Repaso de Cálculo de Probabilidades Básico

1.2. Introducción

Se comienza este tema con la noción de probabilidad y la terminología subyacente. La probabilidad constituye por sí misma un concepto básico que refleja su relación con la faceta del mundo exterior que pretende estudiar: los fenómenos aleatorios, que suponen unas ciertas reglas de comportamiento. De alguna manera el concepto de probabilidad se relaciona o recuerda las propiedades de la frecuencia relativa. A partir de ella, y junto con las definiciones de probabilidad condicionada y la de sucesos independientes, se deducen los resultados fundamentales del Cálculo de Probabilidades.

Luego, se muestra el nexo que une la teoría de la probabilidad y la estadística aplicada: la noción de variable aleatoria, mostrando de esta manera cómo puede emplearse la teoría de la probabilidad para sacar conclusiones precisas acerca de una población sobre la base de una muestra extraída de ella. Muchos de los análisis estadísticos son, de hecho, estudio de las propiedades de una o más variables aleatorias.

En las aplicaciones prácticas es importante poder describir los rasgos principales de una distribución, es decir, caracterizar los resultados de un experimento aleatorio mediante unos parámetros. Se llega así al estudio de las características asociadas a una variable aleatoria, introduciendo los conceptos de esperanza y varianza matemática y relacionándolos con los conceptos de media y varianza de una variable estadística.

1.3. Experimentos y sucesos aleatorios

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

- Se puede repetir indefinidamente, siempre en las mismas condiciones;

- Antes de realizarlo, no se puede predecir el resultado que se va a obtener;

- El resultado que se obtenga, e, pertenece a un conjunto conocido previamente de resultados posibles. A este conjunto, de resultados posibles, lo denominaremos **espacio muestral** y lo denotaremos normalmente mediante la letra E. Los elementos del espacio muestral se denominan **sucesos elementales**.

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

  Cualquier subconjunto[1] de E se denominará **suceso aleatorio**, y se denotará normalmente con las letras $A, B, ...$

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

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

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

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

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

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

[1] En lo que sigue, no nos preocuparemos de cuestiones de medibilidad.

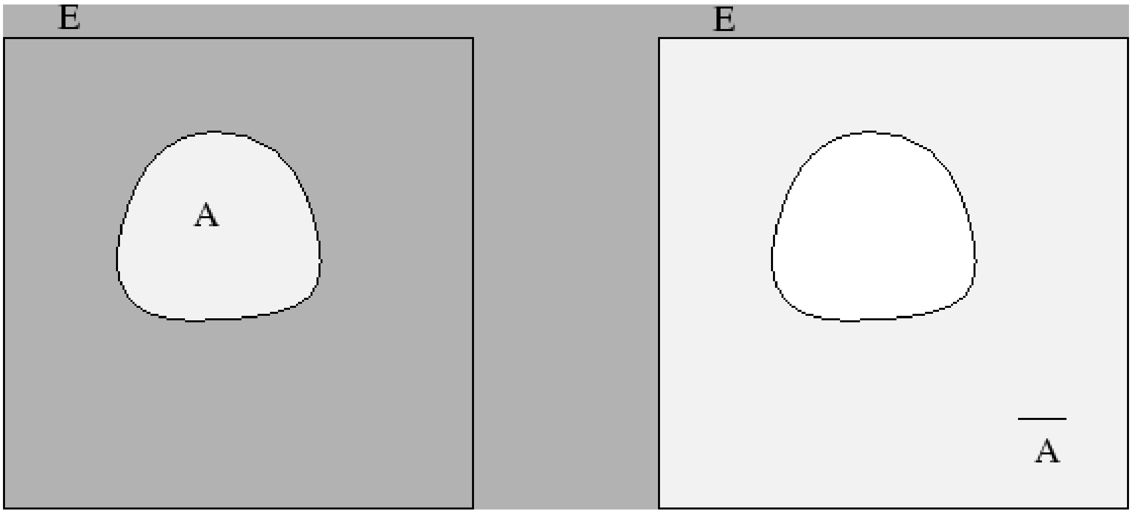


Figura 4.1: Representación gráfica de un suceso aleatorio $A \subset E$ y de su suceso contrario

1.3.1. Ejemplo

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

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

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

1.4. Operaciones básicas con sucesos aleatorios

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

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

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

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

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

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

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

A veces por comodidad se omite el símbolo $\cap$ para denotar la intersección de conjuntos, sobre todo cuando el número de conjuntos que intervienen en la expresión es grande. En particular, podremos usar la siguiente notación como equivalente a la intersección:

$$A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_{n-1} \cap A_n \stackrel{def}{\equiv} A_1A_2A_3 \cdots A_{n-1}A_n.$$

