Capítulo 2

Probabilidad

2.1. Introducción

Recordemos que en el Tema 1 definimos **Experimento** como **cualquier procedimiento de obtención de un dato en el que mantenemos fijos ciertos factores**. De esta forma se puede hablar de repetir el experimento si volvemos a obtener un nuevo dato mientras se mantiene constante el efecto de esos mismos factores. Esta definición nos fue útil para entender la diferencia entre muestra y población. La población es el conjunto de resultados que se obtiene de repetir el experimento todas las veces posibles, mientras que cualquier número inferior de repeticiones nos proporciona sólo una muestra.

En este tema estamos interesados en aquellos experimentos en los que no estamos seguros del resultado que se va a obtener, y usaremos el cocepto de probabilidad para dar una medida de la incertidumbre sobre dicho resultado.

Al conjunto de factores que controlamos en un experimento le denominaremos **condiciones de experimentación**. Si dichas condiciones cambian, el experimento será diferente; no estaríamos repitiendo el mismo experimento, sino realizando otro experimento diferente. Nuestro interés en este tema es sobre los resultados que se obtienen al repetir **el mismo experimento**. En estadística es importante distinguir entre dos tipos de experimentos:

- **Experimento determinista:** Un experimento es **determinista** cuando al repetirse **siempre se observa el mismo resultado**. De esta forma, en un experimento determinista puede predecirse exactamente el dato que se va a obtener. La razón por la que se obtiene el mismo resultado es porque **en el experimento se controlan absolutamente todos los factores** que influyen sobre el resultado. Es decir, las condiciones de experimentación son completamente restrictivas e incluyen a todos los factores susceptibles de influir en el resultado del experimento. De esta forma si dichos factores se mantienen fijos, se obtiene siempre el mismo valor de la variable, pues no habrá nada que lo altere. Por ejemplo, el resultado de una operación matemática es determinista. El resultado de un modelo matemático construido para describir algún fenómeno también es determinista. En la realidad, es difícil tener este tipo de experimentos, pues habrá factores imposibles de controlar, y no podrán incluirse dentro de las condiciones de experimentación.

- **Experimento aleatorio:** Un experimento es **aleatorio** si al repetirle **no siempre se obtiene el mismo resultado**. Un experimento aleatorio es un esquema de experimentación más realista que un experimento determinista. En la realidad, será difícil diseñar experimentos en los que todos los factores estén bajo control, siendo la situación más frecuente aquella en las que las condiciones de experimentación (es decir, el conjunto de factores que decidimos controlar) supongan sólo una porción de los factores que influyan en el resultado. De esta forma, al repetir el experimento habrá circunstancias que habrán cambiado, lo que posibilita que el resultado sea diferente cada vez. Como el resultado del experimento aleatorio depende precisamente de los factores que no controlamos, **habrá incertidumbre sobre el resultado final**. La incertidumbre será tanto mayor cuanto más importantes sean los factores que no controlamos.

  En estadística, al efecto de los factores no controlados se le denomina azar. Por tanto, en un experimento aleatorio hay varios resultados posibles y en el valor finalmente observado interviene en mayor o menor medida el azar. Por ejemplo, si el experimento consiste en lanzar una moneda y observar el resultado, hay dos posibles resultados: cara y cruz, y en el resultado final intervienen factores imposibles de controlar: impulso en el lanzamiento, velocidad de giro de la moneda, tiempo hasta que se detiene, etc. Por tanto, no sabremos a ciencia cierta qué saldrá finalmente (dentro de las opciones posibles, naturalmente). Otro experimento podría consistir en medir cuánto tiempo tardará una máquina en realizar una tarea. En este segundo ejemplo hay infinitos resultados posibles, por ser el tiempo una variable continua, y hay igualmente incertidumbre de cuánto se tardará finalmente, pues dependerá de factores como rozamientos, desgastes, etc.. Una vez lanzada la moneda o una vez realizado el proceso la incertidumbre desaparecerá y observaremos el dato final.

En estadística usaremos el concepto de probabilidad para medir la indertidumbre de observar un determinado resultado **antes** de ejecutar el experimento. El conocimiento de dicha probabilidad será esencial para poder extraer conclusiones generalizables a futuras repeticiones del experimento. La probabilidad de un suceso puede utilizarse de dos formas principales:

1. El conocimiento de la probabilidad de un suceso ayudará a **valorar el riesgo de nuestras decisiones** o **anticipar los recursos** que nos preparen para dicho suceso. Esta actividad es puramente deductiva.

2. Una vez observado un conjunto de resultados de un experimento aleatorio, podemos utilizar dichas observaciones para **valorar si nuestras hipótesis sobre lo que esperábamos obtener eran o no razonables**. Esta valoración se realiza comparando los resultados obtenidos con la probabilidad que habíamos calculado para su aparición. Esta actividad combina tanto deducción como inducción (o inferencia). (¿por qué?). Por ejemplo, si después de lanzar un dado 1000 veces vemos que obtenemos siempre un dos, empezaremos a sospechar del dado.

Antes de entrar a definir el concepto de probabilidad continuaremos introduciendo algunas definiciones útiles.

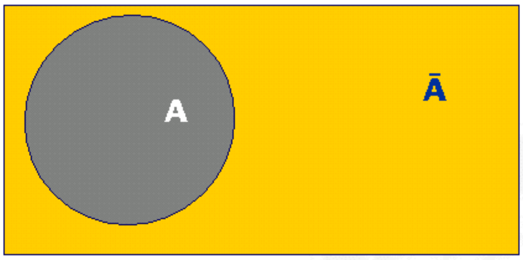
- **Suceso:** es el resultado o conjunto de resultados de un experimento que comparte alguna característica definida por el experimentador. Por ejemplo, en el lanzamiento de un dado, un suceso puede ser sacar un número par, o bien sacar exactamente el número 6, o sacar un número inferior a 3. Cada vez que al realizar un experimento obtenemos un valor contenido en la definición del suceso, diremos que hemos observado dicho suceso.

