Probabilidad y Estadística (Borradores, Curso 23)
Espacios de Probabilidad
Elementos de Análisis Combinatorio

Sebastian Grynberg

9-23 de marzo 2011

Andrei Nikolaevich Kolmogorov (1903-1987)
Estableció los fundamentos de la Teoría de Probabilidad en 1933

"se aprende a pensar abstractamente
mediante el pensamiento abstracto."

(G.W.F. Hegel)

Índice

1. Teoría elemental

1.1. Los axiomas de Kolmogorov

Sean Ω un conjunto no vacío cuyos elementos ω serán llamados *eventos elementales* y $\mathcal{A}$ una familia de subconjuntos de Ω que serán llamados *eventos*.

Definición 1.1. $\mathcal{A}$ es un *álgebra de eventos* si contiene a Ω y es cerrada por complementos y uniones finitas[1] :

(i) $\Omega \in \mathcal{A}$;

(ii) $A \in \mathcal{A}$ implica $A^c \in \mathcal{A}$;

(iii) $A, B \in \mathcal{A}$ implica $A \cup B \in \mathcal{A}$.

Definición 1.2. Una *medida de probabilidad* $\mathbb{P}$ sobre $(\Omega, \mathcal{A})$ es una función $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ que satisface los axiomas siguientes:

I. Para cada $A \in \mathcal{A}$, $\mathbb{P}(A) \geq 0$;

II. $\mathbb{P}(\Omega) = 1$;

III. Si los eventos A y B no tienen elementos en común, entonces

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Definición 1.3. Un *espacio de probabilidad* es una terna $(\Omega, \mathcal{A}, \mathbb{P})$ formada por un conjunto no vacío Ω, llamado *el espacio muestral*; un álgebra $\mathcal{A}$ de subconjuntos de Ω; llamados *los eventos aleatorios*; y una medida de probabilidad $\mathbb{P}$ definida sobre los eventos aleatorios.

---

[1]**Nomenclatura y definiciones previas.** Sean A y B eventos.

1. Escribiremos A^c para designar al evento que no ocurre A:

$$A^c := \{\omega \in \Omega : \omega \notin A\}.$$

   El evento A^c se llama el *complemento* de A.

2. Escribiremos $A \cup B$ para designar al evento que ocurre al menos uno de los eventos A o B:

$$A \cup B := \{\omega \in \Omega : \omega \in A \text{ o } \omega \in B\}.$$

   El evento $A \cup B$ se llama la *unión* de A y B.

3. Escribiremos $A \cap B$ para designar al evento ocurren ambos A y B:

$$A \cap B := \{\omega \in \Omega : \omega \in A \text{ y } \omega \in B\}.$$

   El evento $A \cap B$ se llama la *intersección* de A y B.

A veces escribiremos $A \setminus B$ en lugar de $A \cap B^c$, esto es, el evento que A ocurre, pero B no lo hace. Cuando dos eventos A y B no tienen elementos en común, esto es $A \cap B = \emptyset$, diremos que A y B son *disjuntos*. Una colección de eventos $A_1, A_2, \dots$ se dice *disjunta dos a dos*, si $A_i \cap A_j = \emptyset$ para todo $i \neq j$.

Nota Bene (Consistencia). El sistema de axiomas I-III es *consistente*. Esto se prueba mediante un ejemplo. Sea Ω un conjunto que consiste de un solo elemento y sea $\mathcal{A} = \{\emptyset, \Omega\}$ la familia de todos los subconjuntos de Ω. $\mathcal{A}$ es un álgebra y la función $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ definida por $\mathbb{P}(\Omega) := 1$ y $\mathbb{P}(\emptyset) := 0$ es una medida de probabilidad. $\square$

Construcción de espacios de probabilidad finitos. Los espacios de probabilidad más simples se construyen de la siguiente manera. Se considera un conjunto finito Ω y una función $p : \Omega \to [0, 1]$ tal que

$$\sum_{\omega\in\Omega} p(\omega) = 1.$$

La función p se llama *función de probabilidad* y los números $p(\omega)$, $\omega \in \Omega$, se llaman las *probabilidades de los eventos elementales* $\omega \in \Omega$ o simplemente las *probabilidades elementales*.

El álgebra de eventos, $\mathcal{A}$, se toma como el conjunto de todos los subconjuntos de Ω y para cada $A \in \mathcal{A}$ se define

$$\mathbb{P}(A) := \sum_{\omega\in A} p(\omega),$$

donde la suma vacía se define como 0.

Todos los espacios de probabilidad finitos en los que $\mathcal{A}$ es la familia de todos los subconjuntos de Ω se construyen de esta manera.

Ejemplo 1.4 (Lanzar una moneda equilibrada)**.** Se lanza una moneda. Los resultados posibles son cara o ceca y pueden representarse mediante las letras H (*head*) y T (*tail*). Adoptando esa representación el espacio muestral correspondiente es

$$\Omega = \{H, T\}.$$

Decir que una moneda es equilibrada significa que la función de probabilidad asigna igual probabilidad a los dos resultados posibles:

$$p(H) = p(T) = 1/2.$$

$\square$

Ejemplo 1.5 (Lanzar un dado equilibrado)**.** Cuando se lanza un dado, la elección obvia del espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$, y la función probabilidad debe estar dada por $p(i) = 1/6$, $i = 1, \dots, 6$.

Sean A y B los eventos definidos por "*el resultado es impar*" y "*el resultado es a lo sumo* 4", respectivamente. Entonces $A = \{1, 3, 5\}$, $B = \{1, 2, 3, 4\}$ y sus probabilidades son

$$\mathbb{P}(A) = p(1) + p(3) + p(5) = \frac{3}{6}, \qquad \mathbb{P}(B) = p(1) + p(2) + p(3) + p(4) = \frac{4}{6}.$$

Las probabilidades de los eventos $A^c = \{2, 4, 6\}$, $A \cup B = \{1, 2, 3, 4, 5\}$ y $A \cap B = \{1, 3\}$ son

$$\mathbb{P}(A^c) = \frac{3}{6}, \qquad \mathbb{P}(A \cup B) = \frac{5}{6}, \qquad \mathbb{P}(A \cap B) = \frac{2}{6}.$$

Notar que

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A), \qquad \text{y} \qquad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

□

1.2. Relación con los datos experimentales

La teoría de probabilidad se aplica al mundo real de los experimentos de la manera siguiente:

(1) Consideramos un sistema de condiciones, $\mathcal{S}$, que pueden repetirse cualquier cantidad de veces.

(2) Estudiamos una familia determinada de eventos que pueden ocurrir como resultado de realizar las condiciones $\mathcal{S}$. En los casos individuales donde se realizan las condiciones $\mathcal{S}$, los eventos ocurren, generalmente, de distintas maneras. En el conjunto Ω incluimos, *a priori*, todos los resultados que podrían obtenerse al realizar las condiciones $\mathcal{S}$.

(3) Si al realizar las condiciones $\mathcal{S}$ el resultado pertenece al conjunto A (definido de alguna manera), diremos que ocurre el evento A.

Ejemplo 1.6 (Dos monedas)**.** Las condiciones $\mathcal{S}$ consisten en lanzar una moneda dos veces. El conjunto de los eventos mencionados en (2) resultan del hecho de que en cada lanzamiento puede obtenerse una cara (H) o una ceca (T). De aquí resulta que solamente son posibles cuatro resultados (los eventos elementales), a saber: HH, HT, TH, TT. Si el evento A se define por la ocurrencia de una repetición, entonces A consistirá en que suceda el primero o el cuarto de los cuatro eventos elementales. Esto es, $A = \{HH, TT\}$. De esta manera, todo evento puede considerarse como un conjunto de eventos elementales. □

(4) Bajo ciertas condiciones se puede suponer que, dado el sistema de condiciones $\mathcal{S}$, un evento A que a veces ocurre y a veces no, tiene asignado un número real $\mathbb{P}(A)$ que tiene las características siguientes:

(a) Se puede estar prácticamente seguro que si el sistema de condiciones $\mathcal{S}$ se repite un gran número de veces, n, entonces si $n(A)$ es la cantidad de veces que ocurre el evento A, la proporción $n(A)/n$ diferirá muy poco de $\mathbb{P}(A)$.

(b) Si $\mathbb{P}(A)$ es muy pequeña, se puede estar prácticamente seguro de que cuando se realicen las condiciones $\mathcal{S}$ solo una vez, el evento A no ocurrirá.

Deducción empírica de los axiomas. En general, se puede suponer que la familia $\mathcal{A}$ de los eventos observados $A, B, C, \dots$ que tienen probabilidades asignadas, constituye un álgebra de eventos. Está claro que $0 \leq n(A)/n \leq 1$ de modo que el axioma I es bastante natural. Para el evento Ω, $n(\Omega)$ siempre es igual a n de modo que es natural definir $\mathbb{P}(\Omega) = 1$ (Axioma II). Si finalmente, A y B son incompatibles (i.e., no tienen elementos en común), entonces $n(A \cup B) = n(A) + n(B)$ y de aquí resulta que

$$\frac{n(A \cup B)}{n} = \frac{n(A)}{n} + \frac{n(B)}{n}.$$

Por lo tanto, es apropiado postular que $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (Axioma III).

Nota Bene 1. La afirmación de que un evento A ocurre en las condiciones $\mathcal{S}$ con una determinada probabilidad $\mathbb{P}(A)$ equivale a decir que en una serie suficientemente larga de experimentos (es decir, de realizaciones del sistema de condiciones $\mathcal{S}$), las frecuencias relativas

$$\hat{p}_k(A) := \frac{n_k(A)}{n_k}$$

de ocurrencia del evento A (donde n_k es la cantidad de experimentos realizados en la k-ésima serie y $n_k(A)$ la cantidad de ellos en los que ocurre A) son aproximadamente idénticas unas a otras y están próximas a $\mathbb{P}(A)$. □

Ejemplo 1.7. Las condiciones $\mathcal{S}$ consisten en lanzar una moneda (posiblemente cargada). Podemos poner $\Omega = \{H, T\}$ y $\mathcal{A} = \{\emptyset, H, T, \Omega\}$, y las posibles medidas de probabilidad $\mathbb{P} : \mathcal{A} \to [0, 1]$ están dadas por

$$\mathbb{P}(\emptyset) = 0, \qquad \mathbb{P}(H) = p, \qquad \mathbb{P}(T) = 1 - p, \qquad \mathbb{P}(\Omega) = 1,$$

donde p es un número real fijo perteneciente al intervalo $[0, 1]$.

Si en 10 series, de 1000 lanzamientos cada una, se obtienen las siguientes frecuencias relativas de ocurrencia del evento $A = \{H\}$

$$0.753;\ \ 0.757;\ \ 0.756;\ \ 0.750;\ \ 0.746;\ \ 0.758;\ \ 0.751;\ \ 0.748;\ \ 0.749;\ \ 0.746, \tag{1}$$

parece razonable asignarle a p el valor 0.75. □

Nota Bene 2. Si cada una de dos afirmaciones diferentes es prácticamente segura, entonces podemos decir que simultáneamente son ambas seguras, aunque el grado de seguridad haya disminuido un poco. Si, en cambio, el número de tales afirmaciones es muy grande, de la seguridad práctica de cada una, no podemos deducir nada sobre la validez simultánea de todos ellas. En consecuencia, del principio enunciado en (a) no se deduce que en una cantidad muy grande de series de n experimentos cada una, en *cada uno de ellos* la proporción $n(A)/n$ diferirá sólo un poco de $\mathbb{P}(A)$.

En los casos más típicos de la teoría de probabilidades, la situación es tal que en una larga serie de pruebas es posible obtener uno de los dos valores extremos para la frecuencia

$$\frac{n(A)}{n} = \frac{n}{n} = 1 \qquad \text{y} \qquad \frac{n(A)}{n} = \frac{0}{n} = 0.$$

Así, cualquiera sea el número de ensayos n, es imposible asegurar con absoluta certeza que tendremos, por ejemplo, la desigualdad

$$\left|\frac{n(A)}{n} - \mathbb{P}(A)\right| < \frac{1}{10}.$$

Por ejemplo, si el evento A es sacar un seis con un dado, entonces en n tiradas la probabilidad de obtener un seis en cada una de ellas es $(1/6)^n > 0$; en otras palabras, con probabilidad $(1/6)^n$ tendremos una frecuencia relativa de sacar seis igual a *uno*; y con probabilidad $(5/6)^n$ no saldrá un seis en ninguna, es decir, la frecuencia relativa de sacar seis será igual a *cero*. □

Nota Bene 3. De acuerdo con nuestros axiomas a un evento imposible (un conjunto vacío) le corresponde la probabilidad $\mathbb{P}(\emptyset) = 0$, pero la recíproca no es cierta: $\mathbb{P}(A) = 0$ no implica la imposibilidad de A. Cuando $\mathbb{P}(A) = 0$, del principio (b) todo lo que podemos asegurar es que cuando se realicen las condiciones $\mathcal{S}$ una sola vez, el evento A será prácticamente imposible. Sin embargo, esto no asegura de ningún modo que en una sucesión suficientemente grande de experimentos el evento A no ocurrirá. Por otra parte, del principio (a) solamente se puede deducir que cuando $\mathbb{P}(A) = 0$ y n es muy grande, la proporción $n(A)/n$ debe ser muy pequeña (por ejemplo, $1/n$). $\square$

1.3. Corolarios inmediatos de los axiomas

De $A \cup A^c = \Omega$ y los axiomas II y III se deduce que

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A). \tag{2}$$

En particular, debido a que $\Omega^c = \emptyset$, tenemos que $\mathbb{P}(\emptyset) = 0$.