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

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

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

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

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

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

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

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

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

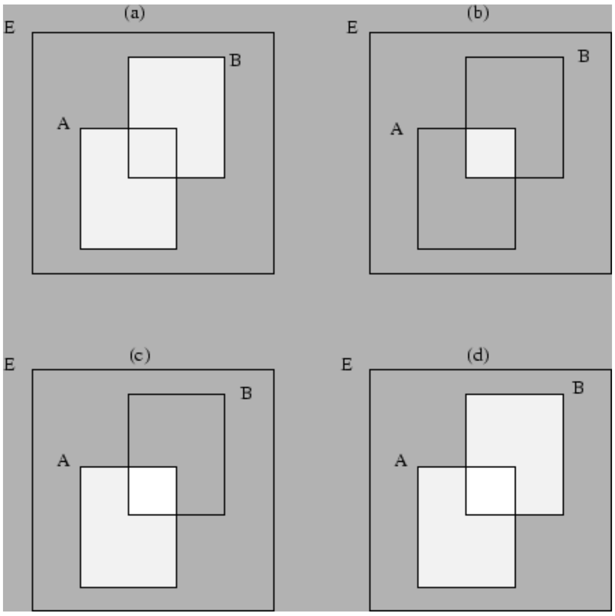


Figura 4.2: Dados dos sucesos aleatorios $A, B \subset E$ se representa: en (a) $A \cup B$; en (b) $A \cap B$; en (c) $A - B$; en (d) $A \triangle B$

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

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

1.4.1. Experimentos aleatorios y probabilidad

Se denominan experimentos deterministas aquellos que realizados de una misma forma y con las mismas condiciones iniciales, ofrecen siempre el mismo resultado. Como ejemplo, tenemos que un objeto de cualquier masa partiendo de un estado inicial de reposo, y dejado caer al vacío desde una torre, llega siempre al suelo con la misma velocidad $v = \sqrt{2gh}$.

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

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

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

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

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

1.5. Definición axiomática de probabilidad

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

- La probabilidad sólo puede tomar valores comprendidos entre 0 y 1 (no puede haber sucesos cuya probabilidad de ocurrir sea del 200 % ni del −5 %);
- La probabilidad del suceso seguro es 1, es decir, el 100 %;
- La probabilidad del suceso imposible debe ser 0;
- La probabilidad de la intersección de dos sucesos debe ser menor o igual que la probabilidad de cada uno de los sucesos por separado,

$$\begin{aligned} P(A\cap B) &\leq P(A) \\ P(A\cap B) &\leq P(B) \end{aligned}$$

- La probabilidad de la unión de sucesos debe ser mayor o igual que la de cada uno de los sucesos por separado,

$$\begin{aligned} P(A\cup B) &\geq P(A) \\ P(A\cup B) &\geq P(B) \end{aligned}$$

  Más aún, si los sucesos son disjuntos (incompatibles) debe ocurrir que

$$A\cap B = \emptyset \Longrightarrow P(A\cup B) = P(A) + P(B)$$

- La probabilidad del suceso contrario de A, debe valer

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

Esto en realidad puede deducirse del siguiente razonamiento:

$$1 = P(E) = P(A \cup A^c) = P(A) + P(A^c) \Longrightarrow P(A^c) = 1 - P(A).$$

En las últimas líneas se han esbozado ciertas propiedades que debería cumplir una función que se comporte como la *probabilidad*. Hemos de tener en cuenta entonces que siguiendo esos puntos:

1. La función de probabilidad debe calcularse sobre subconjuntos de E. No es estrictamente necesario que sean todos, pero sí es necesario que si se puede calcular sobre un conjunto, lo pueda ser también sobre su complementario, y que si se puede calcular sobre dos conjuntos A y B, también se pueda calcular sobre su unión y su intersección. Para ello, introduciremos el concepto de σ-álgebra de sucesos, que será una clase de subconjuntos de E sobre los que podamos aplicar las reglas de la probabilidad.

2. Entre las leyes que debe cumplir una función de probabilidad y que hemos descrito antes, hemos observado que algunas son redundantes, ya que se pueden deducir de las demás. Con la definición axiomática de la probabilidad, pretendemos dar el menor conjunto posible de estas reglas, para que las demás se deduzcan como una simple consecuencia de ellas.

Se precisan, a continuación, los conceptos de σ-álgebra de sucesos y de probabilidad.

1.5.1. Concepto de σ-álgebra de sucesos

Sea $\mathcal{A}$ una clase no vacía formada por ciertos subconjuntos del espacio muestral E. Diremos que esta clase es una σ-álgebra de sucesos si los sucesos complementarios de aquéllos que están en $\mathcal{A}$, también están en $\mathcal{A}$, así como sus uniones numerables (sean finitas o infinitas). Esto se puede enunciar como:

$$\begin{aligned} \forall A &\in \mathcal{A} \Longrightarrow A^c \in \mathcal{A} \\ \forall A_1, \ldots, A_n &\in \mathcal{A} \Longrightarrow \bigcup_{i=1}^{n} A_i \in \mathcal{A} \end{aligned}$$

1.5.2. Concepto axiomático de probabilidad

Dado un espacio muestral E, y una σ-álgebra de sucesos $\mathcal{A}$ sobre él, diremos que $\mathcal{P}$ es una probabilidad sobre $\mathcal{A}$ si se verifica lo siguiente (*axiomas*):

Axioma 1 La probabilidad es una función definida sobre $\mathcal{A}$ y que sólo toma valores positivos comprendidos entre 0 y 1.

