Manual Básico de Estadística

Alfredo Sánchez Alberca (asalber@ceu.es)

Sep 2017

Departamento de Matemática Aplicada y Estadística
CEU San Pablo

Términos de la licencia

Contenidos

Teoría de la Probabilidad

Introducción

La estadística descriptiva permite describir el comportamiento y las relaciones entre las variables en la muestra, pero no permite sacar conclusiones sobre el resto de la población.

Ha llegado el momento de dar el salto de la muestra a la población y pasar de la estadística descriptiva a la inferencia estadística, y el puente que lo permite es la **teoría de la probabilidad**.

Hay que tener en cuenta que el conocimiento que se puede obtener de la población a partir de la muestra es limitado, y que para obtener conclusiones válidas para la población la muestra debe ser representativa de esta. Por esta razón, para garantizar la representatividad de la muestra, esta debe extraerse *aleatoriamente*, es decir, al *azar*.

La teoría de la probabilidad precisamente se encarga de controlar ese azar para saber hasta qué punto son fiables las conclusiones obtenidas a partir de una muestra.

El estudio de una característica en una población se realiza a través de experimentos aleatorios.

Definición (Experimento aleatorio)

Un *experimento aleatorio* es un experimento que cumple dos condiciones:

1. El conjunto de posibles resultados es conocido.
2. No se puede predecir con absoluta certeza el resultado del experimento.

Ejemplo. Un ejemplo típico de experimentos aleatorios son los juegos de azar. El lanzamiento de un dado, por ejemplo, es un experimento aleatorio ya que:

- Se conoce el conjunto posibles de resultados $\{1, 2, 3, 4, 5, 6\}$.
- Antes de lanzar el dado, es imposible predecir con absoluta certeza el valor que saldrá.

Otro ejemplo de experimento aleatorio sería la selección de un individuo de una población al azar y la determinación de su grupo sanguíneo.

En general, la obtención de cualquier muestra mediante procedimientos aleatorios será un

Definición (Espacio muestral)

Al conjunto Ω de todos los posibles resultados de un experimento aleatorio se le llama *espacio muestral*.

Algunos ejemplos de espacios muestrales son:

- Lanzamiento de una moneda: $\Omega = \{c, x\}$.
- Lanzamiento de un dado: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Grupo sanguíneo de un individuo seleccionado al azar: $\Omega = \{\text{A}, \text{B}, \text{AB}, 0\}$.
- Estatura de un individuo seleccionado al azar: $\Omega = \mathbb{R}^+$.

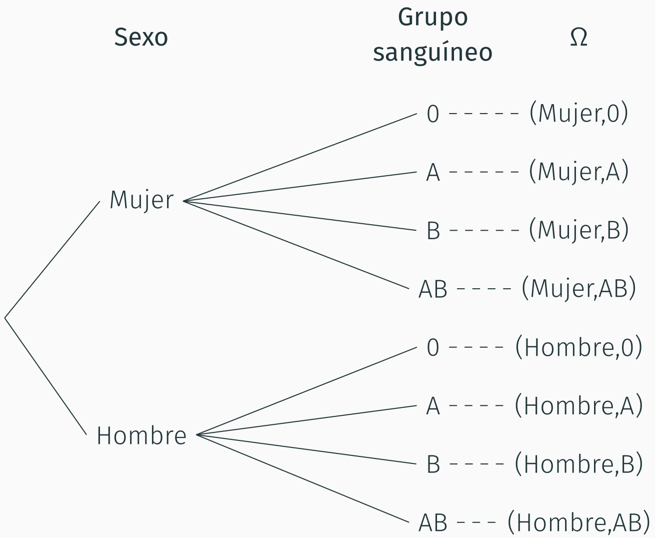
Diagrama de árbol

En experimentos donde se mide más de una variable, la determinación del espacio muestral puede resultar compleja. En tales casos es recomendable utilizar un **diagrama de árbol** para construir el espacio muestral.

En un diagrama de árbol cada variable se representa en un nivel del árbol y cada posible valor de la variable como una rama.

Diagrama de árbol

Ejemplo de sexo y grupo sanguíneo



Definición (Suceso aleatorio)

Un *suceso aleatorio* es cualquier subconjunto del espacio muestral Ω de un experimento aleatorio.

Existen distintos tipos de sucesos:

Suceso imposible: Es el suceso vacío $\emptyset$. Este suceso nunca ocurre.

Sucesos elementales: Son los sucesos formados por un solo elemento.

Sucesos compuestos: Son los sucesos formados por dos o más elementos.

Suceso seguro: Es el suceso que contiene el propio espacio muestral Ω. Este suceso siempre ocurre.

Definición (Espacio de sucesos)

Dado un espacio muestral Ω de un experimento aleatorio, el conjunto formado por todos los posibles sucesos de Ω se llama *espacio de sucesos de* Ω y se denota $\mathcal{P}(\Omega)$.

Ejemplo. Dado el espacio muestral $\Omega = \{a, b, c\}$, su espacio de sucesos es

$$\mathcal{P}(\Omega) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$$

Operaciones entre sucesos

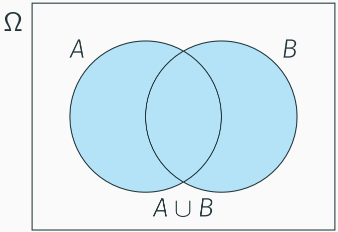
Puesto que los sucesos son conjuntos, por medio de la teoría de conjuntos se pueden definir las siguientes operaciones entre sucesos:

- Unión.
- Intersección.
- Complementario.
- Diferencia.