Teorema de aditividad. Si los eventos $A_1, A_2, \dots, A_n$ son disjuntos dos a dos, entonces del axioma III se deduce la fórmula

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i). \tag{3}$$

Ejercicios adicionales

1. Sean A y B dos eventos. Mostrar que

(**a**) Si $A \subseteq B$, entonces $\mathbb{P}(A) \leq \mathbb{P}(B)$.
Más precisamente, se tiene que $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$.
Sugerencia. Expresar el evento B como la unión disjunta de los eventos A y $B \setminus A$ y usar el axioma III.

(**b**) La probabilidad de que ocurra al menos uno de los eventos A o B es

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

Sugerencia. La unión $A \cup B$ de dos eventos puede expresarse como la unión de dos eventos disjuntos: $A \cup (B \setminus (A \cap B))$.

2. Mostrar que para eventos A, B y C vale que

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

3. Mostrar que para eventos $A_1, A_2, \dots, A_n$ vale que

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_j) - \cdots$$
$$+(-1)^n \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n).$$

1.4. Probabilidad condicional

Definición 1.8 (Probabilidad condicional)**.** Si $A \subset \Omega$ es un evento de probabilidad positiva, la *probabilidad condicional del evento B dado que ocurrió el evento A* se define como el cociente

$$\mathbb{P}(B|A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}. \tag{4}$$

Nota Bene (La probabilidad condicional induce una medida de probabilidad sobre los eventos aleatorios). Sea $A \subset \Omega$ un evento de probabilidad positiva, arbitrario pero fijo. Es fácil ver que valen las siguientes propiedades:

1. Para cada $B \in \mathcal{A}$, $\mathbb{P}(B|A) \geq 0$;
2. $\mathbb{P}(\Omega|A) = 1$;
3. Si los eventos B y C no tienen elementos en común, entonces

$$\mathbb{P}(B \cup C|A) = \mathbb{P}(B|A) + \mathbb{P}(C|A).$$

Comparando las propiedades 1-3 con los axiomas I-III, encontramos que la función $\mathbb{P}(\cdot|A) : \mathcal{A} \to \mathbb{R}$ es una medida de probabilidad sobre los eventos aleatorios. Por lo tanto, todos los resultados generales referidos a la propiedades de $\mathbb{P}(\cdot)$ también valen para la probabilidad condicional $\mathbb{P}(\cdot|A)$. $\square$

Ejemplo 1.9. Se lanza un dado, ¿cuál es la probabilidad condicional de observar un 3, condicionada a que el resultado del dado es a lo sumo 4? Denotando mediante A al evento "*el resultado es a lo sumo* 4" y mediante B el evento "*el resultado es* 3". Tenemos que $\mathbb{P}(A) = 4/6$, $\mathbb{P}(B) = 1/6$ y $\mathbb{P}(A \cap B) = \mathbb{P}(A) = 1/6$. Así

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{1/6}{4/6} = \frac{1}{4},$$

lo que intuitivamente tiene sentido (¿por qué?). $\square$

Probabilidad compuesta. De la definición de la probabilidad condicional del evento B dado que ocurrió el evento A resulta inmediatamente la siguiente fórmula

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A). \tag{5}$$

denominada *regla del producto*.

El siguiente Teorema generaliza la regla del producto (5) y se obtiene por inducción.

Teorema 1.10 (Regla del producto)**.** Suponiendo que todos los eventos condicionantes tienen probabilidad positiva, tenemos que

$$\mathbb{P}\left(\cap_{i=1}^{n} A_i\right) = \mathbb{P}\left(A_n \left| \cap_{i=1}^{n-1} A_i\right.\right) \cdots \mathbb{P}\left(A_3 | A_1 \cap A_2\right) \mathbb{P}(A_2|A_1)\mathbb{P}(A_1). \tag{6}$$

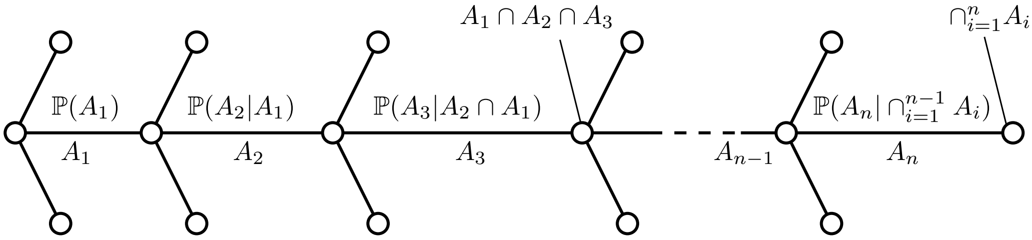


Figura 1: Ilustración de la regla del producto: el evento $\cap_{i=1}^{n} A_i$ tiene asociada una única trayectoria sobre un árbol que describe la historia de un experimento aleatorio realizado por etapas sucesivas. Las aristas de esta trayectoria corresponden a la ocurrencia sucesiva de los eventos $A_1, A_2, \ldots, A_n$ y sobre ellas registramos la correspondiente probabilidad condicional. El nodo final de la trayectoria corresponde al evento $\cap_{i=1}^{n} A_i$ y su probabilidad se obtiene multiplicando las probabilidades condicionales registradas a lo largo de las aristas de la trayectoria: $\mathbb{P}(\cap_{i=1}^{n} A_i) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_2 \cap A_1) \cdots \mathbb{P}(A_n| \cap_{i=1}^{n-1} A_i)$. Notar que cada nodo intermedio a lo largo de la trayectoria también corresponde a un evento intersección y su probabilidad se obtiene multiplicando las probabilidades condicionales registradas desde el inicio de la trayectoria hasta llegar al nodo. Por ejemplo, el evento $A_1 \cap A_2 \cap A_3$ corresponde al nodo indicado en la figura y su probabilidad es $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)$.

Ejemplo 1.11. Una urna contiene 5 bolas rojas y 10 bolas negras. Se extraen dos bolas al azar sin reposición. ¿Cuál es la probabilidad que ambas bolas sean negras?

Sean N_1 y N_2 los eventos definidos por "*la primer bola extraída es negra*" y "*la segunda bola extraída es negra*", respectivamente. Claramente $\mathbb{P}(N_1) = 10/15$. Para calcular $\mathbb{P}(N_2|N_1)$ observamos que si ocurrió N_1, entonces solo 9 de las 14 bolas restantes en la urna son negras. Así $\mathbb{P}(N_2|N_1) = 9/14$ y

$$\mathbb{P}(N_2 \cap N_1) = \mathbb{P}(N_2|N_1)\mathbb{P}(N_1) = \frac{10}{15} \cdot \frac{9}{14} = \frac{3}{7}.$$

$\square$

Teorema 1.12 (Fórmula de probabilidad total)**.** Sean $A_1, A_2, \dots, A_n$ eventos incompatibles dos a dos y tales que $\cup_{i=1}^n A_i = \Omega$. Para cada $B \in \mathcal{A}$ vale la siguiente fórmula

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i). \tag{7}$$

denominada *fórmula de probabilidad total.*

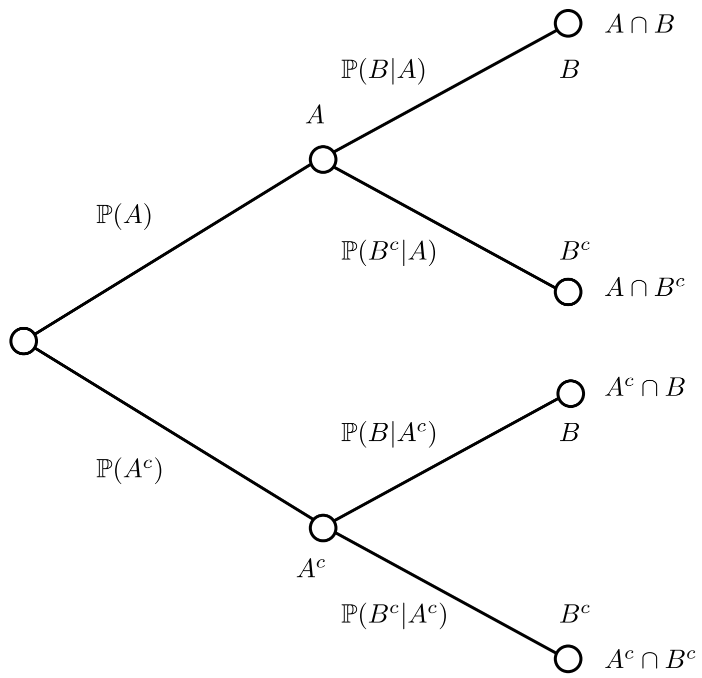


Figura 2: Ilustración de la fórmula de probabilidad total: *un experimento de dos etapas binarias y su correspondiente diagrama de árbol.* La primera ramificación (de izquierda a derecha) se basa en el resultado de la primer etapa del experimento (A o A^c) y la segunda en su resultado final (B o B^c). Multiplicando las probabilidades registradas a lo largo de cada trayectoria se obtiene la probabilidad del evento intersección representado por el nodo final. Sumando las probabilidades de las trayectorias que corresponden al evento B se obtiene: $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)$.

Nota Bene: Cálculo mediante condicionales. Si se dispone de una colección de eventos $A_1, A_2, \dots, A_n$ de los cuales uno y solamente uno debe ocurrir, la fórmula de probabilidad total (7) permite calcular la probabilidad de cualquier evento B condicionando a saber cuál de los eventos A_i ocurrió. Más precisamente, la fórmula (7) establece que la probabilidad $\mathbb{P}(B)$ es igual al promedio ponderado de las probabilidades condicionales $\mathbb{P}(B|A_i)$ donde cada término se pondera por la probabilidad del evento sobre el que se condicionó. Esta fórmula es útil debido a que a veces es más fácil evaluar las probabilidades condicionales $\mathbb{P}(B|A_i)$ que calcular directamente la probabilidad $\mathbb{P}(B)$. $\square$

Ejemplo 1.13 (Experimentos de dos etapas)**.** La primera etapa del experimento produce una partición $A_1, A_2, \dots, A_n$ del espacio muestral Ω. La segunda etapa produce el evento B. La fórmula (7) se utiliza para calcular la probabilidad de B.

Ejemplo 1.14. Una urna contiene 5 bolas rojas y 10 bolas negras. Se extraen dos bolas sin reposición. ¿Cuál es la probabilidad que la segunda bola sea negra?

El espacio muestral de este experimento aleatorio se puede representar como las trayectorias a lo largo de un árbol como se muestra en la Figura 3.

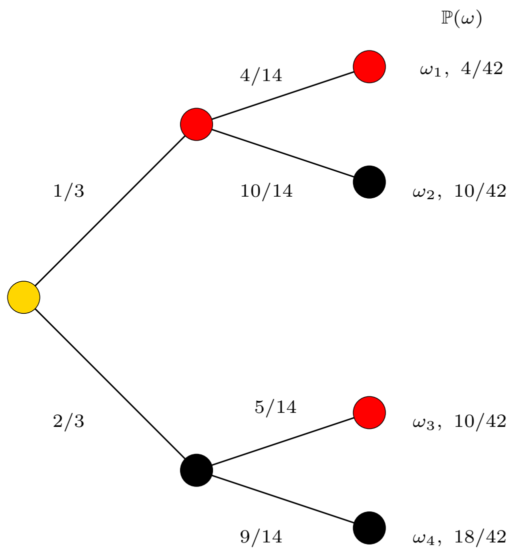


Figura 3: Observando el árbol se deduce que la probabilidad de que la segunda bola sea negra es: $\frac{1}{3} \cdot \frac{10}{14} + \frac{2}{3} \cdot \frac{9}{14} = \frac{2}{3}$.

Formalmente, el problema se resuelve mediante la fórmula de probabilidad total. Sean N_i y R_i los eventos definidos por "*la i-ésima bola extraída es negra*" y "*la i-ésima bola extraída es roja*", respectivamente $(i = 1, 2)$. Vale que

$$\mathbb{P}(N_1) = \frac{10}{15}, \qquad \mathbb{P}(R_1) = \frac{5}{15}, \qquad \mathbb{P}(N_2|R_1) = \frac{10}{14}, \qquad \mathbb{P}(N_2|N_1) = \frac{9}{14}.$$

Usando la fórmula de probabilidad total obtenemos

$$\begin{aligned} \mathbb{P}(N_2) &= \mathbb{P}(N_2 \cap R_1) + \mathbb{P}(N_2 \cap N_1) \\ &= \mathbb{P}(N_2|R_1)\mathbb{P}(R_1) + \mathbb{P}(N_2|N_1)\mathbb{P}(N_1) \\ &= \frac{10}{14} \cdot \frac{1}{3} + \frac{9}{14} \cdot \frac{2}{3} = \frac{28}{42} = \frac{2}{3}. \end{aligned}$$

$\square$

Primera versión de la regla de Bayes. Sean A y B dos eventos de probabilidad positiva. De la regla del producto (5) y su análoga $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$ obtenemos la siguiente fórmula importante

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}, \tag{8}$$

que contiene, en esencia, el Teorema de Bayes.