  En general. usaremos las letras mayúsculas para designar a los sucesos. Por ejemplo, sea el suceso A: obtener un número impar al lanzar un dado. Si lanzamos un dado 3 veces y obtenemos $\{1{,}5{,}3\}$ hemos observado el suceso A sólo una vez en esas tres repeticiones del experimento. Otro ejemplo, sea el suceso C: tardar menos de una hora en ejecutar la máquina M la tarea T. Si la máquina realiza la tarea 10 veces y en todas ellas ha tardado más de una hora, no habremos observado nunca dicho suceso.

- **Sucesos elementales:** Cada uno de los resultados elementales de un experimento aleatorio. Es decir, son los valores diferentes de la variable de interés que se obtienen al repetir el experimento. Por ejemplo, al lanzar un dado, los sucesos elementales son seis: $\{1{,}2{,}3{,}4{,}5{,}6\}$. Al lanzar una moneda, los sucesos elementales son dos: {cara, cruz}. Al medir el tiempo que la máquina M tarda en realizar la tarea T, los sucesos elementales son infinitos, al ser el tiempo una variable continua.

- **Sucesos compuestos:** cualquier unión de sucesos elementales es un suceso compuesto. Un suceso compuesto se suele definir mediante el conjunto de resultados o sucesos elementales que lo forman. Por ejemplo, el suceso A: obtener un valor par al lanzar un dado es un suceso compuesto, y se escribirá como $A : \{2, 4, 6\}$. Observar en la máquina anterior una duración superior a diez minutos en ejecutar la tarea es también un suceso compuesto y puede escribirse como $B : \{t \mid t > 10\}$, donde el símbolo '$|$' se lee 'dado que' o 'condicionado a'.

- **Suceso contrario o complementario:** Sea A un suceso. Llamaremos $\bar{A}$ al suceso que ocurre cuando no ocurre A. Por ejemplo, si A es el suceso: obtener un número par al lanzar un dado, entonces $\bar{A}$ será el suceso: obtener un número impar al lanzar un dado. Si A es el suceso: la máquina tarda más de 10 minutos en ejecutar la tarea, entonces $\bar{A}$ será el suceso: la máquina tarda 10 minutos o menos en ejecutar la tarea. Cuando observamos A, entonces no observarenos $\bar{A}$, y cuando no observamos A, entonces lo que observamos es $\bar{A}$. Al suceso contrario $\bar{A}$ también se le denomina suceso complementario. La siguiente figura muestra dos sucesos complementarios mediante un diagrama de Venn

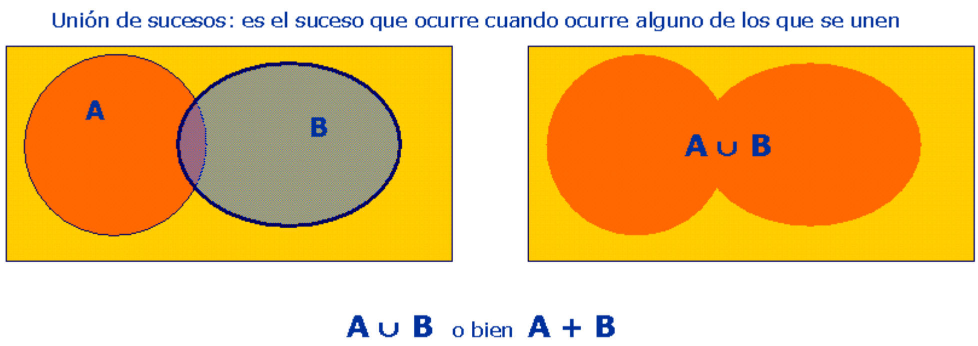


- **Espacio muestral:** es el conjunto de todos los sucesos que es posible observar al realizar un experimento. **El espacio muestral asociado a un experimento se construye uniendo todos los sucesos elementales**. Cualquer suceso observado, elemental o compuesto, estará dentro del espacio muestral. Por ejemplo, el suceso A: obener un 2 al lanzar un dado está dentro del espacio muestral del resultado del lanzamiento de un dado, así como el suceso B: obtener 1 ó 3; pero el suceso C: obtener un número mayor que 12 no está dentro del espacio muestral de dicho experimento.

- **Suceso seguro:** diremos que un suceso es seguro si siempre se observa. A este suceso le denotaremos por E. El espacio muestral es un suceso seguro. Por eso al espacio muestral se le suele denotar por la letra E.

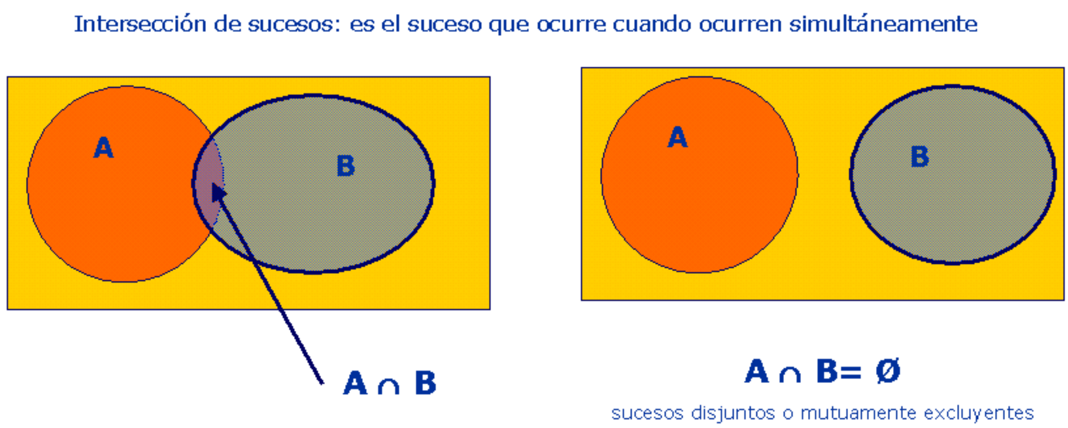
- **Suceso imposible:** es un suceso que nunca se puede observar, por estas fuera del espacio muestral se denomina suceso imposible, y se denota por $\emptyset$. Por ejemplo, obtener un 10 al lanzar un dado es un suceso imposible. Observar una duración negativa en la ejecución de una tarea por una máquina es también un suceso imposible.