Definición (Suceso unión)

Dados dos sucesos $A, B \subseteq \Omega$, se llama *suceso unión* de A y B, y se denota $A \cup B$, al suceso formado por los elementos de A junto a los elementos de B, es decir,

$$A \cup B = \{x \,|\, x \in A \text{ o } x \in B\}.$$

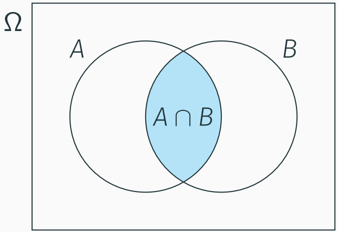


El suceso unión $A \cup B$ ocurre siempre que ocurre A o B.

Definición (Suceso intersección)

Dados dos sucesos $A, B \subseteq \Omega$, se llama *suceso intersección* de A y B, y se denota $A \cap B$, al suceso formado por los elementos comunes de A y B, es decir,

$$A \cap B = \{x \,|\, x \in A \text{ y } x \in B\}.$$



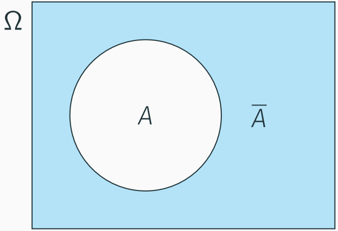
El suceso intersección $A \cap B$ ocurre siempre que ocurren A y B.

Diremos que dos sucesos son **incompatibles** si su intersección es vacía.

Definición (Suceso contrario)

Dado suceso $A \subseteq \Omega$, se llama *suceso contrario o complementario* de A, y se denota $\overline{A}$, al suceso formado por los elementos de Ω que no pertenecen a A, es decir,

$$\overline{A} = \{x \,|\, x \notin A\}.$$

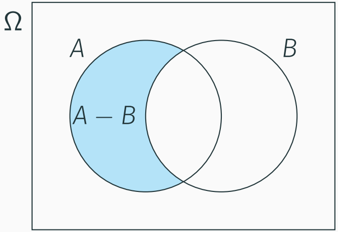


El suceso contrario $\overline{A}$ ocurre siempre que no ocurre A.

Definición (Suceso diferencia)

Dados dos sucesos $A, B \subseteq \Omega$, se llama *suceso diferencia* de A y B, y se denota $A - B$, al suceso formado por los elementos de A que no pertenecen a B, es decir,

$$A - B = \{x \,|\, x \in A \text{ y } x \notin B\} = A \cap \overline{B}.$$



El suceso diferencia $A - B$ ocurre siempre que ocurre A pero no ocurre B, y también puede expresarse como $A \cap \overline{B}$.

Dado el espacio muestral correspondiente al lanzamiento de un dado $\Omega = \{1, 2, 3, 4, 5, 6\}$ y los sucesos $A = \{2, 4, 6\}$ y $B = \{1, 2, 3, 4\}$,

- La unión de A y B es $A \cup B = \{1, 2, 3, 4, 6\}$.
- La intersección de A y B es $A \cap B = \{2, 4\}$.
- El contrario de A es $\overline{A} = \{1, 3, 5\}$.
- Los eventos A y $\overline{A}$ son incompatibles.
- La diferencia de A y B es $A - B = \{6\}$, y la diferencia de B y A es $B - A = \{1, 3\}$.

Álgebra de sucesos

Dados los sucesos $A, B, C \in \mathcal{P}(\Omega)$, se cumplen las siguientes propiedades:

1. $A \cup A = A$, $A \cap A = A$ (idempotencia).
2. $A \cup B = B \cup A$, $A \cap B = B \cap A$ (conmutativa).
3. $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$ (asociativa).
4. $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$ (distributiva).
5. $A \cup \emptyset = A$, $A \cap E = A$ (elemento neutro).
6. $A \cup E = E$, $A \cap \emptyset = \emptyset$ (elemento absorbente).
7. $A \cup \overline{A} = E$, $A \cap \overline{A} = \emptyset$ (elemento simétrico complementario).
8. $\overline{\overline{A}} = A$ (doble contrario).
9. $\overline{A \cup B} = \overline{A} \cap \overline{B}$, $\overline{A \cap B} = \overline{A} \cup \overline{B}$ (leyes de Morgan).
10. $A \cap B \subseteq A \cup B$.

Definición clásica de probabilidad

Definición (Probabilidad — Laplace)

Dado un espacio muestral Ω de un experimento aleatorio donde todos los elementos de Ω son equiprobables, la *probabilidad* de un suceso $A \subseteq \Omega$ es el cociente entre el número de elementos de A y el número de elementos de Ω

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{nº casos favorables a A}}{\text{nº casos posibles}}$$

Esta definición es ampliamente utilizada, aunque tiene importantes restricciones:

- Es necesario que todos los elementos del espacio muestral tengan la misma probabilidad de ocurrir (*equiprobabilidad*).
- No puede utilizarse con espacios muestrales infinitos, o de los que no se conoce el número de casos posibles.

¡Ojo! Esto no se cumple en muchos experimentos aleatorios reales.

Teorema (Ley de los grandes números)

Cuando un experimento aleatorio se repite un gran número de veces, las frecuencias relativas de los sucesos del experimento tienden a estabilizarse en torno a cierto número, que es precisamente su probabilidad.

De acuerdo al teorema anterior, podemos dar la siguiente definición

Definición (Probabilidad frecuentista)

Dado un espacio muestral Ω de un experimento aleatorio reproducible, la *probabilidad* de un suceso $A \subseteq \Omega$ es la frecuencia relativa del suceso A en infinitas repeticiones del experimento

$$P(A) = lim_{n\to\infty}\frac{n_A}{n}$$

Aunque esta definición es muy útil en experimentos científicos reproducibles, también tiene serios inconvenientes, ya que

- Sólo se calcula una aproximación de la probabilidad real.