Ejemplo 1.15. Un test de sangre es 95 % efectivo para detectar una enfermedad cuando una persona realmente la padece. Sin embargo, el test también produce un "falso positivo" en el 1 % de las personas saludables testeadas. Si el 0, 5 % de la población padece la enfermedad, cuál es la probabilidad de que una persona tenga la enfermedad si su test resultó positivo?

Sea A el evento definido por *"la persona testeada tiene la enfermedad"* y sea B el evento definido por *"el resultado de su test es positivo"*. La probabilidad que nos interesa es $\mathbb{P}(A|B)$ y se puede calcular de la siguiente manera. Sabemos que

$$\mathbb{P}(A) = 0.005, \qquad \mathbb{P}(A^c) = 0.995,$$

$$\mathbb{P}(B|A) = 0.95, \qquad \mathbb{P}(B|A^c) = 0.01,$$

y usando esa información queremos calcular

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

El numerador, $\mathbb{P}(A \cap B)$, se puede calcular mediante la regla del producto

$$\mathbb{P}(A \cap B) \;=\; \mathbb{P}(B|A)\mathbb{P}(A) = (0.95)(0.005)$$

y el denominador, $\mathbb{P}(B)$, se puede calcular usando la fórmula de probabilidad total

$$\mathbb{P}(B) \;=\; \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c) = (0.95)(0.005) + (0.01)(0.995).$$

Por lo tanto,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)} = \frac{95}{294} \approx 0.323.$$

En otras palabras, sólo el 32 % de aquellas personas cuyo test resultó positivo realmente tienen la enfermedad. $\square$

Teorema 1.16 (Bayes)**.** Sean $A_1, A_2, \ldots, A_n$ eventos incompatibles dos a dos y tales que $\cup_{i=1}^n A_i = \Omega$. Sea B un evento de probabilidad positiva. Entonces,

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\sum_{i=1}^n \mathbb{P}(A_i)\mathbb{P}(B|A_i)}, \qquad i = 1, 2, \ldots, n. \tag{9}$$

Si los eventos $A_1, A_2, \ldots, A_n$ se llaman "hipótesis", la fórmula (9) se considera como la probabilidad de la hipótesis A_i después de la ocurrencia del evento B. En tal caso, $\mathbb{P}(A_i)$ es la probabilidad *a priori* de la hipótesis A_i y la fórmula (9) para $\mathbb{P}(A_i|B)$ se llama la *regla de Bayes para la probabilidad a posteriori* de la hipótesis A_i.

Nota Bene. Advertimos al lector que no trate de memorizar la fórmula (9). Matemáticamente, solo se trata de una forma especial de escribir la fórmula (8) y de nada más. $\square$

Ejemplo 1.17 (Canal de comunicación binario)**.** Un canal de comunicación binario simple transporta mensajes usando solo dos señales: 0 y 1. Supongamos que en un canal de comunicación binario dado el 40 % de las veces se transmite un 1; que si se transmitió un 0 la probabilidad de recibirlo correctamente es 0.90; y que si se transmitió un 1 la probabilidad de recibirlo correctamente es 0.95. Queremos determinar

(a) la probabilidad de recibir un 1;

(b) dado que se recibió un 1, la probabilidad de que haya sido transmitido un 1;

Solución. Consideramos los eventos A="*se transmitió un* 1" y B="*se recibió un* 1". La información dada en el enunciado del problema significa que $\mathbb{P}(A) = 0.4$, $\mathbb{P}(A^c) = 0.6$, $\mathbb{P}(B|A) = 0.95$, $\mathbb{P}(B|A^c) = 0.1$, $\mathbb{P}(B^c|A) = 0.05$, $P(B^c|A^c) = 0.90$ y puede representarse en la forma de un diagrama de árbol tal como se indicó en la sección 1.4.

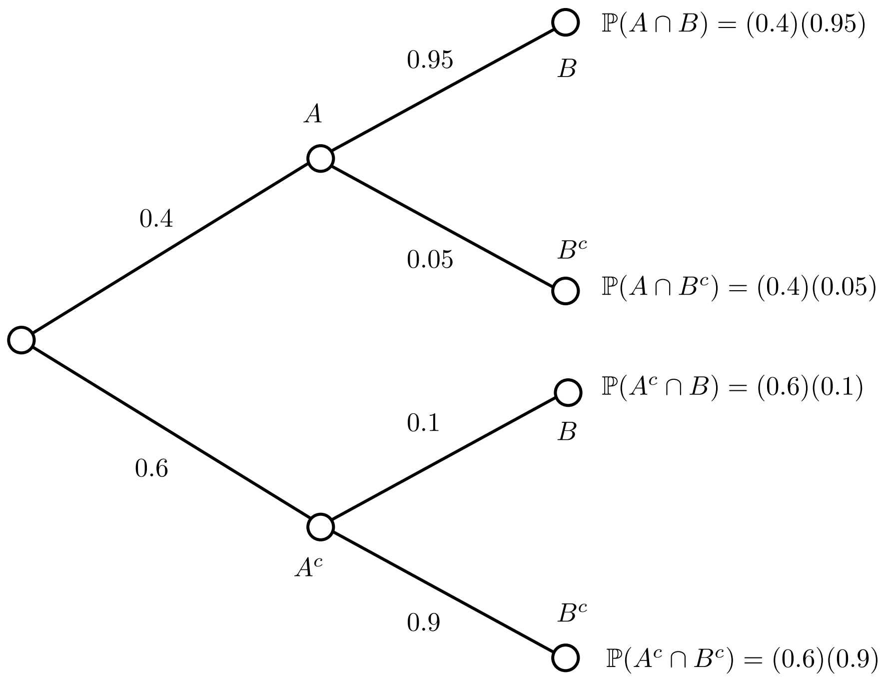


Figura 4: Observando el árbol se deduce que la probabilidad de recibir un 1 es $\mathbb{P}(B) = (0.4)(0.95) + (0.6)(0.1) = 0.44$. También se deduce que la probabilidad de que haya sido transmitido un 1 dado que se recibió un 1 es $\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{(0.4)(0.95)}{0.44} = 0.863...$ $\square$

Ejercicios adicionales

4. Los dados de Efron. Se trata de cuatro dados A, B, C, D como los que se muestran en la Figura 5.

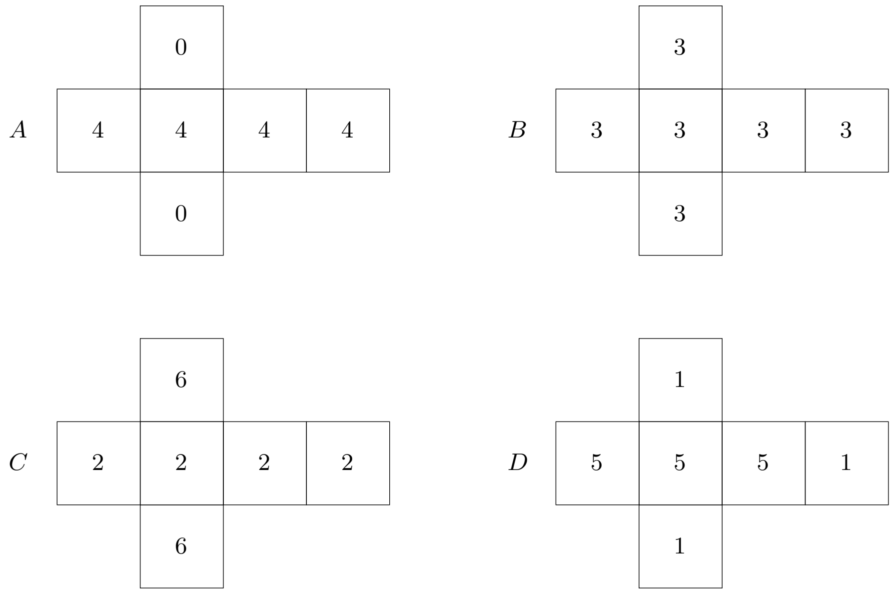


Figura 5: Dados de Efron

Las reglas del juego son las siguientes: juegan dos jugadores, cada jugador elige un dado, se tiran los dados y gana el que obtiene el número más grande.

(**a**) Calcular las siguientes probabilidades: que A le gane a B; que B le gane a C; que C le gane a D; que D le gane a A.

(**b**) ¿Cuál es la mejor estrategia para jugar con los dados de Efron?.

(**c**) Lucas y Monk jugaran con los dados de Efron eligiendo los dados al azar. Calcular las siguientes probabilidades:

- que Lucas pierda la partida si Monk obtiene un 3,
- que Lucas gane la partida si le toca el dado A.

(**d**) ¿Qué ocurre con el juego cuando los dados se eligen al azar?

(**e**) ¿Qué ocurre con el juego si a un jugador se le permite elegir un dado y el otro debe elegir al azar uno entre los restantes tres?

(**f**) Lucas y Monk jugaron con los dados de Efron, eligiendo los dados al azar. Lucas ganó, ¿cuál es la probabilidad de que le haya tocado el dado C?

1.5. Independencia

Definición 1.18 (Independencia)**.** Los eventos $A_1, A_2, \ldots, A_n$ son *mutuamente independientes* si satisfacen las siguientes $2^n - n - 1$ ecuaciones:

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_m}), \tag{10}$$

donde $m = 1, 2, \ldots, n$, y $1 \leq i_1 < i_2 < \ldots < i_m \leq n$.

Nota Bene 1. Para $n = 2$ el sistema de ecuaciones (10) se reduce a una condición: dos eventos A_1 y A_2 son independientes si satisfacen la ecuación

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2). \tag{11}$$

Ejemplo 1.19.

(a) Se extrae un naipe al azar de un mazo de naipes de poker. Por razones de simetría esperamos que los eventos "*corazón*" y "*As*" sean independientes. En todo caso, sus probabilidades son $1/4$ y $1/13$, respectivamente y la probabilidad de su realización simultánea es $1/52$.

(b) Se arrojan dos dados. Los eventos "*as en el primer dado*" y "*par en el segundo*" son independientes pues la probabilidad de su realización simultánea, $3/36 = 1/12$, es el producto de sus probabilidades respectivas: $1/6$ y $1/2$.

(c) En una permutación aleatoria de las cuatro letras a, b, c, d los eventos "*a precede a b*" y "*c precede a d*" son independientes. Esto es intuitivamente claro y fácil de verificar. $\square$

Nota Bene 2. Para $n > 2$, los eventos $A_1, A_2, \ldots, A_n$ pueden ser independientes de a pares: $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$, $1 \leq i < j \leq n$, pero no ser mutuamente independientes.

Ejemplo 1.20. Sea Ω un conjunto formado por cuatro elementos: $\omega_1, \omega_2, \omega_3, \omega_4$; las correspondientes probabilidades elementales son todas iguales a $1/4$. Consideramos tres eventos:

$$A_1 = \{\omega_1, \omega_2\}, \qquad A_2 = \{\omega_1, \omega_3\}, \qquad A_3 = \{\omega_1, \omega_4\}.$$

Es fácil ver que los eventos A_1, A_2, A_3 son independientes de a pares, pero no son mutuamente independientes:

$$\begin{gathered} \mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = 1/2, \\ \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_2 \cap A_3) = 1/4 = (1/2)^2, \\ \mathbb{P}(A_1 \cap A_2 \cap A_3) = 1/4 \neq (1/2)^3. \end{gathered}$$

$\square$

Independencia y probabilidades condicionales. En la introducción del concepto de independencia no utilizamos el de probabilidad condicional. Sin embargo, sus aplicaciones dependen generalmente de las propiedades de ciertas probabilidades condicionales.

Para fijar ideas, supongamos que $n = 2$ y que las probabilidades de los eventos A_1 y A_2 son positiva. En tal caso, los eventos A_1 y A_2 son independientes si y solamente si

$$\mathbb{P}(A_2|A_1) = \mathbb{P}(A_2) \qquad \text{y} \qquad \mathbb{P}(A_1|A_2) = \mathbb{P}(A_1).$$

El siguiente Teorema expresa la relación general entre el concepto de independencia y las probabilidades condicionales.

Teorema 1.21. Si todas las probabilidades $\mathbb{P}(A_i)$ son positivas, entonces una condición necesaria y suficiente para la mutua independencia de los eventos $A_1, A_2, \dots, A_n$ es la satisfacción de las ecuaciones

$$\mathbb{P}(A_i|A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_i) \tag{12}$$

cualesquiera sean $i_1, i_2, \dots, i_k, i$ distintos dos a dos.

Ejercicios adicionales

5. Se tira una moneda honesta n veces. Sea A el evento que se obtenga al menos una cara y sea B el evento que se obtengan al menos una cara y al menos una ceca. Analizar la independencia de los eventos A y B.