$$\begin{aligned} \mathcal{P} \quad &: \quad \mathcal{A} \longrightarrow [0,1] \subset \mathbb{R} \\ \text{Dado } A \quad &\subset \quad E,\ A \in \mathcal{A} \text{ se cumple que } 0 \leq P(A) \leq 1 \end{aligned}$$

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

Axioma 3. La probabilidad de la unión numerable de sucesos disjuntos es la suma de sus probabilidades (figura 4.4):

$$A_1, \ldots A_n, \ldots \in \mathcal{A} \Longrightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

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

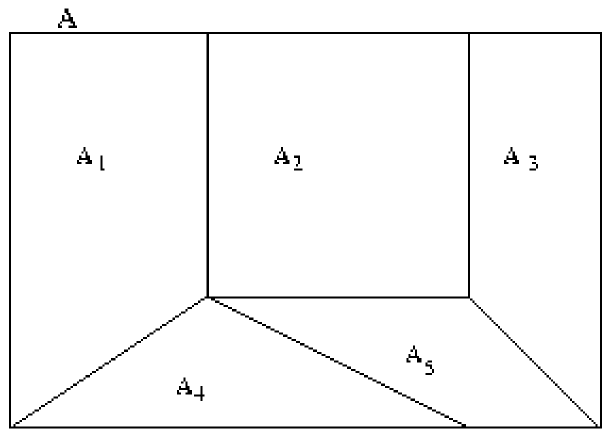


Figura 4.4: Visualización del axioma 3

1.5.3. Observación

La introducción de la definición de σ-álgebra puede parecer innecesaria a primera vista, ya que es una clase formada por subconjuntos de E que verifican ciertas propiedades relativas a la complementariedad y a las uniones finitas que ya verifica de antemano el conjunto denominado partes de $E, P(E)$, formado por todos los subconjuntos de E. Cuando el conjunto E de los posibles resultados de un experimento aleatorio sea finito, normalmente consideraremos como σ-álgebra de sucesos al conjunto $P(E)$. Esto ocurre cuando por ejemplo realizamos el experimento aleatorio de lanzar un dado: $E = \{1, 2, 3, 4, 5, 6\}$

$$\mathcal{A} = P(E) = \{\emptyset, E, \{1\}, \{2\}, \ldots, \{1, 2\}, \{1, 3\}, \ldots, \{1, 2, 3\}, \ldots\}.$$

Cuando E es infinito no numerable, la estructura del conjunto $P(E)$ puede presentar propiedades extremadamente engorrosas. Entonces es más conveniente utilizar como σ-álgebra un subconjunto más pequeño suyo que nos permita realizar las operaciones de complementariedad o de uniones finitas que se precisan en la definición de un σ-álgebra. Por ejemplo, si realizamos el experimento aleatorio de esperar el tiempo que hace falta para que un átomo de carbono catorce, C^{14}, se desintegre de modo natural, se tiene que $E = \mathbb{R}^+$. Sin embargo, la σ-álgebra de sucesos que consideramos no es $P(\mathbb{R}^+)$, que es una clase demasiado compleja para definir sobre sus elementos una medida de probabilidad. En su lugar, consideramos la σ-álgebra formada por todos los intervalos, abiertos o cerrados, y sus uniones numerables:

$$\mathcal{A} = \left\{\emptyset, \mathbb{R}^+, (2, 4], (4, 5], \ldots\right\}$$

lo que por supuesto incluye a los puntos de $\mathbb{R}^+$, ya que por ejemplo $\{2\} = [2, 2]$.

Este tipo de conjuntos (los intervalos) son los que interesan en la práctica por lo que se considera el siguiente convenio:

No haremos en general referencia a la σ-álgebra de sucesos más que cuando sea estrictamente necesario. De este modo cuando a partir de ahora se diga $A \subset E$, nos referiremos ímplicitamente a que $A \in \mathcal{A}$ donde $\mathcal{A}$ es una σ-álgebra de sucesos asociado a E y sobre la que se ha definido la función de probabilidad.

Si el espacio muestral es finito o infinito numerable, entenderemos que el σ-álgebra de sucesos es por defecto $P(E)$.

Si E es un conjunto infinito no numerable como $\mathbb{R}$ ó $\mathbb{R}^+$, o subconjuntos suyos en forma de intervalos, entenderemos que la σ-álgebra asociada es la mencionada en el ejemplo anterior, es decir, la generada por todos los intervalos abiertos, cerrados o semi-abiertos (lo que incluye en particular a los puntos), y sus uniones finitas. De este modo podremos calcular probabilidades como las de dichos intervalos.

1.5.4. Interpretación clásica de la probabilidad

En muchas ocasiones, por razones de simetría física o lógica, encontramos todos los resultados igualmente verosímiles y se apela al concepto clásico de probabilidad que se define mediante el cociente entre el número de casos favorables al suceso y al número de casos posibles. Si un experimento cualquiera puede dar lugar a un número finito de resultados posibles, y no existe ninguna razón que privilegie unos resultados en contra de otros, se calcula la probabilidad de un suceso aleatorio A, según la *regla de Laplace*, como el cociente entre el número de casos favorables a A, y el de todos los posibles resultados del experimento:

$$P(A) = \frac{\text{número de casos favorables a } A}{\text{número de casos posibles}}.$$

Ejemplo

Calcular la probabilidad de que al lanzar un dado se obtenga un número impar.