Definición (Kolmogórov)

Dado un espacio muestral Ω de un experimento aleatorio, una función de *probabilidad* es una aplicación que asocia a cada suceso $A \subseteq \Omega$ un número real $P(A)$, conocido como probabilidad de A, que cumple los siguientes axiomas:

1. La probabilidad de un suceso cualquiera es positiva o nula,
$$P(A) \geq 0.$$
2. La probabilidad del suceso seguro es igual a la unidad,
$$P(\Omega) = 1.$$
3. La probabilidad de la unión de dos sucesos incompatibles ($A \cap B = \emptyset$) es igual a la suma de las probabilidades de cada uno de ellos,
$$P(A \cup B) = P(A) + P(B).$$

Consecuencias de los axiomas de probabilidad

A partir de los axiomas de la definición de probabilidad se pueden deducir los siguientes resultados:

Consecuencias de los axiomas de probabilidad

A partir de los axiomas de la definición de probabilidad se pueden deducir los siguientes resultados:

1. $P(\overline{A}) = 1 - P(A)$.

Consecuencias de los axiomas de probabilidad

A partir de los axiomas de la definición de probabilidad se pueden deducir los siguientes resultados:

1. $P(\overline{A}) = 1 - P(A)$.
2. $P(\emptyset) = 0$.

Consecuencias de los axiomas de probabilidad

A partir de los axiomas de la definición de probabilidad se pueden deducir los siguientes resultados:

1. $P(\overline{A}) = 1 - P(A)$.
2. $P(\emptyset) = 0$.
3. Si $A \subseteq B$ entonces $P(A) \leq P(B)$.

Consecuencias de los axiomas de probabilidad

A partir de los axiomas de la definición de probabilidad se pueden deducir los siguientes resultados:

1. $P(\overline{A}) = 1 - P(A)$.
2. $P(\emptyset) = 0$.
3. Si $A \subseteq B$ entonces $P(A) \leq P(B)$.
4. $P(A) \leq 1$.

Consecuencias de los axiomas de probabilidad

A partir de los axiomas de la definición de probabilidad se pueden deducir los siguientes resultados:

1. $P(\overline{A}) = 1 - P(A)$.
2. $P(\emptyset) = 0$.
3. Si $A \subseteq B$ entonces $P(A) \leq P(B)$.
4. $P(A) \leq 1$.
5. Si A y B son sucesos compatibles, es decir, su intersección no es vacía, entonces

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

Consecuencias de los axiomas de probabilidad

A partir de los axiomas de la definición de probabilidad se pueden deducir los siguientes resultados:

1. $P(\overline{A}) = 1 - P(A)$.
2. $P(\emptyset) = 0$.
3. Si $A \subseteq B$ entonces $P(A) \leq P(B)$.
4. $P(A) \leq 1$.
5. Si A y B son sucesos compatibles, es decir, su intersección no es vacía, entonces

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

6. Si el suceso A está compuesto por los sucesos elementales $e_1, e_2, ..., e_n$, entonces

$$P(A) = \sum_{i=1}^{n} P(e_i).$$

Interpretación de la probabilidad

Como ha quedado claro en los axiomas anteriores, la probabilidad de un evento A es un número real $P(A)$ que está siempre entre 0 y 1.

En cierto modo, este número expresa la verosimilitud del evento, es decir, la confianza que hay en que ocurra A en el experimento. Por tanto, también nos da una medida de la incertidumbre sobre el suceso.

- La mayor incertidumbre corresponde a $P(A) = 0.5$ (Es tan probable que ocurra A como que no ocurra).
- La menor incertidumbre corresponde a $P(A) = 1$ (A sucederá con absoluta certeza) y $P(A) = 0$ (A no sucederá con absoluta certeza).

Cuando $P(A)$ está más próximo a 0 que a 1, la confianza en que no ocurra A es mayor que la de que ocurra A. Por el contrario, cuando $P(A)$ está más próximo a 1 que a 0, la confianza en que ocurra A es mayor que la de que no ocurra A.

Experimentos condicionados

En algunas ocasiones, es posible que tengamos alguna información sobre el experimento antes de su realización. Habitualmente esa información se da en forma de un suceso B del mismo espacio muestral que sabemos que es cierto antes de realizar el experimento.

En tal caso se dice que el suceso B es un suceso *condicionante*, y la probabilidad de otro suceso A se conoce como **probabilidad condicionada** y se expresa

$$P(A|B).$$

Esto debe leerse como *probabilidad de A dado B* o *probabilidad de A bajo la condición de B.*

Experimentos condicionados

Los condicionantes suelen cambiar el espacio muestral del experimento y por tanto las probabilidades de sus sucesos.

Supongamos que tenemos una muestra de 100 hombres y 100 mujeres con las siguientes frecuencias

	No fumadores	Fumadores
Mujeres	80	20
Hombres	60	40

Entonces, usando la definición frecuentista de probabilidad, la probabilidad de que una persona elegida al azar sea fumadora es

$$P(\text{Fumadora}) = \frac{60}{200} = 0.3.$$

Experimentos condicionados

Los condicionantes suelen cambiar el espacio muestral del experimento y por tanto las probabilidades de sus sucesos.

Supongamos que tenemos una muestra de 100 hombres y 100 mujeres con las siguientes frecuencias

	No fumadores	Fumadores
Mujeres	80	20
Hombres	60	40

Entonces, usando la definición frecuentista de probabilidad, la probabilidad de que una persona elegida al azar sea fumadora es

$$P(\text{Fumadora}) = \frac{60}{200} = 0.3.$$

Sin embargo, si se sabe que la persona elegida es mujer, entonces la muestra se reduce a la primera fila, y la probabilidad de ser fumadora es

$$P(\text{Fumadora}|\text{Mujer}) = \frac{20}{100} = 0.2.$$

Definición (Probabilidad condicionada)

Dado un espacio muestral Ω de un experimento aleatorio, y dos dos sucesos $A, B \subseteq \Omega$, la probabilidad de A *condicionada* por B es

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

siempre y cuando, $P(B) \neq 0$.