---

2. Teoría general

2.1. El axioma de continuidad

Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad.

En todo lo que sigue, además de suponer que valen los axiomas I-III enunciados en la Sección 1.1, también supondremos que vale el siguiente:

IV. *Axioma de continuidad.* Para cada sucesión decreciente de eventos

$$A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots \tag{13}$$

de $\mathcal{A}$, tal que

$$\bigcap_{n=1}^{\infty} A_n = \emptyset$$

vale que

$$\lim_{n\to\infty} \mathbb{P}(A_n) = 0.$$

Nota Bene 1. Si la familia de eventos $\mathcal{A}$ es finita el axioma IV se deduce de los axiomas I-III. En tal caso, en la sucesión (13) solo hay una cantidad finita de eventos diferentes. Si A_k es el menor de ellos, entonces todos los conjuntos A_{k+m} coinciden con A_k. Tenemos que $A_k = A_{k+m} = \bigcap_{n=1}^{\infty} A_n = \emptyset$ y $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\emptyset) = 0$. Por lo tanto, todos los ejemplos de espacios de probabilidad *finitos* satisfacen los axiomas I-IV.

Nota Bene 2. Se puede probar que para espacios muestrales infinitos, el axioma de continuidad IV es independiente de los axiomas I-III. Este nuevo axioma es esencial solamente para espacios de probabilidad infinitos y es casi imposible elucidar su significado empírico en la forma en que lo hicimos con los axiomas I-III.

Ejemplo 2.1. Sean $\Omega = \mathbb{Q} \cap [0,1] = \{r_1, r_2, r_3, \dots\}$ y $\mathcal{A}_0$ la familia de los subconjuntos de Ω de la forma $[a,b]$, $[a,b)$, $(a,b]$ o (a,b). La familia, $\mathcal{A}$ de todas las uniones finitas de conjuntos disjuntos de $\mathcal{A}_0$ es un álgebra de eventos. La medida de probabilidad definida por

$$
\begin{aligned}
\mathbb{P}(A) &:= b - a, && \text{si } A \in \mathcal{A}_0,\\
\mathbb{P}(A) &:= \textstyle\sum_{i=1}^{k} \mathbb{P}(A_i) && \text{si } A = \bigcup_{i=1}^{k} A_i, \text{ para } A_i \in \mathcal{A}_0 \text{ y } A_i \cap A_j = \emptyset,
\end{aligned}
$$

satisface los axiomas I-III pero no satisface el axioma de continuidad. En efecto, para cada $r \in \Omega$, $\{r\} \in \mathcal{A}$ y $\mathbb{P}(\{r\}) = 0$. Los eventos $A_n := \Omega \setminus \{r_1, \dots, r_n\}$, $n \in \mathbb{N}$, son decrecientes y $\bigcap_{n=1}^{\infty} A_n = \emptyset$, sin embargo $\lim_{n\to\infty} \mathbb{P}(A_n) = 1$, puesto que $\mathbb{P}(A_n) = 1$ para todo $n \geq 1$. $\square$

Teorema 2.2.

(a) Si $A_1 \supset A_2 \supset \cdots$ y $A = \bigcap_{n=1}^{\infty} A_n$, entonces $\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

(b) Si $A_1 \subset A_2 \subset \cdots$ y $A = \bigcup_{n=1}^{\infty} A_n$, entonces $\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

Demostración.

(a) Considerar la sucesión $B_n = A_n \setminus A = A_n \cap A^c$. Observar que $B_1 \supset B_2 \supset \cdots$ y $\bigcap_{n=1}^{\infty} B_n = \emptyset$. Por el axioma de continuidad se obtiene $\lim_{n\to\infty} \mathbb{P}(B_n) = 0$. Observando que $\mathbb{P}(B_n) = \mathbb{P}(A_n \setminus A) = \mathbb{P}(A_n) - \mathbb{P}(A)$ se deduce que

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A).$$

(b) Considerar la sucesión $B_n = A_n^c$. Observar que $B_1 \supset B_2 \supset \cdots$ y $\bigcap_{n=1}^{\infty} B_n = A^c$. Por el inciso (a) se obtiene $\lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. Observando que $\mathbb{P}(B_n) = \mathbb{P}(A_n^c) = 1 - \mathbb{P}(A_n)$ se deduce que

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A).$$

$\square$

Ejemplo 2.3 (Ternario de Cantor)**.** Se sortea número al azar U dentro del intervalo $(0,1]$, ¿cuál es la probabilidad de que el 1 no aparezca en el desarrollo en base 3 de U?

Si

$$U = \sum_{k\geq 1} \frac{a_k(U)}{3^k},$$

donde $a_k(U) \in \{0,1,2\}$ para todo $k \geq 1$, representa el desarrollo en base 3 de U, lo que se quiere calcular es la probabilidad del evento $A = \{a_k(U) \neq 1, \forall\, k \geq 1\}$. Observando que

$$A = \bigcap_{n=1}^{\infty} A_n,$$

donde $A_n = \{a_k(U) \neq 1, \forall\, 1 \leq k \leq n\}$ y notamos que $A_1 \supset A_2 \supset \cdots$ y usando el inciso (a) del **Teorema 2.2** tenemos que $\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$. El problema se reduce a calcular la sucesión de probabilidades $\mathbb{P}(A_n)$ y su límite.

Geométricamente el evento A_1 se obtiene eliminando el segmento $(1/3, 2/3)$ del intervalo $(0,1]$:

$$A_1 = (0, 1/3] \cup [2/3, 1].$$

Para obtener A_2 eliminamos los tercios centrales de los intervalos dos intervalos que componen A_1:

$$A_2 = (0, 1/9] \cup [2/9, 3/9] \cup [6/9, 7/9] \cup [8/9, 1].$$

Continuando de este modo obtenemos una caracterización geométrica de los eventos A_n: A_n es la unión disjunta de 2^n intervalos, cada uno de longitud 3^{-n}. En consecuencia,

$$\mathbb{P}(A_n) = 2^n \frac{1}{3^n} = \left(\frac{2}{3}\right)^n$$

Por lo tanto, $\mathbb{P}(A) = \lim_{n\to\infty} (2/3)^n = 0$. $\square$

Teorema 2.4 (σ-aditividad)**.** Si $A_1, A_2, \ldots,$ es una sucesión de eventos disjuntos dos a dos (i.e., $A_i \cap A_j = \emptyset$ para todos los pares i, j tales que $i \neq j$) y $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, entonces

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) \tag{14}$$

Demostración. La sucesión de eventos $R_n := \bigcup_{m>n} A_m$, $n \geq 1$, es decreciente y tal que $\bigcap_{n=1}^{\infty} R_n = \emptyset$. De acuerdo con el axioma IV tenemos que

$$\lim_{n\to\infty} \mathbb{P}(R_n) = 0. \tag{15}$$

Por otra parte, por el teorema de adición

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{k=1}^{n} \mathbb{P}(A_k) + \mathbb{P}(R_n). \tag{16}$$

De (16) y (15) se obtiene (14). $\square$

Corolario 2.5 (Fórmula de Probabilidad total)**.** Sea B un evento y sea $A_1, A_2, \ldots$ una sucesión de eventos disjuntos dos a dos tal que $\bigcup\limits_{n=1}^{\infty} A_n = \Omega$. En otras palabras, uno y solo uno de los eventos A_n debe ocurrir. Entonces

$$\mathbb{P}(B) = \sum_{n=1}^{\infty} \mathbb{P}(B|A_n)\mathbb{P}(A_n).$$

Rigurosamente, $\mathbb{P}(B|A_n)$ está definida cuando $\mathbb{P}(A_n) > 0$, por lo cual en la fórmula anterior interpretaremos que $\mathbb{P}(B|A_n)\mathbb{P}(A_n) = 0$ cuando $\mathbb{P}(A_n) = 0$.

Demostración. De la identidad de conjuntos

$$B = B \cap \Omega = B \cap \left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty}(B \cap A_n)$$

y la σ-aditividad de la medida de probabilidad $\mathbb{P}$ se deduce que

$$\mathbb{P}(B) = \sum_{n=1}^{\infty} \mathbb{P}(B \cap A_n).$$

Si $\mathbb{P}(A_n) = 0$, $\mathbb{P}(B \cap A_n) = 0$ porque $B \cap A_n \subset A_n$. Si $\mathbb{P}(A_n) > 0$, entonces $\mathbb{P}(B \cap A_n) = \mathbb{P}(B|A_n)\mathbb{P}(A_n)$. $\square$

Corolario 2.6 (Cubrimiento)**.** Si $A, A_1, A_2, \ldots$ pertenecen a $\mathcal{A}$ y $A \subset \bigcup\limits_{n=1}^{\infty} A_n$, entonces

$$\mathbb{P}(A) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Demostración. Una cuenta:

$$\begin{aligned}
A &= A \cap \left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty}\left(A \cap \left(A_n \setminus \bigcup_{k=1}^{n-1}(A_n \cap A_k)\right)\right) \\
\mathbb{P}(A) &= \sum_{n=1}^{\infty} \mathbb{P}\left(A \cap \left(A_n \setminus \bigcup_{k=1}^{n-1}(A_n \cap A_k)\right)\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n).
\end{aligned}$$

$\square$

Ejercicios adicionales

6. Sean Ω un conjunto no vacío y $\mathcal{A}$ una familia de subconjuntos de Ω cerrada por complementos y uniones finitas tal que $\Omega \in \mathcal{A}$. Sea $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ una función tal que

I. Para cada $A \in \mathcal{A}$, $\mathbb{P}(A) \geq 0$;

II. $\mathbb{P}(\Omega) = 1$;

III. Si A y B no tienen elementos en común, entonces $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

IV'. Si $(A_n)_{n\geq 1}$ es una sucesión de conjuntos disjuntos dos a dos y $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}$, entonces

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Mostrar que bajo esas condiciones resulta que cualquiera sea la sucesión decreciente $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$ tal que $\bigcap\limits_{n=1}^{\infty} A_n = \emptyset$ vale que

$$\lim_{n\to\infty} \mathbb{P}(A_n) = 0.$$

2.2. σ-álgebras y teorema de extensión

El álgebra $\mathcal{A}$ se llama una *σ-álgebra*, si toda unión numerable $\bigcup\limits_{n=1}^{\infty} A_n$ de conjuntos $A_1, A_2, \dots \in \mathcal{A}$, disjuntos dos a dos, también pertenece a $\mathcal{A}$.

De la identidad

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty}\left(A_n \setminus \bigcup_{k=1}^{n-1}(A_n \cap A_k)\right)$$

se deduce que la σ-álgebra también contiene todas las uniones numerables de conjuntos $A_1, A_2, \dots \in \mathcal{A}$. De la identidad

$$\bigcap_{n=1}^{\infty} A_n = \Omega \setminus \bigcup_{n=1}^{\infty} A_n^c$$

lo mismo puede decirse de las intersecciones.

Nota Bene. Solamente cuando disponemos de una medida de probabilidad, $\mathbb{P}$, definida sobre una σ-álgebra, $\mathcal{A}$, obtenemos libertad de acción total, sin peligro de que ocurran eventos que no tienen probabilidad.

Lema 2.7 (σ-álgebra generada)**.** Dada un álgebra $\mathcal{A}$ existe la menor σ-álgebra, $\sigma(A)$, que la contiene, llamada la σ-álgebra generada por $\mathcal{A}$.

Teorema 2.8 (Extensión)**.** Dada una función de conjuntos, $\mathbb{P}$, no negativa y σ-aditiva definida sobre un álgebra $\mathcal{A}$ se la puede extender a todos los conjuntos de la σ-álgebra generada por $\mathcal{A}$, $\sigma(\mathcal{A})$, sin perder ninguna de sus propiedades (no negatividad y σ-aditividad) y esta extensión puede hacerse de una sola manera.

Esbozo de la demostración. Para cada $A \subset \Omega$ definimos

$$\mathbb{P}^*(A) := \inf_{A \subset \cup_n A_n} \sum_n \mathbb{P}(A_n),$$

donde el ínfimo se toma respecto a todos los cubrimientos del conjunto A por colecciones finitas o numerables de conjuntos A_n pertenecientes a $\mathcal{A}$. De acuerdo con el Teorema de cubrimiento $\mathbb{P}^*(A)$ coincide con $\mathbb{P}(A)$ para todo conjunto $A \in \mathcal{A}$.

La función $\mathbb{P}^*$ es no negativa y σ-aditiva sobre $\sigma(\mathcal{A})$. La unicidad de la extensión se deduce de la propiedad minimal de $\sigma(\mathcal{A})$. □

3. Simulación de experimentos aleatorios

3.1. Números aleatorios.

Teóricamente, los números aleatorios son realizaciones independientes del experimento conceptual que consiste en *"elegir al azar"* un número U del intervalo $[0, 1]$. Aquí la expresión *"elegir al azar"* significa que el número U tiene la distribución uniforme sobre el intervalo $(0, 1]$, i.e., la probabilidad del evento $U \in (a, b)$ es igual a $b - a$, cualquiera sea la pareja de números reales a y b tales que $0 < a < b \leq 1$.