Solución: El espacio muestral es $E = \{1, 2, 3, 4, 5, 6\}$. Vamos a llamar A, al suceso consistente en que el resultado es impar, $A = \{1, 3, 5\}$. Como no suponemos que ninguna de las caras ofrece una probabilidad de ocurrencia diferente a las demás, podemos aplicar la regla de Laplace para obtener que

$$\begin{aligned} P(A) &= \frac{\text{número de casos favorables a } A}{\text{número de casos posibles}} = \\ &= \frac{\text{número de elementos en } A}{\text{número de elementos en } E} = \frac{3}{6} = 0{,}5 \end{aligned}$$

Ejemplo

Deseamos calcular la probabilidad de que al extraer una carta de una baraja de póker (52 cartas) obtengamos una *pica*. Si A es el suceso *obtener pica* se tiene

$$P(A) = \frac{13}{52} = \frac{1}{4}.$$

Ejemplo

Tiramos dos dados, uno blanco y otro negro. Se desea calcular las probabilidades de los sucesos:

- Sacar un 4 en el dado negro:

$$P(A) = \frac{6}{36} = \frac{1}{6}.$$

- Sacar más puntuación en el dado blanco que en el negro:

$$P(B) = \frac{5+4+3+2+1}{36} = \frac{15}{36} = \frac{5}{12}.$$

- Sacar 2, 3 ó 12

$$P(C) = \frac{1+2+1}{36} = \frac{1}{9}.$$

- La mayor puntuación es 5:

$$P(D) = \frac{9}{36} = \frac{1}{4}.$$

Es fácil comprobar que las probabilidades así definidas son, en efecto, probabilidades. Si n_A designa el número de casos favorables al suceso A y n al número de casos posibles, resulta que

- $P(A) = \frac{n_A}{n} \geq 0$, pues $n_A \geq 0$.
- $P(E) = \frac{n_E}{n} = \frac{n}{n} = 1$
- $P(A \cup B) = \frac{n_{A \cup B}}{n} = \frac{n_A + n_B}{n} = \frac{n_A}{n} + \frac{n_B}{n} = P(A) + P(B)$, si $A \cap B = \emptyset$.

Sin embargo, se observa que en muchas ocasiones este concepto de probabilidad no es aplicable (por ejemplo, si se desea calcular las probabilidades de quién ganará la liga de fútbol este año).

1.5.5. Interpretación frecuentista de la probabilidad

Un concepto más general de probabilidades es el frecuentista. Se aplica a experimentos que se pueden repetir indefinidamente bajo condiciones similares, adoptándose la hipótesis implícita de que las frecuencias relativas se estabilizan al repetirse el experimento.

Por ejemplo, en la Figura 4.3 se presenta la evolución de la frecuencia relativa del número de caras obtenido en el lanzamiento de una moneda en 100 ocasiones (simulado por un ordenador). En principio la evolución de las frecuencias relativas es errática, pero a medida que el número de tiradas aumenta, tiende a lo que entendemos por probabilidad de cara.

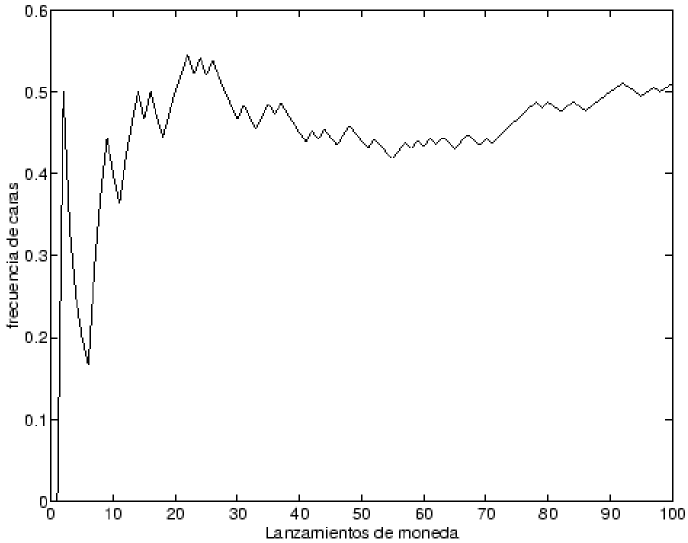


Figura 4.3: Convergencia a 1/2 de la frecuencia relativa del número de caras obtenido en lanzamientos sucesivos de una moneda (simulación en ordenador).