- **Suceso unión $A \cup B$:** El suceso unión $A \cup B$ o también $A + B$ es el suceso que se observa si suceden alguno de los sucesos A y B. Es decir, puede observarse sólo A, sólo B o ambos. Al suceso unión también se le denomina A **ó** B. Por ejemplo, sea A: observar un número par al lanzar un dado, y B: observar un número mayor que 3 al lanzar un dado. Entonces $A \cup B = \{2, 4, 5, 6\}$. La figura siguiente ilustra la unión de dos sucesos mediante diagramas de Venn.



La unión de todos los sucesos elementales dará el espacio muestral E. Asímismo, la unión de un suceso y su complementario también dará el espacio muestral: $A + \bar{A} = E$.

- **Suceso intersección** $A \cap B$**:** El suceso intersección $A \cap B$ o AB es el suceso que se observa cuando se observan A **y** B simultáneamente. También se le denomina A y B. Utilizando el ejemplo anterior, sea A: observar un número par al lanzar un dado, y B: observar un número mayor que 3 al lanzar un dado. Entonces $A \cap B = \{4, 6\}$. **Cuando dos sucesos no comparten elementos, se les denomina disjuntos o mutuamente excluyentes**, y su intersección será el suceso imposible ($A \cap B = \emptyset$). El siguiente diagrama de Venn muestra el suceso intersección de dos sucesos.



Las operaciones unión e intersección verifican las siguientes propiedades:

	Unión	**Intersección**
Conmutativa	$A \cup B = B \cup A$	$A \cap B = B \cap A$
Asociativa	$A \cup (B \cup C) = (A \cup B) \cup C$	$A \cap (B \cap C) = (A \cap B) \cap C$
Idempotente	$A \cup A = A$	$A \cap A = A$
Simplificación	$A \cup (B \cap A) = A$	$A \cap (B \cup A) = A$
Distributiva	$A \cup (B \cap C) = (A \cup B) \cap (A \cup B)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap B)$
Elemento neutro	$A \cup \varnothing = A$	$A \cap E = A$
Absorción	$A \cup E = E$	$A \cap \varnothing = \varnothing$

A las familias de conjuntos que verifican las propiedades anteriores se les denomina álgebras de Boole. En el álgebra de Boole anterior se verifican las siguientes propiedades, conocidas como leyes de De Morgan:

- El suceso contrario de la unión de dos sucesos es la intersección de sus sucesos contrarios: $\overline{A \cup B} = \bar{A} \cap \bar{B}$
- El suceso contrario de la intersección de dos sucesos es la unión de sus sucesos contrarios: $\overline{A \cap B} = \bar{A} \cup \bar{B}$

2.2. Definición de probabilidad y propiedades

Supongamos que estamos interesados en la observación de un suceso, resultado de la ejecución de un experimento aleatorio. Salvo que dicho suceso sea el suceso seguro o un suceso imposible, nunca sabremos si ese suceso será finalmente observado o no. Habrá una incertidumbre sobre la observación de dicho suceso. El grado de incertidumbre, o análogamente, certidumbre, será mayor o menor dependiendo de cada caso concreto. Usaremos el concepto de probabilidad de un suceso para medir dicha incertidumbre. *Definiremos probabilidad de un suceso en un experimento aleatorio como la frecuencia relativa de aparición de dicho suceso si repitiésemos el experimento indefinidamente.* A veces esta probabilidad será fácil de cuantificar. Por ejemplo, la probabilidad de observar el suceso A: cara, al lanzar una moneda es de 0.5; es básicamente un razonamiento lógico una vez que conocemos las propiedades físicas de la moneda. En general, el cálculo de probabilidades es sencillo si todos los sucesos elementales son equiprobables.

Otras veces, el cálculo de la probabilidad de un suceso requerirá un proceso de experimentación para obtener dicha probabilidad empíricamente. Por ejemplo, sabremos la probabilidad de que una máquina tarde menos de 10 minutos en medir una tarea si medimos muchas veces dicha tarea. Será imposible repetir la tarea indefinidamente, como requeriría la aplicación estricta de la definición de probabilidad; pero tras un número elevado de repeticiones podemos conseguir una aproximación satisfactoria.

Ejemplo 1 *Un laboratorio ha diseñado dos tipos de aislante, aislante Tipo A y aislante Tipo B. El destino del aislante es cubrir un componente electrónico que ha de estar colocado en una atmósfera muy corrosiva durante un periodo continuado de 100 horas. Para evaluar la probabilidad de que un aislante resista a dicha atmósfera durante ese tiempo, se colocan un conjunto grande de elementos de ambos tipos de aislante durante 100 horas en dicha atmósfera. Después del experimento se observa que 80 de cada 100 aislantes de Tipo A siguen en buen estado, mientas que sólo 60 de cada 100 aislantes de Tipo B siguen en buen estado. De esta forma, puede concluirse que,* ***aproximadamente****, la probabilidad de que el aislante de Tipo A resista es $P(A) = 0{,}8$, y la probabilidad (aproximada) de que el aislante de Tipo B resista es $P(B) = 0{,}6$. Por lo tanto, el aislante Tipo A parece más aconsejable que el aislante Tipo B. (¿por qué esas probabilidades que se han obtenido se les califica como aproximadas?)*

En este ejemplo, la observación del estado de cada elemento de cada tipo de aislante puede considerarse una repetición de un experimento. Para hacer el cálculo exacto de la probabilidad de que un tipo de aislante sea adecuado, necesitaríamos infinitasrepeticiones, lo que es imposible en la práctica. Por eso, cuando obtenemos una probabilidad de forma empírica, sólo podremos asignar un valor de probabilidad aproximado basado en un número suficientemente grande de repeticiones.