Prácticamente, toda computadora tiene instalado un algoritmo para simular números aleatorios y se pueden obtener mediante una instrucción "random". En el *software* Octave, por ejemplo, la sentencia *rand* simula un número aleatorio y $rand(1, n)$ simula un vector de n números aleatorios. En algunas calculadoras (llamadas científicas) la instrucción *Ran#* permite simular números aleatorios de tres dígitos. En algunos libros de texto se pueden encontrar tablas de números aleatorios (p. ej., Meyer, P. L.: Introductory Probability and Statistical Applications. Addison-Wesley, Massachusetts. (1972))

Como usar los números aleatorios. La idea principal puede presentarse mediante un ejemplo muy simple. Queremos construir un mecanismo aleatorio para simular el lanzamiento de una moneda cargada con probabilidad p de obtener de obtener "cara". Llamemos X al resultado del lanzamiento: $X \in \{0, 1\}$ con la convención de que "cara"$= 1$ y "ceca"$= 0$.

Para construir X usamos un número aleatorio U, uniformemente distribuido sobre el intervalo $[0,1]$ y definimos

$$X := \mathbf{1}\{1-p < U \leq 1\}. \tag{17}$$

Es fácil ver X satisface las condiciones requeridas. En efecto,

$$\mathbb{P}(X=1) = \mathbb{P}(1-p < U \leq 1) = 1-(1-p) = p.$$

La ventaja de la construcción es que puede implementarse casi de inmediato en una computadora. Por ejemplo, si $p = 1/2$, una rutina en Octave para simular X es la siguiente

Rutina para simular el lanzamiento de una moneda

```
U = rand;
if U>1/2
   X=1;
else
   X=0;
end
X
```

Nota Bene. El ejemplo anterior es el prototipo para construir y simular experimentos aleatorios. Con la misma idea podemos construir experimentos aleatorios tan complejos como queramos.

3.2. Simulación de experimentos aleatorios

Supongamos que $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ representa el espacio muestral correspondiente a un experimento aleatorio y que cada evento elemental $\omega_k \in \Omega$ tiene asignada la probabilidad $p(\omega_k) = p_k$. Usando un número aleatorio, U, uniformemente distribuido sobre el intervalo $(0,1]$ podemos construir un mecanismo aleatorio, X, adecuado para simular los resultados del experimento aleatorio considerado. Basta con definir

$$X = \sum_{k=1}^{m} k\mathbf{1}\left\{L_{k-1} < U \leq L_k\right\}, \text{ donde } L_0 := 0 \text{ y } L_k := \textstyle\sum_{i=1}^{k} p_i, \text{ para } 1 \leq k \leq m \tag{18}$$

e identificar cada evento elemental $\omega_k \in \Omega$ con su correspondiente subíndice k. En efecto, de la definición (18) se deduce que para cada $k = 1, \dots, m$ vale que

$$\mathbb{P}(X=k) = \mathbb{P}(L_{k-1} < U \leq L_k) = L_k - L_{k-1} = p_k.$$

Nota Bene. El mecanismo aleatorio definido en (18) puede construirse "gráficamente" de la siguiente manera:

1. Partir el intervalo $(0, 1]$ en subintervalos sucesivos $I_1, \dots, I_m$ de longitudes $p_1, \dots, p_m$, respectivamente.

2. Sortear un número aleatorio, U, y observar en cuál de los intervalos de la partición cae.

3. Si U cae en el intervalo I_k, producir el resultado ω_k.

Ejemplo 3.1 (Lanzar un dado equilibrado)**.** Se quiere simular el lanzamiento de un dado equilibrado. El espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$ y la función de probabilidades es $p(k) = 1/6$, $k = 1, \dots, 6$. El mecanismo aleatorio $X = X(U)$, definido en (18), se construye siguiendo los siguientes pasos: **1.** Partir el intervalo $(0, 1]$ en 6 intervalos sucesivos de longitud 1/6: $I_1 = (0, 1/6], I_2 = (1/6, 2/6], I_3 = (2/6, 3/6], I_4 = (3/6, 4/6], I_5 = (4/6, 5/6]$ e $I_6 = (5/6, 6/6]$. **2.** Sortear un número aleatorio U. **3.** Si $U \in I_k$, $X = k$. En pocas palabras,

$$X = \sum_{k=1}^{6} k\mathbf{1}\left\{\frac{k-1}{6} < U \leq \frac{k}{6}\right\}. \tag{19}$$

Por ejemplo, si sorteamos un número aleatorio, U y se obtiene que $U = 0.62346$, entonces el valor simulado del dado es $X = 4$. Una rutina en Octave para simular X es la siguiente

Rutina para simular el lanzamiento de un dado

```
U=rand;
k=0;
do
  k++;
until((k-1)/6<U & U<=k/6)
X=k
```

□

3.3. Estimación de probabilidades

Formalmente, un experimento aleatorio se describe mediante un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$. Todas las preguntas asociadas con el experimento pueden reformularse en términos de este espacio. En la práctica, decir que un evento A ocurre con una determinada probabilidad $\mathbb{P}(A) = p$ equivale a decir que en una serie suficientemente grande de experimentos las frecuencias relativas de ocurrencia del evento A

$$\hat{p}_k(A) = \frac{n_k(A)}{n_k}$$

(donde n_k es la cantidad de ensayos realizados en la k-ésima serie y $n_k(A)$ es la cantidad en los que ocurre A) son aproximadamente idénticas unas a otras y están próximas a p. Las series de experimentos se pueden simular en una computadora utilizando un *generador de números aleatorios*.

Ejemplo 3.2. El experimento consiste en lanzar 5 monedas equilibradas y registrar la cantidad N de *caras* observadas. El conjunto de todos los resultados posibles es $\Omega = \{0, 1, 2, 3, 4, 5\}$. El problema consiste en asignarle probabilidades a los eventos elementales.

La solución experimental del problema se obtiene realizando *una serie suficientemente grande de experimentos* y asignando a cada evento elemental su frecuencia relativa.

Sobre la base de una rutina similar a la que presentamos en la sección 3.1 para simular el resultado del lanzamiento de una moneda se pueden simular $n = 10000$ realizaciones del experimento que consiste en lanzar 5 monedas equilibradas. Veamos como hacerlo. Usamos la construcción (17) para simular el lanzamiento de 5 monedas equilibradas X_1, X_2, X_3, X_4, X_5. La cantidad de caras observadas es la suma de las X_i: $N = X_1 + X_2 + X_3 + X_4 + X_5$.

Repitiendo la simulación 10000 veces (o genéricamente n veces), obtenemos una tabla que contiene la cantidad de veces que fué simulado cada valor de la variable N. Supongamos que obtuvimos la siguiente tabla:

valor simulado	0	1	2	3	4	5
cantidad de veces	308	1581	3121	3120	1564	306

(20)

En tal caso diremos que se obtuvieron las siguientes estimaciones

$$\begin{array}{lll} \mathbb{P}(N=0) \approx 0.0308, & \mathbb{P}(N=1) \approx 0.1581, & \mathbb{P}(N=2) \approx 0.3121, \\ \mathbb{P}(N=3) \approx 0.3120, & \mathbb{P}(N=4) \approx 0.1564, & \mathbb{P}(N=5) \approx 0.0306. \end{array}$$

Para finalizar este ejemplo, presentamos un programa en Octave que simula diez mil veces el lanzamiento de cinco monedas equilibradas, contando en cada una la cantidad de caras observadas y que al final provee una tabla como la representada en (20)

```
n = 10000;
N = zeros(1,n);
for i=1:n
    U=rand(1,5);
    X=[U<=(1/2)];
    N(i)=sum(X);
end
for j=1:6
    T(j)=sum([N==j-1]);
end
T
```

□

Nota Bene. Usando las herramientas que proporciona el cálculo combinatorio se puede demostrar que para cada $k \in \{0, 1, 2, 3, 4, 5\}$ vale que $\mathbb{P}(N = k) = \binom{5}{k}\frac{1}{32}$. En otros términos,

$$\mathbb{P}(N = 0) = 0.03125, \qquad \mathbb{P}(N = 1) = 0.15625, \qquad \mathbb{P}(N = 2) = 0.31250,$$
$$\mathbb{P}(N = 3) = 0.31250, \qquad \mathbb{P}(N = 4) = 0.15625, \qquad \mathbb{P}(N = 5) = 0.03125.$$

□

Ejemplo 3.3 (Paradoja de De Mere). ¿Cuál de las siguientes apuestas es más conveniente?

- Obtener al menos un as en 4 tiros de un dado.
- Obtener al menos un doble as en 24 tiros de dos dados.

La construcción (19) permite simular 4 tiros de un dado usando 4 números aleatorios independientes U_1, U_2, U_3, U_4. La cantidad de ases obtenidos en los 4 tiros es la suma $S = \sum_{i=1}^{4} \mathbf{1}\{0 < U_i \leq 1/6\}$. El evento A_1 = "*obtener al menos un as en 4 tiros de un dado*" equivale al evento $S \geq 1$. Si repetimos la simulación 10000 veces podemos obtener una estimación (puntual) de la probabilidad del evento A_1 calculando su frecuencia relativa.

La siguiente rutina (en Octave) provee una estimación de la probabilidad del evento A_1 basada en la repetición de 10000 simulaciones del experimento que consiste en tirar 4 veces un dado.

Rutina 1

```
n=10000;
A1=zeros(1,n);
for i=1:n
    U=rand(1,4);
    S=sum(U<=1/6);
    if S>=1
       A1(i)=1;
    else
       A1(i)=0;
    end
end
hpA1=sum(A1)/n
```

Ejecutando 10 veces la **Rutina 1** se obtuvieron los siguientes resultados para la frecuencia relativa del evento A_1

0.5179 0.5292 0.5227 0.5168 0.5204 0.5072 0.5141 0.5177 0.5127 0.5244

Notar que los resultados obtenidos se parecen entre sí e indican que la probabilidad de obtener al menos un as en 4 tiros de un dado es mayor que 0.5.

La construcción (19) permite simular 24 tiros de dos dados usando 48 números aleatorios independientes U_1, U_2, U_3, $U_4, \dots, U_{47}, U_{48}$. La cantidad de veces que se obtiene un doble as en los 24 tiros es la suma $S = \sum_{i=1}^{24} \mathbf{1}\{0 < U_{2i-1} \leq 1/6, 0 < U_{2i} \leq 1/6\}$. El evento $A_2 =$ "*obtener al menos un doble as en 24 tiros de dos dados*" equivale al evento $S \geq 1$. Si repetimos la simulación 10000 veces podemos obtener una estimación (puntual) de la probabilidad del evento A_2 calculando su frecuencia relativa.

La siguiente rutina (en Octave) provee una estimación de la probabilidad del evento A_2 basada en la repetición de 10000 simulaciones del experimento que consiste en tirar 24 veces dos dado.

Rutina 2

```
n=10000;
A2=zeros(1,n);
for i=1:n
    U=rand(2,24);
    V=(U<=1/6);
    S=sum(V(1,:).*V(2,:));
    if S>=1
       A2(i)=1;
    else
       A2(i)=0;
    end
end
hpA2=sum(A2)/n
```

Ejecutando 10 veces la **Rutina 2** se obtuvieron los siguientes resultados para la frecuencia relativa del evento A_2

0.4829   0.4938   0.4874   0.4949   0.4939   0.4873   0.4882   0.4909   0.4926   0.4880

Notar que los resultados obtenidos se parecen entre sí e indican que la probabilidad de obtener al menos un as en 4 tiros de un dado es menor que 0.5.

Conclusión. Los resultados experimentales obtenidos indican que es mejor apostar a que se obtiene al menos un as en 4 tiros de un dado que apostar a que se obtiene al menos un doble as en 24 tiros de un dado.

La Figura 6 permite comparar las frecuencias relativas de los eventos A_1 y A_2 obtenidas repitiendo 100 las rutinas 1 y 2 respectivamente. Notar que las frecuencias relativas del evento A_1 fluctúan alrededor del valor 0.5177 y las del evento A_2 fluctúan alrededor del valor 0.4914. Se podría concluir que $\mathbb{P}(A_1) \approx 0.5177$ y $\mathbb{P}(A_2) \approx 0.4914$.