Supongamos, en general, que un suceso A ocurre A_n veces en n repeticiones del experimento. Resultaría, entonces, la definición frecuentista de la probabilidad de A como

$$P(A) = \lim_n \frac{A_n}{n},$$

supuesto que existe el límite. Con esta definición se tiene

- $\frac{A_n}{n} \geq 0$, $\forall n$, lo que implica que $P(A) = \lim_n \frac{A_n}{n} \geq 0$
- $A = E$, $\forall n$, lo que implica que $A_n = n$ y $P(E) = \lim_n \frac{n}{n} = 1$
- $P(A \cup B) = \lim_n \frac{(A \cup B)_n}{n} = \lim_n \frac{A_n + B_n}{n} = \lim_n \frac{A_n}{n} + \lim_n \frac{B_n}{n} = P(A) + P(B)$, supuestos A y B disjuntos.

Se observa que no es posible, sin embargo, hablar de una sucesión infinita de repeticiones por lo que, en la práctica hablaremos de sucesiones largas:

La frecuencia a largo plazo de un suceso es la fracción de tiempo que ocurre tal suceso en una sucesión larga de ensayos.

Por ejemplo, si se tira un dado 10000 veces y se obtiene 1510 veces el 6, se puede decir que la probabilidad de obtener 6 al tirar ese dado es 0,15. O, si en una *UCI* infantil ha habido 623 ingresos, de 3200, con la enfermedad A, puedo decir que la probabilidad de que ingrese un niño con la enfermedad A será de $623/3200$.

Obviamente, si el número de ensayos es pequeño, carece de sentido emplear este concepto. Por ejemplo, si tiro un dado y no aparece ninguna vez el 4, en 5 tiradas, debería estimar la probabilidad de que aparezca un 4 en la próxima tirada mediante $0/5$ (lo cual es absurdo).

Otras veces los experimentos aleatorios no pueden ser realizados, como es el caso de calcular la probabilidad de morir jugando a la ruleta rusa con un revólver: no es posible (o no se debe) calcular esta probabilidad repitiendo el experimento un número indefinidamente alto de veces para aproximarla mediante la frecuencia relativa.

1.5.6. Probabilidades como grado de confianza

El concepto que se introduce ahora es completamente distinto a los anteriores, en el sentido de que mientras los anteriores presentaban la probabilidad como una propiedad del sistema que se observa, ahora lo presentamos como una propiedad del observador del sistema: la probabilidad se presenta como una medida del grado de creencia que tiene una persona sobre la ocurrencia del suceso de interés.

La primera cuestión que surge es si las creencias tienen cabida en la Ciencia y la Ingeniería. La respuesta, en nuestra opinión, es afirmativa puesto que en numerosas fases de la historia de la Ciencia, y en distintas ramas, han existido varias teorías contradictorias que defendían distintas concepciones de un fenómeno. En muchas ocasiones, con la recogida de nueva información, algunas teorías han quedado refutadas, mientras que otras han evolucionado y convergido a una teoría mejor.

El segundo punto a destacar es que las creencias dependerán de cada persona y de la información que tengamos en cada momento. Por ejemplo, ante el problema de paternidad del *Cordobés chico,* la madre de éste, una aficionada a las revistas *del corazón,* y yo, tendremos distinta opinión y tendremos distinto grado de creencia sobre el hecho de que el Cordobés sea el padre real. Si discutiésemos entre nosotros y compartiésemos la información, cambiarían, tal vez, nuestras creencias hasta converger eventualmente.

Para nosotros, las dos grandes ventajas de la concepción de la probabilidad como grado de creencia es que, primero, podremos aplicarla en cualquier situación en que una persona tenga una opinión; así, si consideramos la ignorancia como una opinión, podremos aplicar siempre este concepto de probabilidad (frente a lo limitado de los conceptos clásico y frecuentista). La segunda ventaja es que las probabilidades cambian al recibirse nueva información, lo que las hace especialmente útiles en nuestra concepción estadística.

Por ello, definiremos la probabilidad de un suceso como *la medida del grado de creencia que tiene una persona dada en un momento dado sobre la ocurrencia del suceso.*

Para modelarla, necesitamos alguna escala, lo que se consigue con **un experimento de calibración**.

Para ello, nos basta con ser capaces de imaginar un experimento con resultados que encontremos igualmente verosímiles. Supongamos que se escoge un experimento con n resultados, por ejemplo, tirar un dado con 6 posibles resultados. Introduzco el siguiente mecanismo de apuestas: me pides que escoja uno de los n resultados, y me ofreces 100 euros si el experimento produce el resultado escogido; si encuentro las n apuestas posibles indiferentes entre sí, entonces encontraré los n resultados igualmente verosímiles y, **para mí**, el experimento propuesto será de calibración.

Definición

Un experimento es de calibración, para una persona, si encuentra todos los resultados igualmente verosímiles.

Potencialmente hay numerosos experimentos de calibración, como tirar un dado equilibrado; sacar bolas de una bolsa con bolas iguales; usar una rueda de ruleta; usar un generador de números aleatorios, etc.