En temas posteriores analizaremos qué número de repeticiones de un experimento necesitaremos como mínimo para que el cálculo empírico de una probabilidad tenga una fiabilidad suficiente.

habrá ocasiones en que la probabilidad sea simplemente una medida subjetiva. Por ejemplo, la probabilidad de que mañana llueva es una medida subjetiva de la certidumbre de que llueva, pues el mañana sólo lo podremos observar una vez. Sin embargo todo el mundo entiende que si la probabilidad de que mañana llueva es de 0.9 habrá gran riesgo de lluvia sin necesidad de imaginar la repetición de ningún experimento de viajes en el tiempo. En estas situaciones irrepetibles puede interpretarse que la probabilidad es la frecuencia relativa de observación del suceso en situaciones análogas.

En cualquier caso, tanto en situaciones objetivas o subjetivas, la probabilidad tiene las mismas propiedades que la frecuencia relativa. Sea A un suceso (simple o compuesto), resultado de un experimento aleatorio. Entonces la probabilidad de observar A se denotará por $P(A)$ y verifica las siguientes propiedades

1. $0 \leq P(A) \leq 1$
2. $P(E) = 1$
3. $P(\varnothing) = 0$
4. Sea $\bar{A}$ el suceso contrario o complementario de A, entonces $P(\bar{A}) = 1 - P(A)$
5. Si los sucesos A y B son mutuamente excluyentes $P(A + B) = P(A) + P(B)$.
6. Si A y B no son excluyentes $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Si los sucesos elementales son equiprobables, como sucede al lanzar un dado o una moneda, **la probabilidad de cada suceso elemental es $1/n$ donde n es el número de sucesos elementales**. Por eso es fácil deducir que la probabiidad de obtener cara al lanzar una moneda es $1/2$ y la de obtener un 4 al lanzar un dado es $1/6$. A este tipo de situaciones se le denomina **modelo de probabilidad uniforme**.

Siguiendo con este tipo de razonamiento puramente lógico para calcular probabilidades, **si el suceso cuya probabilidad nos interesa calcular es la unión de sucesos elementales, su probabilidad será la suma de las probabilidades de dichos sucesos elementales**, lo que se deduce de la propiedad 5 anterior. Por ejemplo, la probabilidad de obtener un número par al lanzar un dado es igual a la probabilidad de sacar 2 más la de sacar 4 más la de sacar 6, en total, $3/6$. Esta regla de cálculo de probabilidades se denomina **regla de Laplace,** y puede enunciarse como sigue:

- **Regla de Laplace**: Sea un espacio muestral E consistente en n sucesos elementales equiprobables, y sea A un suceso compuesto por k sucesos elementales, entonces

$$P(A) = \frac{k}{n} = \frac{\text{número de sucesos elementales favorables}}{\text{número de sucesos elementales posibles}}.$$

Hay que remarcar nuevamente que **la regla de Laplace sólo es aplicable en contextos en los que cada resultado elemental es equiprobable**. Fuera de este contexto en el que se manejan sucesos elementales equiprobables, el cálculo de probabilidades de sucesos puede complicarse enormemente. La literatura está llena de problemas clásicos de probabilidad realmente endiablados, para cuya resolución no cabe más que analizar con cuidado y paciencia cómo es el espacio muestral y como descomponer el suceso de interés en partes más sencillas. Problemas clásicos de probabilidad 'recreativa' se pueden encontrar, por ejemplo, en www.mathpages.com.

En las secciones siguientes vamos a analizar algunas reglas que nos permitan calcular probabilidades de sucesos complejos en función de la información que se tenga de otros sucesos más sencillos.

2.3. Probabilidad condicionada y total

La incertidumbre sobre la observación de un suceso depende del grado de información que tengamos, y por tanto la probabilidad de un mismo suceso puede variar según el conjunto de información. Por ejemplo, la probabilidad de obtener un 2 al lanzar un dado es $1/6$; sin embargo si alguien nos dice que el número que ha salido es par, entonces la probabilidad de que sea 2 será $1/3$. Podemos decir entonces que la **probabilidad incondicional** de sacar un 2 es $1/6$, pero la **probabilidad condicionada** a que el número ha sido par es $1/3$. La notación para este tipo de probabilidades es la siguiente. Llamemos A al suceso que no sabemos si observaremos o no y cuya probabilidad queremos calcular (obtener un 2 al lanzar un dado). Llamemos B al suceso que ya ha sido observado, y que precisamente por eso afecta a la incertidumbre sobre A (en nuestro ejemplo del dado, el suceso B sería obtener un número par). Entonces la probabilidad de A condicionada a B, o también la probabilidad de A dado B es

$$P(A|B).$$

El cálculo de $P(A|B)$ depende de la relación que haya entre ambos sucesos. Es posible obtenerla si conocemos $P(B)$ y $P(A \cup B)$ a través de la relación

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{2.1}$$

A esta relación la denominaremos **regla de la probabilidad condicionada**. Para entender y justificar esta fórmula usaremos el siguiente ejemplo. En una sala hay 300 personas. La siguiente tabla de frecuencias bivariante establece la clasificación por sexo y por ser o no fumador de esas 300 personas:

	Chicas	Chicos	Total fumadores
Fuma	15	15	30
No fuma	105	165	270
Total por sexo	120	180	300

Sea F el suceso: ser fumador; es decir, que al extraer a una persona al azar de entre los 300 resulte que es una persona fumadora. La probabilidad de ese suceso será la frecuencia relativa de su aparición al repetirse indefinidamente este experimento de extracción de un individuo al azar, es decir $P(F) = 30/300 = 0{,}1$. Nótese que estas repeticiones (imaginarias) del experimento son siempre sobre una base de 300 individuos, porque son extracciones con reposición. Una vez analizado un individuo, éste volvería al grupo. Este valor también puede obtenerse por la regla de Laplace anterior, pues todos los individuos tiene la misma probabilidad de ser seleccionados, pero sólo 30 de los 300 poseen el atributo definido por el suceso. Por tanto, la probabilidad de sepeccionar a un individuo fumador es 30/300=0.1.