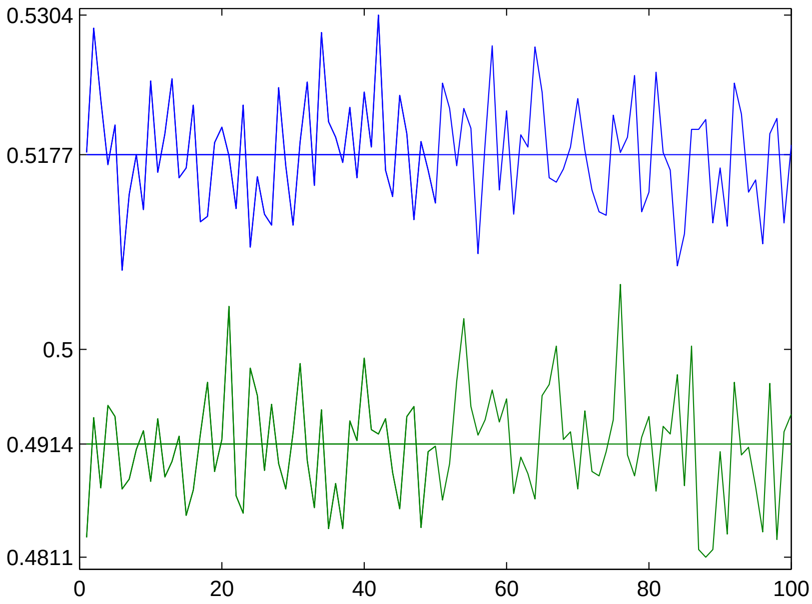


Figura 6: Comparación gráfica: Cada punto de la forma $(k, \hat{p}_k(A_1))$, $k = 1, \dots, 100$ sobre la poligonal azul representa la frecuencia relativa del evento "*obtener al menos un as en 4 lanzamientos de un dado*" en la k-ésima serie de 10000 simulaciones de dicho experimento. Cada punto de la forma $(k, \hat{p}_k(A_2))$, $k = 1, \dots, 100$ sobre la poligonal verde representa la frecuencia relativa del evento "*obtener al menos un doble as en 24 lanzamientos de dos dados*" en la k-ésima serie de 10000 simulaciones de dicho experimento.

Análisis de las fluctuaciones. Para estimar la probabilidad de un evento A, que puede ocurrir como resultado de un experimento aleatorio determinado, podemos realizar N series de n repeticiones independientes del experimento y asignarle al evento A la frecuencia relativa que se obtiene en cada una de las series. Esto produce una muestra de N frecuencias relativas $\hat{p}_k(A) = n_k(A)/n$, $k = 1, \dots, N$. Típicamente esas frecuencias relativas fluctúan alrededor de la probabilidad $\mathbb{P}(A)$ del evento A. La pregunta es: ¿cuan próximas están las frecuencias relativas $\hat{p}_k(A)$ de la probabilidad $\mathbb{P}(A)$? En otras palabras, ¿cómo se comportan las diferencias $\Delta_k(A) = \hat{p}_k(A) - \mathbb{P}(A)$? La respuesta de esa pregunta requiere herramientas teóricas que serán desarrolladas más adelante. Por ahora nos vamos a conformar con una perspectiva "experimental" limitada al ejemplo que estamos analizando.

Ejecutando 100 veces la **Rutina 1** obtuvimos la siguiente tabla de diferencias, $\Delta_k(A_1) = \hat{p}_k(A_1) - 0.5177$, entre las frecuencias relativas y el valor alrededor del que fluctúan, 0.5177.

Tabla 1: Fluctuaciones observadas al estimar la probabilidad del evento $A_1 =$ "*obtener al menos un as en 4 lanzamientos de un dado* mediante sus frecuencias relativas de aparición.

0.0002	0.0115	0.0050	−0.0009	0.0027	**−0.0105**	−0.0036	0.0000
−0.0050	0.0067	−0.0016	0.0019	0.0069	−0.0021	−0.0012	0.0045
−0.0061	−0.0056	0.0011	0.0025	−0.0001	−0.0049	0.0045	−0.0084
−0.0020	−0.0054	−0.0064	0.0061	−0.0010	−0.0064	0.0012	0.0066
−0.0028	0.0111	0.0030	0.0016	−0.0007	0.0043	−0.0021	0.0057
0.0007	**0.0127**	−0.0014	−0.0038	0.0054	0.0019	−0.0059	0.0012
−0.0014	−0.0044	0.0065	0.0042	−0.0010	0.0042	0.0024	−0.0090
0.0010	0.0099	−0.0032	0.0040	−0.0054	0.0018	0.0007	0.0098
0.0057	−0.0021	−0.0025	−0.0013	0.0007	0.0051	0.0004	−0.0032
−0.0052	−0.0055	0.0036	0.0002	0.0016	0.0072	−0.0052	−0.0034
0.0075	0.0002	−0.0014	−0.0101	−0.0072	0.0023	0.0023	0.0032
−0.0062	−0.0012	−0.0065	0.0065	0.0037	−0.0034	−0.0023	−0.0081
0.0019	0.0033	−0.0062	0.0009				

Con un poco de paciencia se puede observar que los valores mínimo y máximo que aparecen en la tabla son, −0.0105 y 0.0127, respectivamente. Esto significa que para todos los valores observados vale $|\hat{p}_k(A_1) - 0.5177| \leq 0.0127$.

Notar que la cantidad de datos presentados en la tabla dificulta su interpretación. En tales casos, resulta útil agrupar los valores en intervalos y graficar la cantidad de datos que pertenecen a cada intervalo. Este procedimiento se denomina construir un *histograma de frecuencias.*

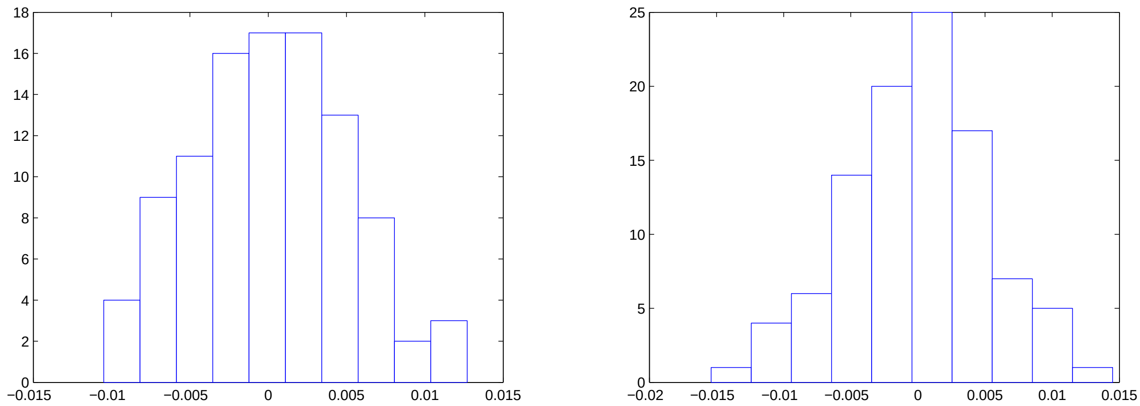
Grosso modo, un *histograma de frecuencias* de una muestra $x_1, \dots, x_N$ se obtiene eligiendo una partición en m intervalos de extremos $a_0 < \cdots < a_m$, de longitudes $a_j - a_{j-1}$; calculando las *frecuencias* $f_j = |x_k \in [a_{j-1}, a_j)|$ y graficando la función igual a f_j en el intervalo $[a_{j-1}, a_j)$ y a 0 fuera de los intervalos.

Dividiendo el intervalo $[-0.0105, 0.0127]$ en 10 intervalos consecutivos de longitudes $(0.0127 - (-0.0105))/10 = 0.0023$ y contando cuántas diferencias $\Delta_k(A_1)$ pertenecen a cada uno, se obtiene la siguiente las siguientes frecuencias: 4, 9, 11, 16, 17, 17, 13, 8, 2, 3, el histograma de frecuencias correspondiente se muestra en la Figura 7 (a).

Ejecutando 100 veces la **Rutina 2** se obtuvo una muestra de diferencias entre las frecuencias relativas con que aparece el evento $A_2 =$ "*obtener al menos un doble as en 24 lanzamientos de dos dados* y el valor alrededor del que fluctúan, 04914.

Los valores mínimo y máximo obtenidos, fueron −0.0154 y 0.0145, respectivamente. Esto significa que para todos los valores observados vale $|\hat{p}_k(A_1) - 0.5177| \leq 0.0154$.

Dividiendo el intervalo $[-0.0154, 0.0145]$ en 10 intervalos consecutivos de igual longitud y contando cuántas diferencias $\Delta_k(A_2) = \hat{p}_k(A_2) - 0.4914$ pertenecen a cada uno, se obtuvieron las siguientes frecuencias: 1, 4, 6, 14, 20, 25, 17, 7, 5, 1. El histograma de frecuencias correspondiente se muestra en la Figura 7 (b).



(a) (b)

Figura 7: (a) Histograma correspondiente a las fluctuaciones observadas al estimar la probabilidad del evento A_1=“*obtener al menos un as en 4 lanzamientos de un dado*” mediante sus frecuencias relativas de aparición. El histograma se obtuvo dividiendo en 10 partes iguales el intervalo $[-0.0105, 0.0127]$ cuyos extremos son el mínimo y el máximo de las diferencias $\Delta_k(A_1) = \hat{p}_k(A_1) - 0.5177$ observadas en una muestra de 100: las cantidades observadas en cada una de las partes son: 4, 9, 11, 16, 17, 17, 13, 8, 2, 3. (b) Histograma correspondiente a las fluctuaciones observadas al estimar la probabilidad del evento A_2=“*obtener al menos un doble as en 24 lanzamientos de un dado*” mediante sus frecuencias relativas de aparición. El histograma se obtuvo dividiendo en 10 partes iguales el intervalo $[-0.0154, 0.0145]$ cuyos extremos son el mínimo y el máximo de las diferencias $\Delta_k(A_2) = \hat{p}_k(A_2) - 0.4914$ observadas en una muestra de 100: las cantidades observadas en cada una de las partes son: 1, 4, 6, 14, 20, 25, 17, 7, 5, 1. $\square$

4. Modelos discretos

Los espacios muestrales más simples son aquellos que contienen un número finito, n, de puntos. Si n es pequeño (como en el caso de tirar algunas monedas), es fácil visualizar el espacio. El espacio de distribuciones de cartas de poker es más complicado. Sin embargo, podemos imaginar cada punto muestral como una ficha y considerar la colección de esas fichas como representantes del espacio muestral. Un evento A se representa por un determinado conjunto de fichas, su complemento A^c por las restantes. De aquí falta sólo un paso para imaginar una bol con infinitas fichas o un espacio muestral con una sucesión infinita de puntos $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$.

Definición 4.1. *Un espacio muestral se llama* discreto *si contiene finitos o infinitos puntos que pueden ordenarse en una sucesión* $\omega_1, \omega_2, \dots$.

Sean Ω un conjunto infinito numerable y $\mathcal{A}$ la σ-álgebra de todos los subconjuntos contenidos en Ω. Todos los espacios de probabilidad que se pueden construir sobre $(\Omega, \mathcal{A})$ se obtienen de la siguiente manera:

1. Tomamos una sucesión de números no negativos $\{p(\omega) : \omega \in \Omega\}$ tal que

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

2. Para cada evento $A \in \mathcal{A}$ definimos $\mathbb{P}(A)$ como la suma de las probabilidades de los eventos elementales contenidos en A:

$$\mathbb{P}(A) := \sum_{\omega \in A} p(\omega). \tag{21}$$

Nombres. La función $p : \Omega \to [0, 1]$ que asigna probabilidades a los eventos elementales $\omega \in \Omega$ se llama *función de probabilidad*. La función $\mathbb{P} : \mathcal{A} \to [0, 1]$ definida en (21) se llama *la medida de probabilidad inducida por* p.

Nota Bene 1. De la definición (21) resultan inmediatamente las siguientes propiedades

(i) Para cada $A \in \mathcal{A}$ vale que $\mathbb{P}(A) \geq 0$

(ii) $\mathbb{P}(\Omega) = 1$.

(iii) *σ-aditividad*. Si $A_1, A_2, \dots$ es una sucesión de eventos disjuntos dos a dos, entonces

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Nota Bene 2. No se excluye la posibilidad de que un punto tenga probabilidad cero. Esta convención parece artificial pero es necesaria para evitar complicaciones. En espacios discretos probabilidad cero se interpreta como imposibilidad y cualquier punto muestral del que se sabe que tiene probabilidad cero puede suprimirse impunemente del espacio muestral. Sin embargo, frecuentemente los valores numéricos de las probabilidades no se conocen de antemano, y se requieren complicadas consideraciones para decidir si un determinado punto muestral tiene o no probabilidad positiva.

Distribución geométrica

Ejemplo 4.2 (Probabilidad geométrica)**.** Sea p un número real tal que $0 < p < 1$. Observando que

$$\sum_{n=1}^{\infty}(1-p)^{n-1} = \frac{1}{p},$$

se deduce que la función $p : \mathbb{N} \to \mathbb{R}$ definida por

$$p(n) := (1-p)^{n-1}p, \qquad n = 1, 2, \dots$$

define una función de probabilidad en $\Omega = \mathbb{N} = \{1, 2, 3, \dots\}$ que se conoce por el nombre de *distribución geométrica de parámetro p*. Esta función de probabilidades está íntimamente relacionada con la cantidad de veces que debe repetirse un experimento aleatorio para que ocurra un evento A (prefijado de antemano) cuya probabilidad de ocurrencia en cada experimento individual es p. □

Ejemplo 4.3. El experimento consiste en lanzar una moneda tantas veces como sea necesario hasta que salga cara. El resultado del experimento será la cantidad de lanzamientos necesarios hasta que se obtenga cara. Los resultados posibles son

$$\Omega = \{1, 2, 3, \dots\} \cup \{\infty\}.$$

El símbolo ∞ está puesto para representar la posibilidad de que todas las veces que se lanza la moneda el resultado obtenido es ceca. El primer problema que debemos resolver es asignar probabilidades a los puntos muestrales. Una forma de resolverlo es la siguiente. Cada vez que se arroja una moneda los resultados posibles son cara (H) o ceca (T). Sean p y q la probabilidad de observar cara y ceca, respectivamente, en cada uno de los lanzamientos. Claramente, p y q deben ser no negativos y

$$p + q = 1.$$

Suponiendo que cada lanzamiento es independiente de los demás, las probabilidades se multiplican. En otras palabras, *la probabilidad de cada secuencia determinada es el producto obtenido de reemplazar las letras H y T por p y q, respectivamente*. Así,

$$\mathbb{P}(H) = p; \qquad \mathbb{P}(TH) = qp; \qquad \mathbb{P}(TTH) = qqp; \qquad \mathbb{P}(TTTH) = qqqp.$$

Puede verse que para cada $n \in \mathbb{N}$ la secuencia formada por $n-1$ letras T seguida de la letra H debe tener probabilidad $q^{n-1}p = (1-p)^{n-1}p$.