Ejemplo

Supongamos que he tenido una discusión con mi pareja y necesito asignar (mi) la probabilidad de que al volver a casa esta noche la encuentre (es decir, el suceso de interés es $A =$ *Mi pareja estará en casa esta noche*). Como experimento de calibración considero el de sacar bolas iguales de una urna:

B	N

Considero las dos apuestas

$$\begin{cases} 10000 \text{ si sale blanco de la urna } (B_1) \\ 0 \text{ en otro caso} \end{cases}$$

$$\begin{cases} 10000 \text{ si está mi pareja } (A) \\ 0 \text{ en otro caso} \end{cases}$$

Así, si digo que prefiero la apuesta sobre A (se escribe como $A \succ B_1$), implícitamente estoy diciendo que, **para mí**, $P(A) \geq 1/2$.

Considero, a continuación, la siguiente urna:

B	B	B	N

y planteo las siguientes apuestas:

$$\begin{cases} 10000 \text{ si sale blanco de la urna } (B_2) \\ 0 \text{ en otro caso} \end{cases}$$

$$\begin{cases} 10000 \text{ si está mi pareja } (A) \\ 0 \text{ en otro caso} \end{cases}$$

Si digo que prefiero la apuesta sobre A (se escribe como $A \succ B_2$), implícitamente estoy diciendo que, **para mí**, $P(A) \geq 3/4$.

Considero, a continuación, la siguiente urna:

B	B	B	B	B	B	B	N

Si considero $B_3 \succ A$, implícitamente estoy diciendo que, **para mí**,

$$\frac{3}{4} \leq P(A) \leq \frac{7}{8}.$$

Se repite, entonces, el proceso con una urna con 13 bolas blancas y 3 negras; si, en este caso, encuentro que $B_4 \sim A$, estaría diciendo que **para mí**, $P(A) = 13/16$.

El método discutido tiene dos inconvenientes aparentemente. En primer lugar, las *apuestas* que se presentan son imaginarias, no tanto por el hecho de que no haya transación monetaria, sino porque muchas veces estaremos hablando de sucesos no observables (bien en ese instante, como cuando hablamos de quién ganará la liga en 2040, o nunca, cuando hablamos de que Aníbal cruzó los Alpes por San Gotardo). Sin embargo, cuando el problema es importante, la gente se implica en el problema de asignación.

En segundo lugar, puede haber un problema de *precisión*; por ejemplo, en cuanto pasamos a urnas con 16 bolas en el experimento que hemos usado, empezamos a tener problemas de discriminación de los grados de creencia. Sin embargo, en las aplicaciones no suele ser necesario requerir tanta finura en la asignación. Se han desarrollado, además, métodos de análisis de sensibilidad que permiten aliviar este problema.

1.5.7. Probabilidad condicionada e independencia de sucesos

A partir de ahora, supondremos que las probabilidades que se emplean son medidas de los grados de creencia de un individuo.

En numerosas ocasiones, tendremos que modelizar una situación en la que se dispone de información adicional, debiendo condicionarse a sucesos o circunstancias. De hecho, en Estadística este será el *letit-motiv* básico: debemos procesar información nueva, o lo que es lo mismo, condicionar a una nueva información. Formalmente, suponemos que estamos interesados en un suceso A; hemos asignado $P(A)$ y nos informan de que ha ocurrido B y queremos saber cómo cambian mis creencias sobre A.

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

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

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

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

1.5.8. Ejemplo

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

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

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

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

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

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

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

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

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

y entonces

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

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

1.5.9. Observación

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

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

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

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

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

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

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

1.6. Ciertos teoremas fundamentales del cálculo de probabilidades

Hay algunos resultados importantes del cálculo de probabilidades que son conocidos bajo los nombres de **teorema de la probabilidad compuesta**, **teorema de la probabilidad total** y **teorema de Bayes**. Veamos cuáles son estos teoremas, pero previamente vamos a enunciar, a modo de recopilación, una serie de resultados elementales:

1.6.1. Proposición

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

- Probabilidad de la unión de sucesos:

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

- Probabilidad de la intersección de sucesos:

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

- Probabilidad del suceso contrario:

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

- Probabilidad condicionada del suceso contrario:

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

1.6.2. Ejemplo

En una universidad el 50 % de los alumnos habla inglés, el 20 % francés y el 5 % los dos idiomas ¿Cuál es la probabilidad de encontrar alumnos que hablen alguna lengua extranjera?

Solución:

Sea A el suceso hablar inglés: $P(A) = 0{,}5$. Sea B el suceso hablar francés: $P(B) = 0{,}2$ y sea $A \cap B$ el suceso hablar francés e inglés: $P(A \cap B) = 0{,}05$. Así,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0{,}5 + 0{,}2 - 0{,}05 = 0{,}65.$$

1.6.3. Ejemplo

En una estación de esquí, para navidades, la experiencia indica que hay tiempo soleado sólo el 15 % de los días. Por otro lado, se ha calculado que cuando un día es soleado, hay una probabilidad del 20 % de que el día posterior también lo sea. Calcular la probabilidad de que, en navidades, un fin de semana completo sea soleado.

Solución: Llamemos S al suceso sábado soleado y D al suceso domingo soleado. La única manera en que un fin de semana completo sea soleado es que lo sea en primer lugar el sábado, y que el domingo posterior también. Es decir:

$$P(S \cap D) = P(S)P(D|S) = 0{,}15 \cdot 0{,}2 = 0{,}03.$$

Luego sólo el 3 % de los fines de semana son soleados.