Sea M el suceso: ser mujer; es decir, que al seleccionar a una persona al azar de entre las 300 resulte ser ua mujer. Entonces $P(M) = 120/300 = 0{,}4$. ¿Y la probabilidad del suceso $P(F|M)$? es decir, suponiendo que la persona seleccionada es una mujer ¿cuál es la probabilidad de que sea fumadora? En este caso, la probabilidad de que una persona fume dado que sea mujer será la frecuencia relativa de aparición de personas fumadoras dentro del colectivo femenino, formado por 120 personas. Las condiciones de experimentación son las de seleccionar personas al azar del grupo de 120 mujeres. Es por tanto 15/120. Se puede escribir entonces que

$$\begin{aligned}
P(F|M) &= \frac{\text{número de mujeres que fuman}}{\text{número de mujeres}} \\
&= \frac{\text{número de mujeres que fuman/número total de personas}}{\text{número de mujeres/número total de personas}} = \frac{15/300}{120/300} \\
&= \frac{P(\text{fumar y ser mujer})}{P(\text{ser mujer})} = \frac{P(F \cap M)}{P(M)},
\end{aligned}$$

que corresponde precisamente con la regla de la probabilidad condicionada expuesta en (2.1).

De (2.1) se obtiene también

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A), \tag{2.2}$$

que proporciona otra expresión útil para calcular probabilidades conjuntas a partir de otras probabilidades que conozcamos.

Ejemplo 2 *La expresión 2.2 nos proporciona otra forma de calcular la probabilidad de que al seleccionar a una persona al azar, del grupo de 300 personas expuestas anteriormente, sea una persona de sexo femenino y además fumadora. Es decir, que verifique $M \cap F$. De la tabla de distribución bivariante tenemos que* $P(M \cap F) = 15/300 = 0{,}05$. *Pero también lo podemos calcular si conocemos la probabilidad de que una chica fume (*$P(F|M) = 15/120$*), y la probabilidad de que se seleccione a una chica (*$P(M) = 120/300$*). Se tiene entonces*

$$P(M \cap F) = P(F|M)P(M) = \frac{15}{120} \times \frac{120}{300} = \frac{15}{300} = 0{,}05$$

En ocasiones estamos interesados en la probabilidad de observar un suceso A que sólo ha sido observado con anterioridad unido a otro suceso B. Por ejemplo, supongamos que sólo sabemos la proporción de hombres y mujeres que fuman ($P(F|M)$ y $P(F|H)$) y sabemos la proporción de hombres y mujeres ($P(M)$ y $P(H) = 1 - P(M)$). ¿Cuál es entonces la proporción de fumadores? es decir ¿qué vale $P(F)$? El objetivo puede interpretarse como el cálculo de una **probabilidad total**, $P(F)$, a partir de las condicionadas $P(F|M)$ y $P(F|H)$. Para ver cómo obtener esta probabilidad total, reordemos que de la definición de suceso seguro E se puede deducir que

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

Estas relaciones nos ayudarán a obtener $P(A)$ en función de la observación del suceso B. El razonamiento es el siguiente

$$P(A) = P(A \cap E) = P(A \cap (B \cup \bar{B}))$$

y de esta forma ya hemos introducido en escena el suceso B del que tenemos información. Entonces

$$\begin{aligned}
P(A \cap (B \cup \bar{B})) &= P(A \cap B) \cup (A \cap \bar{B}) \\
&= P(A \cap B) + P(A \cap \bar{B}) - P(A \cap B \cap A \cap \bar{B}) \\
&= P(A \cap B) + P(A \cap \bar{B})
\end{aligned}$$

pues $P(A \cap B \cap A \cap \bar{B}) = 0$, pues no es posible observar B y $\bar{B}$ simultáneamente, es decir $B \cap \bar{B} = \emptyset$. Usando (2.2) tenemos que

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap \bar{B}) \\ &= P(A|B)P(B) + P(A|\bar{B})P(\bar{B}) \end{aligned}$$

a este resultado se le llama **regla de la probabilidad total**. Este resultado se puede extender al caso en que en lugar de tener los sucesos B y $\bar{B}$ tenemos una separación en más categorías, es decir $B_1 \cup B_2 \cup B_3 \cup \cdots \cup B_k = \cup_{i=1}^{k} B_i = E$. Entonces podemos escribir que

$$P(A) = P\left(A \cap \left(\cup_{i=1}^{k} B_i\right)\right) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(A|B_i)P(B_i) \tag{2.3}$$

que nos permite reconstruir la probabilidad de un suceso después de haber observado la probabilidad de que ocurra cuando se observaban otros.

Ejemplo 3 *En nuestro ejemplo de personas fumadoras o no fumadoras tenemos que*

$$\begin{aligned} P(F|M) &= \frac{15}{120} = 0{,}125; P(M) = \frac{120}{300} = 0{,}4 \\ P(F|H) &= \frac{15}{180} = 0{,}0833 : P(H) = 0{,}6, \end{aligned}$$

y por tanto

$$P(F) = P(F|M)P(M) + P(F|H)P(H) = 0{,}10,$$

que vemos que coincide con el cálculo directo que se obtiene al observar los valores de la tabla, de donde se puede ver que $P(F) = 30/300 = 0{,}10$.

