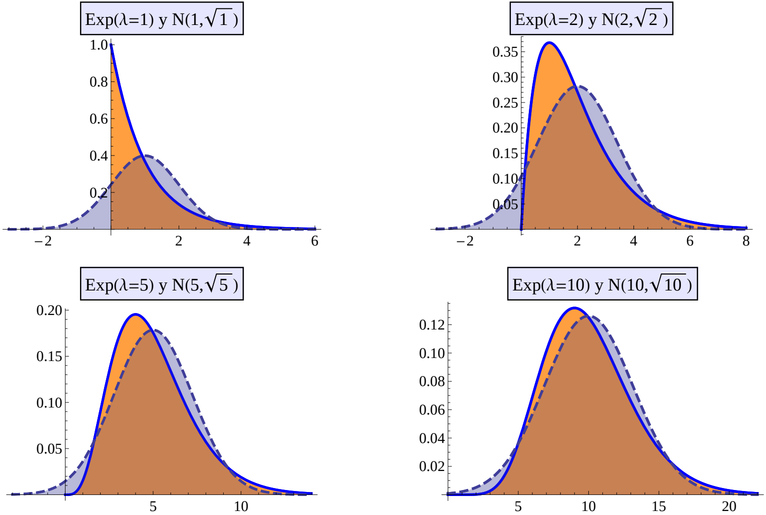
Ejemplo: Un supermercado tiene tres puertas de entrada. Se supone que el número de personas que acuden diariamente por cada una de las puestas son independientes y siguen una distribución de Poisson de medias 200, 150 y 50, respectivamente.

(a) ¿Cuál es la distribución del número total de personas que entra diariamente?

(b) ¿Cuál es la probabilidad de que en 200 días la afluencia de personas supere la cifra de 40260?

Aplicaciones del TCL: Distribución exponencial

Si $X_1, X_2 \cdots \sim$ Exp(1) independientes, $S_n \sim \mathrm{N}(n; \sqrt{n})$ (n grande).



Aplicaciones del TCL: Distribución exponencial

Si $X_1, X_2 \cdots \sim \text{Exp}(1)$ independientes, $S_n \sim \text{N}(n; \sqrt{n})$ (n grande).