El argumento anterior sugiere la siguiente asignación de probabilidades sobre Ω: para cada $n \in \mathbb{N}$, $p(n)$, la probabilidad de que la primera vez que se obtiene cara ocurra en el n-ésimo lanzamiento de la moneda está dada por

$$p(n) = (1-p)^{n-1}p.$$

Como las probabilidades geométricas suman 1 (ver el ejemplo 4.2) al resultado "ceca en todos los tiros" se le debe asignar probabilidad $p(\infty) = 0$. Como el espacio muestral es discreto no hay problema en suprimir el punto ∞.

Consideremos el evento $A =$ "*se necesitan una cantidad par de tiros para obtener la primer cara*". Entonces,

$$A = \{2, 4, 6, 8, \dots\},$$

y

$$\begin{aligned}
\mathbb{P}(A) &= \sum_{\omega \in A} p(\omega) = \sum_{k=1}^{\infty} p(2k) = \sum_{k=1}^{\infty} q^{2k-1} p = pq \sum_{k=0}^{\infty} q^{2k} = pq \left(\frac{1}{1-q^2} \right) \\
&= \frac{pq}{(1-q)(1+q)} = \frac{q}{1+q} = \frac{1-p}{2-p}.
\end{aligned}$$

□

Ejemplo 4.4. Lucas y Monk juegan a la moneda. Lanzan una moneda equilibrada al aire, si sale cara, Lucas le gana un peso a Monk; si sale ceca, Monk le gana un peso a Lucas. El juego termina cuando alguno gana dos veces seguidas.

El espacio muestral asociado a este experimento aleatorio es

$$\Omega = \{HH, TT, HTT, THH, HTHH, THTT, \dots\}.$$

Como podemos tener secuencias de cualquier longitud de caras y cecas alternadas, el espacio muestral es necesariamente infinito.

El evento $A_1 =$"*la moneda fue lanzada como máximo tres veces*" está dado por todos los elementos de Ω que tienen longitud menor o igual que tres:

$$A_1 = \{HH, TT, HTT, THH\}$$

y su probabilidad es

$$\mathbb{P}(A_1) = \mathbb{P}(HH) + \mathbb{P}(TT) + \mathbb{P}(HTT) + \mathbb{P}(THH) = \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = \frac{3}{4}.$$

El evento $A_2 =$"*ceca en el primer lanzamiento*" está dado por todos los elementos de Ω que comienzan con T:

$$A_2 = \{TT, THH, THTT, THTHH, \dots\},$$

y su probabilidad es

$$\begin{aligned}
\mathbb{P}(A_2) &= \mathbb{P}(TT) + \mathbb{P}(THH) + \mathbb{P}(THTT) + \mathbb{P}(THTHH) + \cdots \\
&= \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \frac{1}{2^5} + \cdots = \frac{1}{2}.
\end{aligned}$$

¿Cuál es la probabilidad de que el juego termine alguna vez? Si definimos los eventos $A_n :=$"*el juego termina en la n-ésima jugada*", $n \geq 2$, tendremos que el evento "el juego termina alguna vez" es la unión disjunta de los eventos $A_1, A_2, \dots$, y por lo tanto su probabilidad es la suma de las probabilidades de los eventos A_n. Para cada $n \geq 2$ la probabilidad de A_n es

$$\mathbb{P}(A_n) = \frac{2}{2^n} = \frac{1}{2^{n-1}}$$

En consecuencia la probabilidad de que el juego termine alguna vez es

$$\sum_{n\geq 2} \frac{1}{2^{n-1}} = \sum_{n\geq 1} \frac{1}{2^n} = 1.$$

□

Distribución de Poisson

Ejemplo 4.5 (Probabilidad de Poisson)**.** Sea λ un número real positivo. Observando que

$$e^{\lambda} = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!},$$

se deduce que la función $p : \mathbb{N}_0 \to \mathbb{R}$ definida por

$$p(n) := e^{-\lambda} \frac{\lambda^n}{n!}, \qquad n = 0, 1, 2, \dots$$

define una función de probabilidad en $\Omega = \mathbb{N}_0 = \{0, 1, 2, \dots\}$, conocida como *la distribución de Poisson de intensidad* λ. □

5. Modelos continuos

5.1. Puntos al azar sobre un segmento. La distribución uniforme

Elegir un punto al azar dentro de un segmento de recta de longitud finita es un experimento conceptual intuitivamente claro. Desde el punto de vista teórico el experimento debe describirse mediante un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$.

No se pierde generalidad, si se supone que la longitud del segmento es la unidad y se lo identifica con el intervalo $\Omega = [0, 1]$. La σ-álgebra de eventos $\mathcal{A}$ y la medida de probabilidad $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ se construyen por etapas.

1. Definimos $\mathcal{A}_0$ como la familia de los intervalos contenidos en Ω de la forma $[a, b]$, $[a, b)$, $(a, b]$ o (a, b), $a \leq b$ (notar que $\mathcal{A}_0$ no es un álgebra) y definimos $\mathbb{P}_0 : \mathcal{A}_0 \to \mathbb{R}$ de la siguiente manera:

$$\mathbb{P}_0(A) := \text{longitud}(A) = b - a, \text{ si los extremos del intervalo } A \text{ son } a \text{ y } b.$$

2. La familia $\mathcal{A}_1$ de todas las uniones finitas de conjuntos disjuntos de $\mathcal{A}_0$ es un álgebra de eventos y la función $\mathbb{P}_1 : \mathcal{A}_1 \to \mathbb{R}$ definida por

$$\mathbb{P}_1(A) := \sum_{i=1}^{k} \mathbb{P}_0(A_i), \text{ si } A = \bigcup_{i=1}^{k} A_i,$$

donde $A_1, \ldots, A_k \in \mathcal{A}_0$ y $A_i \cap A_j = \emptyset$ para toda pareja de índices $i \neq j$, es una medida de probabilidad (pues satisface los axiomas I-IV).

3. El teorema de extensión se ocupa del resto: la medida de probabilidad $\mathbb{P}_1$ definida sobre el álgebra $\mathcal{A}_1$ se extiende unívocamente a una medida de probabilidad $\mathbb{P}$ definida sobre la σ-álgebra generada por $\mathcal{A}_1$, $\mathcal{A} := \sigma(A_1)$.

Nota Bene. Esta definición de probabilidad que a cada intervalo $A \subset [0, 1]$ le asigna su respectiva longitud se llama la *distribución uniforme sobre el intervalo* $[0, 1]$ y constituye una generalización de la noción de equiprobabilidad sobre la que se basa la definición de Laplace de la probabilidad para espacios finitos: *"casos favorables sobre casos posibles"*.

5.2. Geometría y probabilidad

Una construcción completamente análoga a la de la sección anterior permite describir teóricamente el experimento conceptual, intuitivamente claro, que consiste en *elegir un punto al azar dentro de una región plana, $\Lambda \subset \mathbb{R}^2$, de área finita y no nula.* Para fijar ideas, se puede imaginar que la región plana es un blanco sobre el que se arroja un dardo.

Ejemplo 5.1 (Dardos)**.** El juego de dardos consiste en tirar un dardo contra un blanco circular. Supongamos que disparamos un dardo (que acertamos al blanco) y observamos dónde se clavó. Naturalmente, los resultados posibles de este experimento son todos los puntos del blanco. No se pierde generalidad si se supone que el centro del blanco es el origen de $\mathbb{R}^2$ y que su radio es 1. En tal caso el espacio muestral de este experimento es

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$

Intuitivamente, la probabilidad de acertarle a un punto predeterminado (arbitrario) debería ser cero. Sin embargo, la probabilidad de que el dardo se clave en cualquier subconjunto ("gordo") A del blanco debería ser proporcional a su área y determinarse por la fracción del área del blanco contenida en A. En consecuencia, definimos

$$\mathbb{P}(A) := \frac{\text{área de } A}{\text{área del blanco}} = \frac{\text{área de } A}{\pi}.$$

Por ejemplo, si $A = \{(x, y) : x^2 + y^2 \leq r^2\}$ es el evento que el dardo caiga a distancia $r < 1$ del centro del blanco, entonces

$$\mathbb{P}(A) = \frac{\pi r^2}{\pi} = r^2.$$

□

"Puntos al azar en regiones planas". Si hacemos abstracción de la forma circular del blanco y de la semántica involucrada en el juego de dardos, obtenemos un modelo probabilístico para el experimento conceptual que consiste en "sortear" o *elegir un punto al azar* en una región plana $\Lambda \subset \mathbb{R}^2$ de área finita y positiva. El espacio muestral es la región plana, $\Omega = \Lambda$, la σ-álgebra de los eventos, $\mathcal{A}$, es la familia de todos los subconjuntos de Λ a los que se les puede medir el área y la probabilidad de cada evento A es la fracción del área de Λ contenida en A. Esto es,

$$\mathbb{P}(A) := \frac{\text{área}(A)}{\text{área}(\Lambda)}. \tag{22}$$

Esta forma de asignar probabilidades es la equivalente para el caso continuo de la fórmula *casos favorables sobre casos posibles* utilizada en espacios muestrales finitos para modelar experimentos aleatorios con resultados equiprobables.

Nota Bene. Si en lugar de elegir un punto al azar dentro del segmento $[a, b]$ elegimos dos puntos de manera independiente, el experimento tendrá por resultado un par de números reales contenidos en $[a, b]$. El espacio muestral será el cuadrado de lado $[a, b]$, $\Omega = [a, b] \times [a, b]$. En este espacio la asignación de probabilidades definida en (22) resulta consistente con la noción de independencia.

Ejemplo 5.2. Se eligen al azar (y en forma independiente) dos puntos x_1 y x_2 dentro de un segmento de longitud L. Hallar la probabilidad de que la longitud del segmento limitado por los puntos x_1 y x_2 resulte menor que $L/2$.

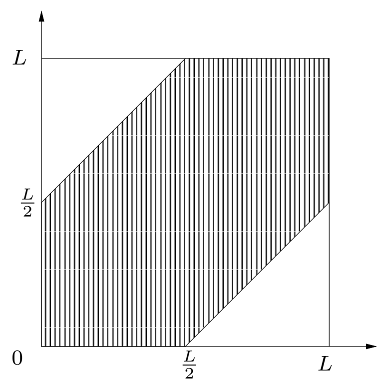


Figura 8: La región sombreada corresponde al evento A="la longitud del segmento limitado por los puntos x_1 y x_2 resulte menor que $L/2$".

El espacio muestral de este experimento es un cuadrado de lado L que puede representarse en la forma $\Omega = \{(x_1, x_2) : 0 \le x_1 \le L, 0 \le x_1 \le L\}$.

El evento A="la longitud del segmento limitado por los puntos x_1 y x_2 resulte menor que $L/2$" puede ocurrir de dos maneras distintas:

(1) si $x_1 \leq x_2$, se debe cumplir la desigualdad $x_2 - x_1 < L/2$;
(2) si $x_2 < x_1$, debe cumplirse la desigualdad $x_1 - x_2 < L/2$.

Observando la Figura 8 está claro que el área del evento A se obtiene restando al área del cuadrado de lado L el área del cuadrado de lado $L/2$:

$$\text{área de } A = L^2 - \frac{L^2}{4} = \frac{3}{4}L^2.$$

Como el área total del espacio muestral es L^2, resulta que $\mathbb{P}(A) = 3/4$. □

Ejemplo 5.3 (Las agujas de Buffon)**.** .Una aguja de longitud $2l$ se arroja sobre un plano dividido por rectas paralelas. La distancia entre rectas es $2a$. Suponiendo que $l < a$, cuál es la probabilidad de que la aguja intersecte alguna de las rectas?

Localizamos la aguja mediante la distancia ρ de su centro a la recta más cercana y el ángulo agudo θ entre la recta y la aguja: $0 \leq \rho \leq a$ y $0 \leq \theta \leq \pi/2$. El rectángulo determinado por esas desigualdades es el espacio muestral Ω. El evento A = *"la aguja interesecta la recta"* ocurre si $\rho \leq l \operatorname{sen} \theta$. La probabilidad de A es el cociente del área de la figura determinada por las tres desigualdades $0 \leq \rho \leq a$, $0 \leq \theta \leq \pi/2$ y $\rho \leq l \operatorname{sen} \theta$ y el área del rectángulo $\pi a/2$.