Esta definición permite calcular probabilidades sin tener que alterar el espacio muestral original del experimento.

Ejemplo. En el ejemplo anterior

$$P(\text{Fumadora}|\text{Mujer}) = \frac{P(\text{Fumadora} \cap \text{Mujer})}{P(\text{Mujer})} = \frac{20/200}{100/200} = \frac{20}{100} = 0.2.$$

Probabilidad del suceso intersección

A partir de la definición de probabilidad condicionada es posible obtener la fórmula para calcular la probabilidad de la intersección de dos sucesos.

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

Ejemplo. En una población hay un 30% de fumadores y se sabe que el 40% de los fumadores tiene cáncer de pulmón. La probabilidad de que una persona elegida al azar sea fumadora y tenga cáncer de pulmón es

$$P(\text{Fumadora} \cap \text{Cáncer}) = P(\text{Fumadora})P(\text{Cáncer}|\text{Fumadora}) = 0.3 \times 0.4 = 0.12.$$

En ocasiones, la ocurrencia del suceso condicionante no cambia la probabilidad original del suceso principal.

Definición (Sucesos independientes)

Dado un espacio muestral Ω de un experimento aleatorio, dos sucesos $A, B \subseteq \Omega$ son *independientes* si la probabilidad de A no se ve alterada al condicionar por B, y viceversa, es decir,

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B),$$

si $P(A) \neq 0$ y $P(B) \neq 0$.

Esto significa que la ocurrencia de uno evento no aporta información relevante para cambiar la incertidumbre sobre el otro.

Cuando dos eventos son independientes, la probabilidad de su intersección es igual al producto de sus probabilidades,

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

Definición (Espacio probabilístico)

Un *espacio probabilístico* de un experimento aleatorio es una terna $(\Omega, \mathcal{F}, P)$ donde

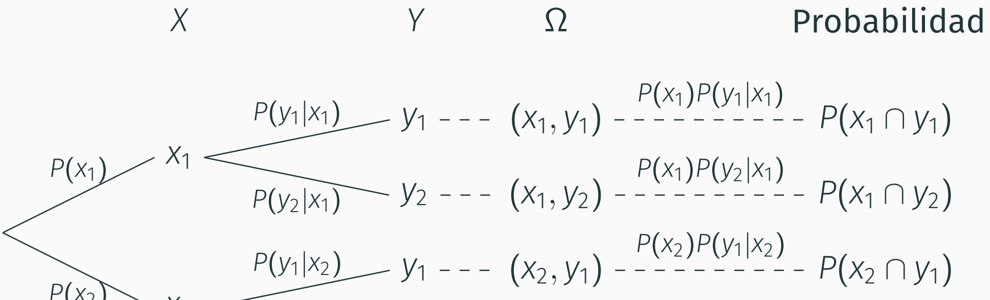
- Ω es el espacio muestral del experimento.
- $\mathcal{F}$ es un un conjunto de sucesos del experimento.
- P es una función de probabilidad.

Si conocemos la probabilidad de todos los elementos de Ω, entonces podemos calcular la proabilidad de cualquier suceso en $\mathcal{F}$ y se puede construir fácilmente el espacio probabilístico.

Construcción del espacio probabilístico

Para determinar la probabilidad de cada suceso elemental se puede utilizar un diagrama de árbol, mediante las siguientes reglas:

1. Para cada nodo del árbol, etiquetar la rama que conduce hasta él con la probabilidad de que la variable en ese nivel tome el valor del nodo, condicionada por los sucesos correspondientes a sus nodos antecesores en el árbol.
2. La probabilidad de cada suceso elemental en las hojas del árbol es el producto de las probabilidades de las ramas que van desde la raíz a la hoja del árbol.

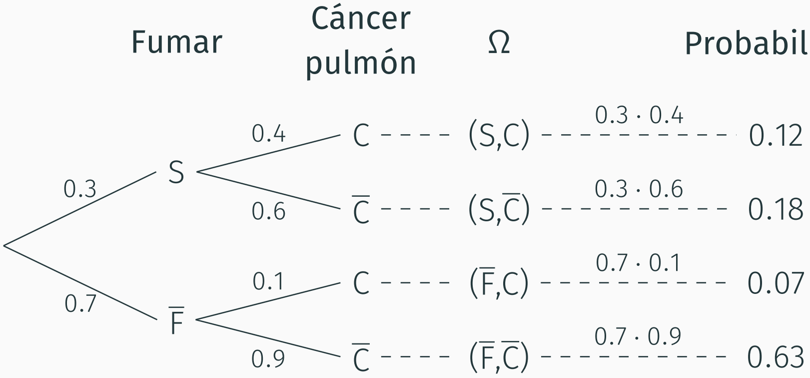


Árboles de probabilidad con variables dependientes

Ejemplo de dependencia del cáncer con respecto al tabaco

Sea una población en la que el 30% de las personas fuman, y que la incidencia del cáncer de pulmón en fumadores es del 40% mientras que en los no fumadores es del 10%.