Ejemplo 4 *Una de las tareas más críticas en la gestión del tráfico de una red informática es la detección de un ataque externo. Dicha detección se hace analizando trazas de los datos que circulan. Se ha de disponer entonces de un algoritmo de detección (AD) que clasifique dicha traza como un ataque o no. Un AD se evalúa en función de dos características: la probabilidad de detectar un ataque,* P_d*, y la probabilidad de dar una falsa alarma.*P_f*. Si llamamos* I *al suceso de sufrir un ataque y* A *a su detección, tendremos que* $P_d = P(A|I)$ *y* $P_f = P(A|\bar{I})$*, donde* $\bar{I}$ *es el suceso complementario a* I*; es decir, que no haya ataque.*

La compañía SSi (www.ebusiness-security.com) comercializa un producto para la detección de ataques (http:// www.ebusiness-security.com/ eTrust_ Intrusion_ detection.htm). El AD que comercializa tiene unas características bastante buenas. La probabilidad de detectar un ataque es $P_d = 0{,}99$*, mientras que la probabilidad de falsa alarma es* $P_f = 0{,}002$*. (Lo ideal sería* $P_d = 1$ *y* $P_f = 0$*).*

Cuando el AD está analizando una unidad de información (packet) hay dos opciones, que dé alarma o que no dé alarma, es decir $P(A)$ *y* $P(\bar{A})$*. Si el sistema recibe por término medio un ataque cada 50.000 unidades de información ¿Cuál es la probabilidad de que el sistema dé una alarma? Si el sistema recibe una media de un millón de packets al día ¿Cuántas alarmas se darán por término medio?*

Solución:

Para calcular esta probabilidad usaremos la regla de la probabilidad total, pues tenemos la probabilidad de alarma condicionada a otro suceso, que se produzca un ataque, así como su probabilidad.

Se tiene entonces que

$$\begin{aligned} P(A) &= P(A|I)P(I) + P(A|\bar{I})P(\bar{I}) \\ &= 0{,}99 \times (1/50000) + 0{,}002 \times (1 - 1/50000) = 0{,}00201976, \end{aligned}$$

que es un número que a primera vista es muy bajo. El número de alarmas en un día será $10^6 \times P(A) \approx 2020$ *alarmas. Puesto que cada alarma ha de ser analizada, 2020 alarmas puede ser un número excesivamente elevado, lo que requerirá que los responsables de la red tengan un sistema de monitorización de alarmas eficiente.*

2.4. Independencia de sucesos

Dos sucesos son independientes si la observación de uno de ellos no aporta información sobre la aparición de otro. Por tanto, la aparición de uno no hace variar la probabilidad del otro suceso. Por tanto, si dos sucesos A y B son independientes se tiene que

$$\begin{aligned} P(A|B) &= P(A) \\ P(B|A) &= P(B). \end{aligned}$$

Por tanto, utilizando la regla de la probabilidad condicionada

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

y por tanto, si hay independencia

$$P(AB) = P(A)P(B). \tag{2.4}$$

A esta expresión se le denomiará **regla de la independencia**, y es con frecuencia utilizada para definir independencia.

Ejemplo 5 *Unas piezas cilíndricas pueden ser defectuosas por tener una longitud inadecuada o por tener un diámetro inadecuado, siendo ambos tipos de defectos independientes. La independencia se debe a que esos defectos aparecen se producen en momentos diferentes del proceso productivo. Si la proporción de cilindros con longitud inadecuada es de 5 % y la de cilindros con diámetro inadecuado es del 3 %. ¿Qué porcentaje de cilindros son defectuosos?*

Solución:

Si llamamos L al suceso: longitud inadecuada, y D al suceso diámetro inadecuado, entonces un cilindro es defectuosos si

$$P(\textit{defectuoso}) = P(L + D) = P(L) + P(D) - P(LD)$$

y al ser ambos sucesos independientes

$$P(LD) = P(L)P(D) = 0{,}05 \times 0{,}03 = 0{,}0015.$$

Por tanto

$$P(\textit{defectuoso}) = 0{,}05 + 0{,}03 - 0{,}0015 = 0{,}0785.$$

No debemos confundir sucesos independientes con sucesos mutuamente excluyentes (o disjuntos). Sucesos mutuamente excluyentes son aquellos que nunca pueden observarse simultáneamente. Dos sucesos mutuamente excluyentes son por tanto dependientes, pues si hemos observado uno de ellos, ya sabemos que el otro suceso no podrá ser observado. Por ejemplo, los sucesos elementales son mutuamente excluyentes. Al lanzar un dado no puede observarse un 2 y un 4 simultáneamente. En sucesos mutuamente excluyentes se verifica que, al ser $A \cap B = \varnothing$, $P(A \cap B) = 0$, por lo que si $P(A) \neq 0$ y $P(B) \neq 0$ se tiene que no se cumple la regla de la independencia y $P(A \cap B) \neq P(A)P(B)$. Lo que se tiene en dos sucesos A y B mutuamente excluyentes es que $P(A|B) = P(B|A) = 0$, pues al observarse uno de ellos, la probabilidad de observar el otro suceso es nula.