El primero de los teoremas que vamos a enunciar es una generalización de la probabilidad de la intersección de dos sucesos a la de un número cualquiera, pero finito, de ellos.

En ocasiones, debemos calcular probabilidades, pero es complicado hacerlo directamente; a veces podemos condicionar a una partición, de manera que las probabilidades condicionadas y condicionantes son de cálculo sencillo, con lo que el cálculo original se facilita.

1.6.4. Teorema (Probabilidad compuesta)

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

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

Demostración:

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

Los teoremas que restan nos dicen cómo calcular las probabilidades de sucesos cuando tenemos que el suceso seguro está descompuesto en una serie de sucesos incompatibles de los que conocemos su probabilidad. Para ello necesitamos introducir un nuevo concepto: se dice que la colección $A_1, A_1, \ldots, A_n \subset E$ es un sistema exhaustivo y excluyente de sucesos si se verifica:

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

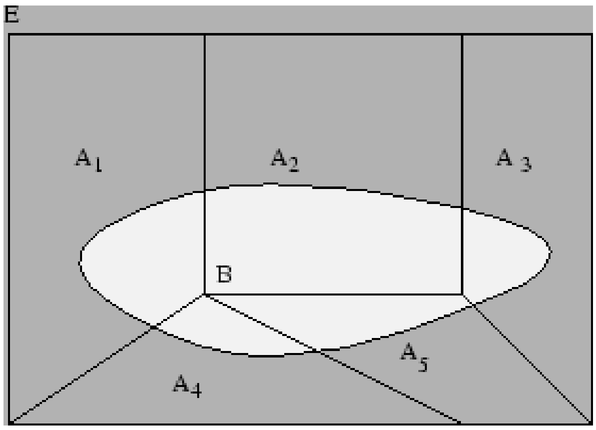
1.6.5. Teorema de la Probabilidad total

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

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

Demostración:

Observando la figura



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

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

1.6.6. Ejemplo

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

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

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

Solución: La situación que tenemos puede ser esquematizada como

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

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

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

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

1.6.7. Teorema de Bayes

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

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

Demostración:

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

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

1.6.8. Ejemplo

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

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

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

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

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

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

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

$0B$	$3R$	U_3	$P(U_3) = 1/3$	$P(B\|U_3) = 0$

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

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

Con respecto a las demás urnas es lo mismo:

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

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

1.6.9. Observación

Obsérvese que en el ejemplo anterior, antes de realizar el experimento aleatorio de extraer una bola para ver su resultado, teníamos que la probabilidad de elegir una urna i cualquiera es $P(U_i)$. Estas probabilidades se denominan probabilidades *a priori.* Sin embargo, después de realizar el experimento, y observar que el resultado del mismo ha sido la extracción de una bola blanca, las probabilidades de cada urna han cambiado a $P\left(U_i|B\right)$. Estas cantidades se denominan probabilidades *a posteriori.* Vamos a representar en una tabla la diferencia entre ambas:

a priori	**a posteriori**
$P(U_1) = 1/3$	$P(U_1\|B) = 9/19$
$P(U_2) = 1/3$	$P(U_2\|B) = 10/19$
$P(U_3) = 1/3$	$P(U_3\|B) = 0$

Las probabilidades *a priori* cambian de tal modo de las *a posteriori*, que una vez observado el resultado del experimento aleatorio, se puede afirmar con certeza que no fue elegida la tercera urna.

Puesto que los denominadores que aparecen en la fórmula de Bayes son comunes, se puede reescribir de la siguiente manera, para $i = 1, \ldots, n$ sucesos, donde $\propto$ es el signo *proporcional a*:

$$P(A_i|G) \propto P(G|A_i)P(A_i),$$

y después se puede normalizar por una constante.

Interpretación

En numerosas ocasiones se necesita disponer de un procedimiento para actualizar creencias, o, lo que es lo mismo, para introducir información nueva que recibimos y modificar nuestras probabilidades. Este es el problema básico de la Estadística.

El ingrediente básico para resolver este problema es la probabilidad condicionada y la forma de resolverlo es a través de la fórmula o regla de Bayes.

Para simplificar la discusión, supongamos que se contemplan dos hipótesis (o teorías o visiones del mundo) que denominamos A y A^c; partimos de unas creencias iniciales o *probabilidades a priori* $P(A)$ y $P(A^c) = 1 - P(A)$.

Se realiza un experimento con resultado G; dependiendo de qué hipótesis sea cierta, tendríamos unas probabilidades para esos resultados que serían $P(G|A)$ (la probabilidad de obtener G en el experimento supuesto que A es cierta) y $P(G|A^c)$, que se denominan verosimilitudes de A y A^c.