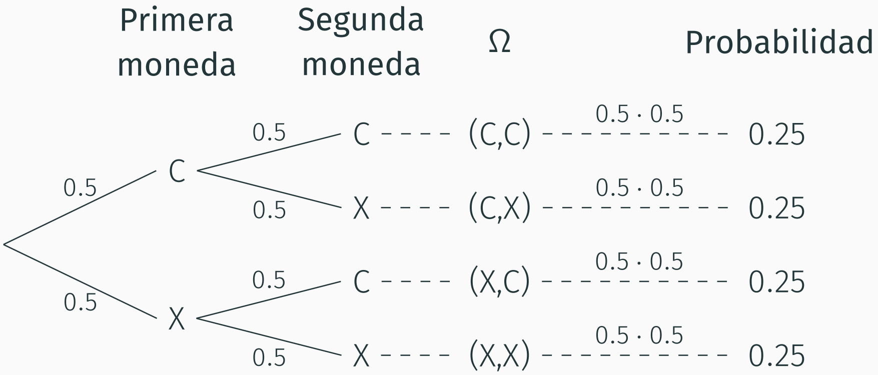
El espacio probabilístico del experimento aleatorio que consiste en elegir una persona al azar y medir las variables Fumar y Cáncer de pulmón se muestra a continuación.



Árboles de probabilidad con variables independientes

Ejemplo de independencia en el lanzamiento de dos monedas

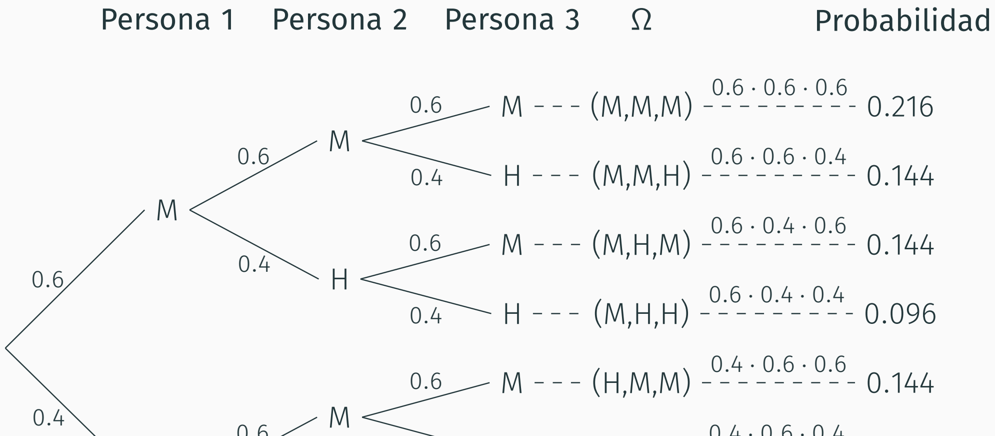
El árbol de probabilidad asociado al experimento aleatorio que consiste en el lanzamiento de dos monedas se muestra a continuación.



Árboles de probabilidad con variables independientes

Ejemplo de independencia en la elección de una muestra aleatoria de tamaño 3

Dada una población en la que hay un 40% de hombres y un 60% de mujeres, el experimento aleatorio que consiste en tomar una muestra aleatoria de tres personas tiene el árbol de probabilidad que se muestra a continuación.



Sistema completo de sucesos

Definición (Sistema completo de sucesos)

Una colección de sucesos $A_1, A_2, \ldots, A_n$ de un mismo espacio muestral Ω es un *sistema completo* si cumple las siguientes condiciones:

1. La unión de todos es el espacio muestral: $A_1 \cup \cdots \cup A_n = \Omega$.
2. Son incompatibles dos a dos: $A_i \cap A_j = \emptyset \ \forall i \neq j$.

En realidad un sistema completo de sucesos es una partición del espacio muestral de acuerdo a algún atributo, como por ejemplo el sexo o el grupo sanguíneo.

Conocer las probabilidades de un determinado suceso en cada una de las partes de un sistema completo puede ser útil para calcular su probabilidad.

Teorema (Probabilidad total)

Dado un sistema completo de sucesos $A_1, \ldots, A_n$ y un suceso B de un espacio muestral Ω, la probabilidad de cualquier suceso B del espacio muestral se puede calcular mediante la fórmula

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

La demostración del teorema es sencilla, ya que al ser $A_1, \ldots, A_n$ un sistema completo tenemos

$$B = B \cap E = B \cap (A_1 \cup \cdots \cup A_n) = (B \cap A_1) \cup \cdots \cup (B \cap A_n)$$

y como estos sucesos son incompatibles entre sí, se tiene

$$P(B) = P((B \cap A_1) \cup \cdots \cup (B \cap A_n)) = P(B \cap A_1) + \cdots + P(B \cap A_n) =$$

$$= P(A_1)P(B/A_1) + \cdots + P(A_n)P(B/A_n) = \sum_{i=1}^{n} P(A_i)P(B/A_i).$$

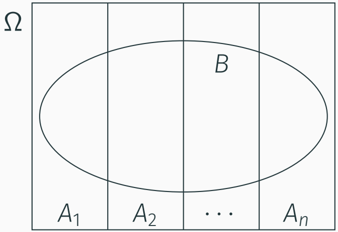
The proof of the theorem is quite simple. As $A_1, \ldots, A_n$ is a partition of Ω, we have

$$B = B \cap \Omega = B \cap (A_1 \cup \cdots \cup A_n) = (B \cap A_1) \cup \cdots \cup (B \cap A_n).$$

And all the events of this union are mutually incompatible as $A_1, \ldots, A_n$ are, thus

$$P(B) = P((B \cap A_1) \cup \cdots \cup (B \cap A_n)) = P(B \cap A_1) + \cdots + P(B \cap A_n) =$$

$$= P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n) = \sum_{i=1}^{n} P(A_i)P(B|A_i).$$



Total probability theorem
Proof

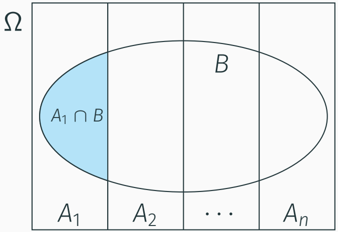
The proof of the theorem is quite simple. As $A_1, \ldots, A_n$ is a partition of Ω, we have

$$B = B \cap \Omega = B \cap (A_1 \cup \cdots \cup A_n) = (B \cap A_1) \cup \cdots \cup (B \cap A_n).$$