2.5. Teorema de Bayes

En las secciones anteriores hemos presentado diferentes formulaciones para calcular la probabilidad de un suceso en función de la diferente información que tengamos. En esdta sección estamos interesados en calcular $P(A|B)$ cuando lo que tenemos es precisamente la probabilidad de la situación inversa, es decir $P(B|A)$. De la fórmula de probabilidad condicionada se tiene que

$$P(A|B) = \frac{P(AB)}{P(B)} \tag{2.5}$$

pero, por otra parte

$$P(B|A) = \frac{P(AB)}{P(A)} \tag{2.6}$$

coincidiendo por tanto el numerador de ambas expresiones (2.5) y (2.6). Despejando en (2.6) y sustituyendo en (2.5) se tiene que

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{2.7}$$

resultado que se conoce como **Teorema de Bayes**. Es frecuente también expresar el Teorema de Bayes sustituyendo el denominador por su expresión respectiva usando el resultado de la probabilidad total, es decir

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}. \tag{2.8}$$

Ejemplo 6 *En sistemas complejos que se averían, no es fácil averiguar qué componente ha fallado, pues eso requeriría un desmontaje que puede ser caro si al final la pieza desmontada no era la causante del fallo. Por esa razón, puede tenerse un conjunto de pruebas indirectas alternativas que pueden ayudar a detectar el origen del fallo. Dichas pruebas pueden tener cierto margen de error. Es decir, pueden detectar fallos inexistentes o no detectarlos cuando los hay.*

Supongamos un componente de un sistema y que la probabilidad de que dicho componente se averíe en un período de tiempo dado es 0,01*. Su estado (averiado, funcionando) se comprueba con un ensayo que cumple que cuando el componente funciona la probabilidad de que el ensayo diga lo contrario es* 0,05*, pero si el componente está averiado el ensayo no se equivoca. Si el ensayo indica que el componente está averiado, ¿cuál es la probabilidad de que realmente lo esté?*

Solución:

Para poder calcular la probabilidad que se pide, lo primero es traducir el texto de nuestro problema en términos estadísticos. Es necesaro darse cuenta qué sucesos se han observado y de qué sucesos tenemos incertidumbre sobre si serán o no observados.

Llamemos A y F a los sucesos 'el componente está averiado' y 'el componente funciona', respectivamente. Llamemos a y f a los respectivos resultados del ensayo; es decir a :'el ensayo dice que el componente está averiado' y b :' el ensayo dice que el componente funciona'. Cuando establecemos que la probabilidad de que se averíe es 0.01, equivale a $P(A) = 0{,}01$. *Si cuando el componente funciona (suceso observado) la probabilidad de que el ensayo diga lo contrario (hay pues incertidumbre) es* $0{,}05$, *equivale a* $P(a|F) = 0{,}05$. *Si cuando el componente está averiado (suceso observado) el ensayo no se equivoca (declaración sobre su incertidumbre) tendremos que* $P(a|A) = 1$. *Y lo que queremos calcular es que si sabemos que el el ensayo da resultado de avería (suceso observado), cuál es la probabilidad de que realmente lo esté (suceso sobre el que hay incertidumbre), que equivale a* $P(A|a)$. *Puede verse que la probabilidad condicionada que queremos calcular* $P(A|a)$ *es la contraria, en el sentido de los sucesos que conocemos y desconocemos, a las probabilidades condicionadas que ya conocemos* $P(a|A)$ *y* $P(a|F)$. *Por tanto puede resolverse con el Teorema de Bayes. Se tiene entonces que*

$$P(A \mid a) = \frac{P(a \mid A)P(A)}{P(a \mid A)P(A) + P(a \mid F)P(F)} = \frac{1 \times 0{,}01}{1 \times 0{,}01 + 0{,}05 \times 0{,}99} = 0{,}168.$$

El numerador de esta fracción representa la probabilidad de que el componente esté averiado y el ensayo así lo indique, y el denominador representa la probabilidad de que el ensayo dé como resultado que el componente está averiado.

Ejemplo 7 *Sigamos con el Ejemplo 4 anterior del AD para detectar intrusiones en un sistema. ¿Cuál es la probabilidad de que al analizar una alarma, ésta sea falsa?*

Aquí se ha de tener mucho cuidado con el lenguaje, pues es fácil confundirse. Antes hemos definido falsa alarma como $P_f = P(A|\bar{I})$. P_f *es la proporción a largo plazo de packets analizados en los que no había ataque y sin embargo sí se dió la alarma. El experimento que se repetía era el análisis de packets sin intrusión, y el resultado era alarma o no-alarma. Ahora analizamos situaciones de alarma, y el resultado es que ha habido o no ha habido intrusión, y lo que queremos es calcular la probabilidad de que al analizar una alarma, el resultado haya sido negativo. Para distinguirlo de la situación anterior de falsa alarma, a esta situación le llamaremos Detección Negativa, y su probabilidad* $P_n = P(\bar{I}|A)$, *mientras que una Detección Positiva se haría con probabilidad* $P_p = P(I|A)$.

Utilizaremos el Teorema de Bayes, pues necesitamos calcular una probabilidad condicionada pero lo que tenemos es precisamente la probabilidad condicionada opuesta. Por el teorema de Bayes tenemos que

$$P(\bar{I}|A) = \frac{P(A|\bar{I})P(\bar{I})}{P(A)} = \frac{0{,}002 \times (1 - 1/50000)}{0{,}00201976} = 0{,}99.$$

Luego la inmensa mayoría de las alarmas analizadas son detecciones negativas. Este resultado puede ser muy frustrante para los técnicos de seguridad, pues quiete decir que invierten la mayoría de su tiempo con alarmas innecesarias. Sin embargo, es una situación difícil de evitar.

Este hecho, el que un técnico de seguridad de una red dedique la mayoría de su tiempo a analizar detecciones negativas es un problema importante pues lleva al técnico a rechazar el AD. Sin embargo, como se vió antes, el AD tenía unas características bastante buenas.

¿Cuál es la solución a este problema? Del análisis anterior se deduce que una posibilidad es reducir aún más la probabilidad de falsa alarma P_f (¿por qué?). Es necesario entonces que el AD tenga una probabilidad de falsa alrma realmente baja. Si $P_f = 0{,}0001$ (veinte veces menor que el anterior) se tendrá que

$$\begin{aligned} P(A) &= P(A|I)P(I) + P(A|\bar{I})P(\bar{I}) \\ &= 0{,}99 \times (1/50000) + 0{,}0001 \times (1 - 1/50000) = 1{,}19798 \times 10^{-4}. \end{aligned}$$

Entonces

$$P(\bar{I}|A) = \frac{P(A|\bar{I})P(\bar{I})}{P(A)} = \frac{0{,}0001 \times (1 - 1/50000)}{1{,}19798 \times 10^{-4}} = 0{,}83,$$

que aunque elevado, es menor que el anterior. Se deduce entonces que una seguridad efectiva ante intrusiones necesita de un sistema de detección altamente preciso así como una labor de análisis de alarmas donde es de esperar un elevado porcentaje de detecciones negativas, sin que ello deba interpretarse como que el AD no funciona.