Puesto que hemos observado G, las nuevas probabilidades relevantes serán las probabilidades a posteriori $P(A|G)$ y $P(A^c|G)$, esto es,

$$P(A|G) = \frac{P(A \cap G)}{P(G)} = \frac{P(G|A)P(A)}{P(G|A)P(A) + P(G|A^c)P(A^c)}.$$

Factor Bayes

La regla de Bayes se puede expresar de otras maneras. Por ejemplo, si calculamos el cociente

$$\frac{P(A|G)}{P(A^c|G)} = \frac{P(G|A)}{P(G|A^c)} \cdot \frac{P(A)}{P(A^c)}.$$

El término

$$\frac{P(A|G)}{P(A^c|G)},$$

se denomina apuestas *a posteriori* a favor de A y

$$\frac{P(A)}{P(A^c)},$$

son las apuestas a favor de A.

Por otra parte,

$$\frac{P(G|A)}{P(G|A^c)},$$

es el *factor Bayes* a favor de A.

De este modo se puede escribir

$$\textit{Apuestas a posteriori} = \textit{Factor Bayes} \times \textit{Apuestas a priori}$$

Predicción

Una vez procesada la información mediante la fórmula de Bayes, la emplearemos para tomar decisiones o realizar predicciones.

El problema de predicción se decribe como sigue. Partiendo de las probabilidades a priori $P(A_i)$ y las verosimilitudes $P(G|A_i)$; hemos hecho una realización del experimento y obtenido G_1; hemos procesado esta información a $P(A_i|G_1)$. Deseamos predecir el resultado de una segunda realización del experimento; para ello emplearemos las probabilidades $P(G_2|G_1)$. Se observa que

$$\begin{aligned} P(G_2|G_1) &= \frac{P(G_2 \cap G_1)}{P(G_1)} = \frac{\sum_j P(G_2 \cap G_1 \cap A_j)}{P(G_1)} = \\ &= \frac{\sum_j P(G_2 \cap G_1 \cap A_j)}{P(A_j \cap G_1)} \cdot \frac{P(A_j \cap G_1)}{P(G_1)} = \\ &= \sum_j P(G_2|G_1 \cap A_j) \cdot P(A_j|G_1). \end{aligned}$$

En muchas ocasiones, si se conoce A_j, G_1 resulta irrelevante para predecir G_2, con lo que

$$P(G_2|G_1 \cap A_j) = P(G_2|A_j)$$

(esto es, G_1 y G_2 son condicionalmente independientes dados A_j), y resulta que

$$P(G_2|G_1) = \sum_j P(G_2|A_j) \cdot P(A_j|G_1).$$

Ejemplo

Tenemos una bolsa con 5 bolas con dos posibles colores, blanco y rojo. Se necesita tener una idea de cuántas bolas blancas hay. Sea j tal número. Este puede ser 0, 1, 2, 3, 4 ó 5. En ausencia de información suponemos que todos los valores son igualmente probables, esto es,

$$P(0) = P(1) = \cdots = P(5) = \frac{1}{6}.$$

Podemos realizar un experimento informativo consistente en sacar una bola y observar su color. En este experimento, si B designa *sacar bola blanca*, se tiene que

$P(B|0) = 0$	$P(B|1) = 1/5$	$P(B|2) = 2/5$	$\cdots$	$P(B|5) = 1.$

Se realiza el experimento y se obtiene bola blanca. Las probabilidades de interés son $P(j|B)$. Se tiene que

$$P(j|B) = \frac{P(j \cap B)}{P(B)} = \frac{P(B|j)P(j)}{\sum_i P(B|i)P(i)}.$$

Por ejemplo,

$$P(4|B) = \frac{4/5 \cdot 1/6}{0 \cdot 1/6 + 1/5 \cdot 1/6 + \cdots + 1 \cdot 1/6}.$$

Se obtiene la siguiente tabla:

j	0	1	2	3	4	5
$P(j)$	1/6	1/6	1/6	1/6	1/6	1/6
$P(j\|B)$	0	1/15	2/15	3/15	4/15	5/15

Así, la composición 5 pasa a ser la más probable.

A continuación, deseamos saber las probabilidades de sacar otra bola blanca, esto es, deseamos calcular $P(B_2|B_1)$, si la extracción se hace *con reemplazamiento*.

En este ejemplo, si sabemos j (la composición de la bolsa), B_1 no aporta información para predecir B_2, por lo que

$$P(B_2|j \cap B_1) = P(B_2|j).$$

Por ser con reemplazamiento, $P(B_2|j) = P(B_1|j)$, como antes, esto es, $P(B_2|j) = \frac{j}{5}$. Así,

$$\begin{aligned} P(B_2|B_1) &= P(B_2|0) \cdot P(0|B_1) + \cdots + P(B_2|5) \cdot P(5|B_1) = \\ &= 0 \cdot 0 + \frac{1}{5} \cdot \frac{1}{15} + \cdots + 1 \cdot \frac{5}{15} = \frac{11}{15}. \end{aligned}$$

Se obtiene fácilmente que

$$P(B_2) = \sum_j P(B_2|j) \cdot P(j) = \frac{1}{2}.$$

Como

$$P(B_2) \neq P(B_2|B_1),$$

B_2 y B_1 son sucesos *dependientes*, más aún, como $P(B_2|B_1) > P(B_2)$, están relacionados positivamente.