And all the events of this union are mutually incompatible as $A_1, \ldots, A_n$ are, thus

$$P(B) = P((B \cap A_1) \cup \cdots \cup (B \cap A_n)) = P(B \cap A_1) + \cdots + P(B \cap A_n) =$$

$$= P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n) = \sum_{i=1}^{n} P(A_i)P(B|A_i).$$



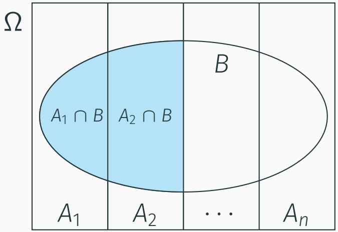
The proof of the theorem is quite simple. As $A_1, \ldots, A_n$ is a partition of Ω, we have

$$B = B \cap \Omega = B \cap (A_1 \cup \cdots \cup A_n) = (B \cap A_1) \cup \cdots \cup (B \cap A_n).$$

And all the events of this union are mutually incompatible as $A_1, \ldots, A_n$ are, thus

$$P(B) = P((B \cap A_1) \cup \cdots \cup (B \cap A_n)) = P(B \cap A_1) + \cdots + P(B \cap A_n) =$$

$$= P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n) = \sum_{i=1}^{n} P(A_i)P(B|A_i).$$



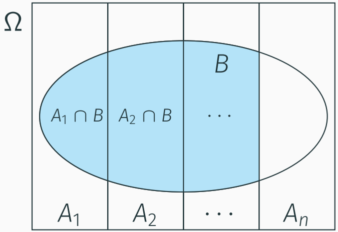
The proof of the theorem is quite simple. As $A_1, \dots, A_n$ is a partition of Ω, we have

$$B = B \cap \Omega = B \cap (A_1 \cup \dots \cup A_n) = (B \cap A_1) \cup \dots \cup (B \cap A_n).$$

And all the events of this union are mutually incompatible as $A_1, \dots, A_n$ are, thus

$$P(B) = P((B \cap A_1) \cup \dots \cup (B \cap A_n)) = P(B \cap A_1) + \dots + P(B \cap A_n) =$$

$$= P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n) = \sum_{i=1}^{n} P(A_i)P(B|A_i).$$



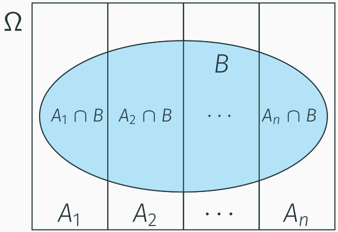
The proof of the theorem is quite simple. As $A_1, \dots, A_n$ is a partition of Ω, we have

$$B = B \cap \Omega = B \cap (A_1 \cup \cdots \cup A_n) = (B \cap A_1) \cup \cdots \cup (B \cap A_n).$$

And all the events of this union are mutually incompatible as $A_1, \dots, A_n$ are, thus

$$P(B) = P((B \cap A_1) \cup \cdots \cup (B \cap A_n)) = P(B \cap A_1) + \cdots + P(B \cap A_n) =$$

$$= P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n) = \sum_{i=1}^{n} P(A_i)P(B|A_i).$$



Un determinado síntoma S puede ser originado por una enfermedad E pero también lo pueden presentar las personas sin la enfermedad. Sabemos que la prevalencia de la enfermedad E es 0.2. Además, se sabe que el 90% de las personas con la enfermedad presentan el síntoma, mientras que sólo el 40% de las personas sin la enfermedad lo presentan.

Si se toma una persona al azar de la población, *¿qué probabilidad hay de que tenga el síntoma?*

Para responder a la pregunta se puede aplicar el teorema de la probabilidad total usando el sistema completo $\{E, \overline{E}\}$:

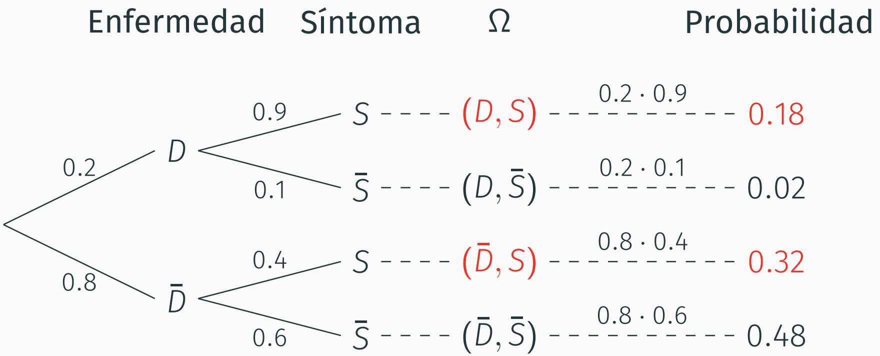
$$P(S) = P(E)P(S|E) + P(\overline{E})P(S|\overline{E}) = 0.2 \cdot 0.9 + 0.8 \cdot 0.4 = 0.5.$$

Es decir, la mitad de la población tendrá el síntoma.

¡En el fondo se trata de una media ponderada de probabilidades!

Teorema de la probabilidad total
Cálculo con el árbol de probabilidad

La respuesta a la pregunta anterior es evidente a la luz del árbol de probabilidad del espacio probabilístico del experimento.



$$P(S) = P(E, S) + P(\bar{E}, S) = P(E)P(S|E) + P(\bar{E})P(S|\bar{E})$$

Teorema de Bayes

Los sucesos de un sistema completo de sucesos $A_1, \cdots, A_n$ también pueden verse como las distintas hipótesis ante un determinado hecho B.

En estas condiciones resulta útil poder calcular las probabilidades a posteriori $P(A_i|B)$ de cada una de las hipótesis.