En las expresiones del Teorema de Bayes (2.7) y (2.8) se ha usado que sólo tenemos el suceso A y su complementario $\bar{A}$. Estas expresoines pueden fácilmente generalizarse para el caso en que tengamos más de dos sucesos elementales, por ejemplo $A_1, A_2, ..., A_J$, tal que $\cup_{j=1}^{J} A_j = E$. Entonces, aplicando el resultado de la probabilidad total (2.3), el Teorema de Bayes se escribiría como

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{J} P(B|A_j)P(A_j)},$$

que es una expresión más general que las anteriores.

La fórmula del Teorema de Bayes también puede escribirse como

$$P(A|B) = \frac{P(B|A)}{P(B)} P(A), \tag{2.9}$$

donde $P(A)$ es la probabilidad de A **antes de** observar B, que denominaremos probabilidad 'apriori', y $P(A|B)$ es la nueva probabilidad de A **después de** haber observado B, y que denominaremos probabilidad 'a posteriori'. La expresión (2.9) muestra así una forma de actualizar la probabilidad del suceso A una vez que observamos B. Si B y A son independientes tendremos que $P(B|A) = P(B)$ y por tanto $P(B|A)/P(B) = 1$. Se tiene entonces que $P(A|B) = P(A)$, y por tanto, el haber observado B no altera nuestro conocimiento de A.

Ejemplo 8 *Una empresa petrolífera ha de decidir si un emplazamiento es adecuado para hacer una prospección petrolífera. La empresa iniciará la propección si después de analizar el terreno, la probabilidad de encontrar petróleo es mayor que 0.5.*

Por una parte se tiene que dadas las condiciones geológicas de la zona, la probabilidad de que en el emplazamiento haya petróleo es de sólo 0.4. Es decir, en 40 de cada 100 emplazamientos con similares condiciones geológicas se ha encontrado petróleo. Existe una forma adicional, aunque más compleja, de obtener más información sobre el potencial del emplazamiento. Es posible contratar a una empresa de ingeniería una prueba sísmica para detectar la presencia de petróleo. Esta prueba sísmica tampoco es del todo concluyente. La experiencia revela que cuando realmente hay petróleo, la prueba sísmica da un resultado positivo el 40 % de las veces, mientas que cuando no hay petróleo, la prueba sísmica detecta erróneamente la presencia de petróleo el 10 % de las veces ¿Debe la empresa petrolífera contratar esa prueba sísmica?

Solución:

Lo que se desea saber es si la prueba sísmica puede dar una mayor evidencia de la existencia de petróleo. Si llamamos A al suceso: hay petróleo, se tiene que usando sólo la información geológica $P(A) = 0{,}4$*, lo que no parece ser suficiente para acometer la inversión de una prospección.*

Si realizamos la prueba sísmica, se pueden tener los resultados a *:'la prueba detecta petróleo', y* $\bar{a}$ *: 'la prueba no detecta petróleo'. Lo que se desea saber es si la prueba puede proporcionar pruebas suficientes a favor de la existencia de petróleo. Es decir, si la prueba puede decirnos si la probabilidad de petróleo es mayor del 50 %. En términos estadísticos, lo que queremos saber es si*

$$P(A|a) > 0{,}5.$$

Lo que sabemos de esa prueba es que cuando hay petróleo, acierta el 40 % de las veces, es decir $P(a|A) = 0{,}4$*. Por otra parte, cuando no hay petróleo, dice que sí lo hay el 10 % de las veces, es decir* $P(a|\bar{A}) = 0{,}10$*. Aplicando Bayes:*

$$P(A|a) = \frac{P(a|A)}{P(a)}P(A)$$

que muestra cómo la realización de la prueba puede actualizar nuestro conocimiento sobre la existencia de petróleo. Aplicando el resultado de la probabilidad total

$$P(a) = P(a|A)P(A) + P(a|\bar{A})P(\bar{A}) = 0{,}4 \times 0{,}4 + 0{,}1 \times 0{,}6 = 0{,}22,$$

luego en terrenos como el que nos ocupa (con $P(A) = 0{,}4$*) el 22 % de las veces que se hace la prueba da positiva. Hay por tanto una probabilidad muy alta (0.78) de que la prueba nos haga desistir de hacer prospección. Pero si da positiva, la probabilidad de que finalmente haya petróleo sería*

$$P(A|a) = \frac{0{,}4}{0{,}22}0{,}4 = 0{,}73 > 0{,}5$$

Por tanto, si la prueba da positivo, la probabilidad de que realmente hubiese petróleo son muy altas. Por tanto, a la empresa petrolífera tendrá mucho interés en hacer el ensayo sísmico en ese tereno. (¿en qué tipo de terrenos no le interesará?)

Si la prueba diese negativa, la probabilidad de que el ensayo acierte y realmente no haya petróleo será

$$P(\bar{A}|\bar{a}) = \frac{P(\bar{a}|\bar{A})}{P(\bar{a})}P(\bar{A}),$$

donde $P(\bar{a}) = 1 - P(a) = 0{,}78$*, y* $P(\bar{a}|\bar{A}) = 1 - P(a|\bar{A}) = 1 - 0{,}1 = 0{,}9$*. Entonces*

$$P(\bar{A}|\bar{a}) = \frac{0{,}9}{0{,}78}0{,}6 = 0{,}69,$$

por lo que la prueba realmente ayuda a tomar una decisión correcta.