Teorema (Bayes)

Dado un sistema completo de sucesos $A_1, \ldots, A_n$ y un suceso B de un espacio muestral Ω y otro suceso B del mismo espacio muestral, la probabilidad de cada suceso A_i $i = 1, \ldots, n$ condicionada por B puede calcularse con la siguiente fórmula

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

En el ejemplo anterior, una pregunta más interesante es qué diagnosticar a una persona que presenta el síntoma.

En este caso se puede interpretar E y $\overline{E}$ como las dos posibles hipótesis para el síntoma S. Las probabilidades a priori para ellas son $P(E) = 0.2$ y $P(\overline{E}) = 0.8$. Esto quiere decir que si no se dispone de información sobre el síntoma, el diagnóstico será que la persona no tiene la enfermedad.

Sin embargo, si al reconocer a la persona se observa que presenta el síntoma, dicha información condiciona a las hipótesis, y para decidir entre ellas es necesario calcular sus probabilidades a posteriori, es decir,

$$P(E|S) \text{ y } P(\overline{E}|S)$$

Para calcular las probabilidades a posteriori se puede utilizar el teorema de Bayes:

$$P(E|S) = \frac{P(E)P(S|E)}{P(E)P(S|E)+P(\bar{E})P(S|\bar{E})} = \frac{0.2\cdot 0.9}{0.2\cdot 0.9+0.8\cdot 0.4} = \frac{0.18}{0.5} = 0.36,$$

$$P(\bar{E}|S) = \frac{P(\bar{E})P(S|\bar{E})}{P(E)P(S|E)+P(\bar{E})P(S|\bar{E})} = \frac{0.8\cdot 0.4}{0.2\cdot 0.9+0.8\cdot 0.4} = \frac{0.32}{0.5} = 0.64.$$

Como se puede ver la probabilidad de tener la enfermedad ha aumentado. No obstante, la probabilidad de no tener la enfermedad sigue siendo mayor que la de tenerla, y por esta razón el diagnóstico seguirá siendo que no tiene la enfermedad.

En este caso se dice que el síntoma *S no es determinante* a la hora de diagnosticar la enfermedad.

En Epidemiología es común el uso de test para diagnosticar enfermedades.

Generalmente estos test no son totalmente fiables, sino que hay cierta probabilidad de acierto o fallo en el diagnóstico, que suele representarse en la siguiente tabla:

	Presencia enfermedad E	Ausencia enfermedad $\overline{D}$
Test positivo $+$	Verdadero positivo VP	Falso positivo FP
Test negativo $-$	Falso negativo FN	Verdadero Negativo VN

La fiabilidad de un test diagnóstico depende de las siguientes probabilidades.

Definición (Sensibilidad)

La *sensibilidad* de un test diagnóstico es la proporción de resultados positivos del test en personas con la enfermedad,

$$P(+|E) = \frac{VP}{VP + FN}$$

Definición (Especificidad)

La *especificidad* de un test diagnóstico es la proporción de resultados negativos del test en personas sin la enfermedad,

$$P(-|\overline{E}) = \frac{VN}{VN + FP}$$

Interpretación de la sensibilidad y la especificidad

Normalmente existe un balance entre la sensibilidad y la especificidad.

Un test con una alta sensibilidad detectará la enfermedad en la mayoría de las personas enfermas, pero también dará más falsos positivos que un test menos sensible. De este modo, un resultado positivo en un test con una gran sensibilidad no es muy útil para confirmar la enfermedad, pero un resultado negativo es útil para descartar la enfermedad, ya que raramente da resultados negativos en personas con la enfermedad.

Por otro lado, un test con una alta especificidad descartará la enfermedad en la mayoría de las personas sin la enfermedad, pero también producirá más falsos negativos que un test menos específico. Así, un resultado negativo en un test con una gran especificidad no es útil para descartar la enfermedad, pero un resultado positivo es muy útil para confirmar la enfermedad, ya que raramente da resultados positivos en personas sin la enfermedad.

Interpretación de la sensibilidad y la especificidad

Decidir entre un test con una gran sensibilidad o un test con una gran especificidad depende del tipo de enfermedad y el objetivo del test. En general, utilizaremos un test sensible cuando:

- La enfermedad es grave y es importante detectarla.
- La enfermedad es curable.
- Los falsos positivos no provocan traumas serios.

Y utilizaremos un test específico cuando:

- La enfermedad es importante pero difícil o imposible de curar.
- Los falsos positivos pueden provocar traumas serios.
- El tratamiento de los falsos positivos puede tener graves consecuencias.

Valores predictivos de un test diagnóstico

Pero el aspecto más importante de un test diagnóstico es su poder predictivo, que se mide con las siguientes probabilidades a posteriori.

Definición (Valor predictivo positivo VPP)

El *valor predictivo positivo* de un test diagnóstico es la proporción de personas con la enfermedad entre las personas con resultado positivo en el test,

$$P(E|+) = \frac{VP}{VP + FP}$$

Definición (Valor predictivo negativo VPN)

El *valor predictivo negativo* de un test diagnóstico es la proporción de personas sin la enfermedad entre las personas con resultado negativo en el test,

$$P(\overline{E}|-) = \frac{VN}{VN + FN}$$

Interpretación de los valores predictivos

Los valores predictivos positivo y negativo permiten confirmar o descartar la enferemedad, respectivamente, si alcanzan al menos el umbral de 0.5.

$$\begin{aligned} VPP > 0.5 &\quad \Rightarrow \quad \text{Diagnosticar la enfermedad} \\ VPN > 0.5 &\quad \Rightarrow \quad \text{Diagnosticar la no enfermedad} \end{aligned}$$

No obstante, estas probabilidades dependen de la proporción de personas con la enfermedad en la población $P(E)$ que se conoce como **prevalencia** de la enfermedad. Pueden calcularse a partir de la sensibilidad y la especificidad del test diagnóstico usando el teorema de Bayes.

$$\begin{aligned} VPP = P(E|+) &= \frac{P(E)P(+|E)}{P(E)P(+|E) + P(\overline{E})P(+|\overline{E})} \\ VPN = P(\overline{E}|-) &= \frac{P(\overline{E})P(-|\overline{E})}{P(E)P(-|E) + P(\overline{E})P(-|\overline{E})} \end{aligned}$$

Así, con enfermedades frecuentes, el valor predictivo positivo aumenta, y con enfermedades raras, el valor predictivo negativo aumenta.

Un test diagnóstico para la gripe se ha aplicado a una muestra aleatoria de 1000 personas. Los resultados aparecen resumidos en la siguiente tabla.

	Presencia de gripe E	Ausencia de gripe $\overline{E}$
Test $+$	95	90
Test $-$	5	810

Según esta muestra, la prevalencia de la gripe puede estimarse como

$$P(E) = \frac{95+5}{1000} = 0.1.$$

La sensibilidad del test diagnóstico es

$$P(+|E) = \frac{95}{95+5} = 0.95.$$

Y la especificidad es

$$P(-|\overline{E}) = \frac{810}{90+810} = 0.9.$$

El valor predictivo positivo del test es

$$VPP = P(E|+) = \frac{95}{95+90} = 0.5135.$$

Como este valor es mayor que 0.5, eso significa que se diagnosticará la gripe si el resultado del test es positivo. No obstante, la confianza en el diagnóstico será baja, ya que el valor es poco mayor que 0.5.

Por otro lado, el valor predictivo negativo es

$$VPN = P(\overline{E}|-) = \frac{810}{5+810} = 0.9939.$$

Como este valor es casi 1, eso significa que es casi seguro que no se tiene la gripe cuando el resultado del test es negativo.

Así, se puede concluir que este test es muy potente para descartar la gripe, pero no lo est tanto para confirmarla.

Razón de verosimilitud de un test diagnóstico

La siguientes medidas también se derivan de la sensibilidad y la especificidad de un test diagnóstico.

Definición (Razón de verosimilitud positiva $RV+$)

La *razón de verosimilitud positiva* de un test diagnóstico es el cociente entre la probabilidad de un resultado positivo en personas con la enfermedad y personas sin la enfermedad, respectivamente.

$$RV+ = \frac{P(+|E)}{P(+|\overline{E})} = \frac{\text{Sensibilidad}}{1 - \text{Especificidad}}$$

Definición (Razón de verosimilitud negativa $RV-$)

La *razón de verosimilitud negativa* de un test diagnóstico es el cociente entre la probabilidad de un resultado negativo en personas con la enfermedad y personas sin la enfermedad, respectivamente.

$$RV- = \frac{P(-|E)}{P(-|\overline{E})} = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}}$$

Interpretación de las razones de verosimilitud

La razón de verosimilitud positiva puede interpretarse como el número de veces que un resultado positivo es más probable en personas con la enfermedad que en personas sin la enfermedad.

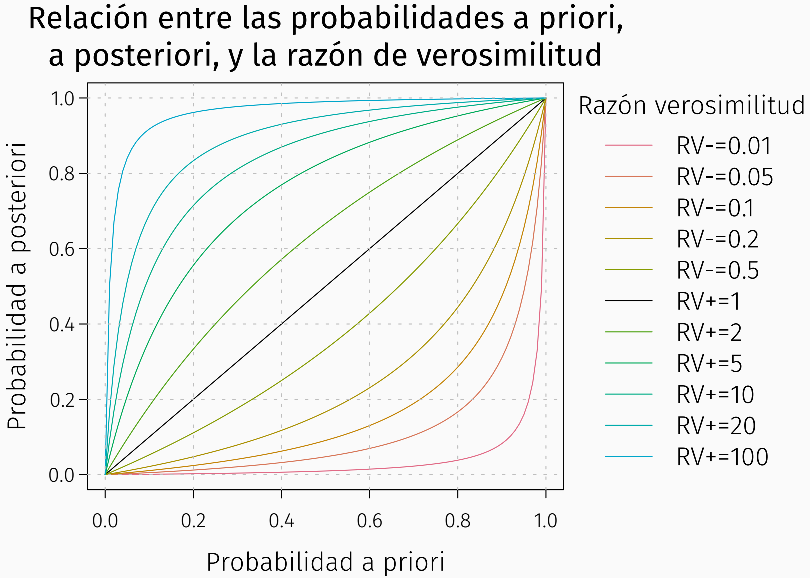
Por otro lado, la razón de verosimilitud negativa puede interpretarse como el número de veces que un resultado negativo es más probable en personas con la enfermedad que en personas sin la enfermedad.

Las probabiliades a posteriori pueden calculares a partir de las probabilidades a priori usando las razones de verosimilitud

$$P(E|+) = \frac{P(E)P(+|E)}{P(E)P(+|E) + P(\overline{E})P(+|\overline{E})} = \frac{P(E)RV+}{1 - P(E) + P(E)RV+}$$

Así,

- Una razón de verosimilitud positiva mayor que 1 aumenta la probabilidad de la enfermedad.
- Una razón de verosimilitud positiva menor que 1 disminuye la probabilidad de la


Relación entre las probabilidades a priori,
a posteriori, y la razón de verosimilitud
Probabilidad a posteriori
Probabilidad a priori
0.0
0.2
0.4
0.6
0.8
1.0
Razón verosimilitud
RV-=0.01
RV-=0.05
RV-=0.1
RV-=0.2
RV-=0.5
RV+=1
RV+=2
RV+=5
RV+=10
RV+=20
RV+=100