El área de la figura es $\int_0^{\pi/2} l \operatorname{sen}(\theta) d\theta = l$. Por lo tanto, la probabilidad de intersección es

$$\mathbb{P}(A) = \frac{2l}{\pi a}. \tag{23}$$

La fórmula (23) indica un método aleatorio para estimar π: arrojar la aguja n veces sobre el plano y contar $n(A)$ la cantidad de veces que la aguja interesectó alguna recta:

$$\hat{\pi} = 2(l/a)(n/n(A)).$$

□

5.3. Paradoja de Bertrand

Se dibuja una cuerda aleatoria CD sobre el círculo de radio 1. ¿Cuál es la probabilidad que la longitud de la cuerda CD supere $\sqrt{3}$, la longitud del lado del triángulo equilátero inscripto en dicho círculo?

Este es un ejemplo de un problema planteado de manera incompleta. La pregunta que debe formularse es la siguiente ¿qué significa elegir "aleatoriamente"? Bertrand propuso tres respuestas diferentes a esa pregunta. Las diferentes respuestas corresponden en realidad a diferentes modelos probabilísticos, i.e., diferentes espacios de probabilidad concretos $(\Omega, \mathcal{A}, \mathbb{P})$.

- *Primer modelo.* Sea Ω_1 la bola de radio 1, $\Omega_1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$, con la σ-álgebra $\mathcal{A}$ de los "subconjuntos cuya área está definida". Para cada $A \in \mathcal{A}$,

$$\mathbb{P}_1(A) = \frac{\text{área}(A)}{\text{área}(\Omega)} = \frac{\text{área}(A)}{\pi}.$$

C y D se construyen del siguiente modo: usando la ley de distribución $\mathbb{P}_1$ se sortea un punto ω sobre la bola de radio 1 y CD es perpendicular al segmento $\overline{0\omega}$ cuyos extremos son $(0,0)$ y ω. La longitud de CD es una función de ω que llamaremos $\ell(\omega)$. Queremos calcular $\mathbb{P}_1(\ell(\omega) \geq \sqrt{3})$. Notar que

$$\ell(\omega) \geq \sqrt{3} \iff \text{longitud}(\overline{0\omega}) \geq \frac{1}{2}.$$

Por lo tanto,

$$\mathbb{P}_1(\ell(\omega) \geq \sqrt{3}) = \frac{\pi - \pi/4}{\pi} = \frac{3}{4}.$$

- *Segundo modelo.* Sea Ω_2 el círculo de radio 1, $\Omega_2 = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, con la σ-álgebra $\mathcal{A}$ de los "subconjuntos cuya longitud está definida". Para cada $A \in \mathcal{A}$,

$$\mathbb{P}_2(A) = \frac{\text{longitud}(A)}{\text{longitud}(\Omega))} = \frac{\text{longitud}(A)}{2\pi}.$$

C y D se construyen del siguiente modo: Se *fija* el punto C; con la ley $\mathbb{P}_2$ se sortea un punto ω sobre el círculo de radio 1 y se pone $D = \omega$. La longitud de CD es una una función de ω que llamaremos $\ell(\omega)$. El conjunto $\{\omega : \ell(\omega)) \geq \sqrt{3}\}$ es el segmento del círculo determinado dos vértices del triángulo equilátero inscripto en el círculo, a saber: los del lado opuesto al vértice C. Por lo tanto,

$$\mathbb{P}_2(\ell(\omega) \geq \sqrt{3}) = \frac{2\pi/3}{2\pi} = \frac{1}{3}.$$

- *Tercer modelo.* Sea Ω_3 el intervalo $[0,1]$ con la σ-álgebra $\mathcal{A}$ de los "subconjuntos cuya longitud está definida". Para cada $A \in \mathcal{A}$,

$$\mathbb{P}_3(A) = \text{longitud}(A).$$

C y D se construyen del siguiente modo: se sortea un punto ω sobre el intervalo $[0,1]$ del eje x y CD es la cuerda perpendicular al eje x que pasa por ω. Es claro que,

$$\ell(\omega) \geq \sqrt{3} \iff \omega \in [1/2, 1].$$

Por lo tanto, la tercer respuesta es $1/2$.

Nota Bene. Obtuvimos 3 respuestas diferentes: $1/4, 1/3$ y $1/2$. Sin embargo, no hay porque sorprenderse debido a que los modelos probabilísticos correspondientes a cada respuesta son diferentes. Cuál de los tres es el "bueno" es otro problema. El modelo correcto depende del mecanismo usado para dibujar la cuerda al azar. Los tres mecanismos anteriores son puramente intelectuales, y muy probablemente, no corresponden a ningún mecanismo físico. Para discriminar entre modelos probabilísticos en competencia se debe recurrir al análisis estadístico que esencialmente se basa en dos resultados de la Teoría de Probabilidad: la ley fuerte de los grandes números y el teorema central del límite. $\square$

5.4. De las masas puntuales a la masa continua

Para concluir está sección mostraremos un par de métodos para construir medidas de probabilidad sobre $\mathbb{R}^n$.

Masas puntuales. Tomamos una sucesión de puntos $\{\mathbf{x}_1, \mathbf{x}_2, \dots\}$ en $\mathbb{R}^n$ y una sucesión de números no negativos $\{p(\mathbf{x}_1), p(\mathbf{x}_2), \dots\}$ tales que

$$\sum_{i=1}^{\infty} p(\mathbf{x}_i) = 1$$

y para cada $A \subset \mathbb{R}^n$ definimos $\mathbb{P}(A)$ como la suma de las *"masas puntuales"*, $p(\mathbf{x}_i)$, de los puntos $\mathbf{x}_i$ contenidos en A:

$$\mathbb{P}(A) := \sum_{\mathbf{x}_i \in A} p(\mathbf{x}_i).$$

Nota Bene. El método de las *masas puntuales* puede generalizarse de la siguiente forma: la suma $\sum_{\mathbf{x}_i}$ se reemplaza por la integral $\int d\mathbf{x}$ y las masas puntuales $p(\mathbf{x}_i)$ por una función $\rho(\mathbf{x})$ denominada *densidad de probabilidades*. Esta metodología es de uso común en mecánica: primero se consideran sistemas con masas puntuales discretas donde cada punto tiene masa finita y después se pasa a la noción de distribución de masa continua, donde cada punto tiene masa cero. En el primer caso, la masa total del sistema se obtiene simplemente sumando las masas de los puntos individuales; en el segundo caso, las masas se calculan mediante integración sobre densidades de masa. Salvo por las herramientas técnicas requeridas, no hay diferencias esenciales entre ambos casos. □

Definición 5.4. Una *densidad de probabilidades sobre* $\mathbb{R}^n$ es una función ("más o menos razonable") no negativa $\rho : \mathbb{R}^n \to \mathbb{R}^+$ tal que

$$\int_{\mathbb{R}^n} \rho(\mathbf{x})\, d\mathbf{x} = 1.$$

Masa continua. Tomamos una densidad de probabilidades $\rho : \mathbb{R}^n \to \mathbb{R}^+$ y para cada subconjunto $A \subset \mathbb{R}^n$ ("más o menos razonable") y definimos $\mathbb{P}(A)$ como la integral de la densidad $\rho(\mathbf{x})$ sobre el conjunto A:

$$\mathbb{P}(A) := \int_A \rho(\mathbf{x}) d\mathbf{x}$$

Ejemplo 5.5 (Gaussiana)**.** La función $\rho : \mathbb{R}^2 \to \mathbb{R}^+$ definida por

$$\rho(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right)$$

es una densidad de probabilidades sobre $\mathbb{R}^2$ denominada *gaussiana bidimensional*. En efecto,

$$\begin{aligned}
\iint_{\mathbb{R}^2} 2\pi\rho(x,y)dxdy &= \iint_{\mathbb{R}^2} \exp\left(-\frac{x^2+y^2}{2}\right) dxdy \\
&= 2\iint_{\mathbb{R}^2} \exp\left(-(x^2+y^2)\right) dxdy \\
&= 2\int_0^{2\pi}\left(\int_0^{\infty} e^{-\rho^2}\rho\, d\rho\right) d\theta \\
&= \int_0^{2\pi}\left(\int_0^{\infty} e^{-\rho^2}2\rho\, d\rho\right) d\theta \\
&= 2\pi. \qquad (24)
\end{aligned}$$

Nota Bene. Observando con cuidado las identidades (24) se puede ver que

$$\int_{\mathbb{R}} e^{-x^2/2}dx = \sqrt{2\pi}.$$

Por lo tanto, la función $\varphi : \mathbb{R} \to \mathbb{R}^+$ definida por

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

es una densidad de probabilidades sobre $\mathbb{R}$. $\square$

6. Elementos de Análisis Combinatorio

Cuando se estudian juegos de azar, procedimientos muestrales, problemas de orden y ocupación, se trata por lo general con espacios muestrales finitos Ω en los que a todos los eventos elementales se les atribuye igual probabilidad. Para calcular la probabilidad de un evento A tenemos que dividir la cantidad de eventos elementales contenidos en A (llamados *casos favorables*) entre la cantidad de total de eventos elementales contenidos en Ω (llamados *casos posibles*). Estos cálculos se facilitan por el uso sistemático de unas pocas reglas.

6.1. Regla del Producto

Sean A y B dos conjuntos cualesquiera. El producto cartesiano de A y B se define por $A \times B = \{(a,b) : a \in A \text{ y } b \in B\}$. Si A y B son finitos, entonces $|A \times B| = |A||B|$.

	b_1	b_2	$\dots$	b_n
a_1	(a_1, b_1)	(a_1, b_2)	$\dots$	(a_1, b_n)
a_2	(a_2, b_1)	(a_2, b_2)	$\dots$	(a_2, b_n)
$\vdots$	$\vdots$	$\vdots$		$\vdots$
a_m	(a_m, b_1)	(a_m, b_2)	$\dots$	(a_m, b_n)

Cuadro 1: Esquema rectangular del tipo *tabla de multiplicar* con m filas y n columnas: en la intersección de fila i y la columna j se encuentra el par (a_i, b_j). Cada par aparece una y sólo una vez.

Demostración. Supongamos que $A = \{a_1, a_2, \dots, a_m\}$ y $B = \{b_1, b_2, \dots, b_n\}$. Basta observar el Cuadro siguiente

En palabras, con m elementos $a_1, \dots, a_m$ y n elementos $b_1, \dots, b_n$ es posible formar $m \cdot n$ pares (a_i, b_j) que contienen un elemento de cada grupo. □

En general vale el siguiente

Teorema 6.1 (Regla del producto)**.** *Dados n_1 elementos $a_1, \dots, a_{n_1}$, y n_2 elementos $b_1, \dots, b_{n_2}$, etc., hasta n_k elementos $x_1, \dots, x_{n_k}$; es posible formar $\prod_{i=1}^{k} n_i = n_1 n_2 \cdots n_k$ k-uplas ordenadas $(a_{i_1}, b_{i_2}, \dots, x_{i_k})$ que contienen un elemento de cada tipo.*

Demostración. Si $k = 2$ ya lo demostramos. Si $k = 3$, tomamos los pares (a_i, b_j) como elementos de un nuevo tipo. Hay $n_1 n_2$ elementos de ese tipo y n_3 elementos c_l. Cada terna (a_i, b_j, c_l) es un par formado por un elemento (a_i, b_j) y un elemento c_l; la cantidad de ternas es por lo tanto $n_1 n_2 n_3$. Etcétera. □

Ejemplo 6.2 (Clasificaciones múltiples)**.** Supongamos que las personas se clasifican de acuerdo a sexo, estado civil, y profesión. Las distintas categorías juegan el papel de los elementos. Si hay 17 profesiones, entonces tenemos $2 \cdot 4 \cdot 17 = 136$ clases en total. □

Ejemplo 6.3 (Distribuir r bolas en n cajas)**.** Los resultados posibles del experimento pueden representarse mediante el conjunto

$$\Omega = \{1, 2, \dots, n\}^r = \{(x_1, x_2, \dots, x_r) : x_i \in \{1, 2, \dots, n\}, 1 \leq i \leq r\},$$

donde $x_i = j$ representa el resultado *"la bola i en la caja j"*. El problema es equivalente al de elegir una caja para cada bola. Con r bolas tenemos r elecciones independientes, y en consecuencia *r bolas pueden ubicarse en n cajas de n^r formas distintas.*

Usamos el lenguaje figurado de bolas y cajas, pero el mismo espacio muestral admite muchas interpretaciones distintas. Para ilustrar el asunto *listaremos una cantidad de situaciones en las que aunque varía el contenido intuitivo son todas abstractamente equivalentes al esquema de poner r bolas en n cajas, en el sentido que los resultados difieren solamente en su descripción verbal.*

1. *Nacimientos.* Las configuraciones posibles de los nacimientos de r personas corresponde a los diferentes arreglos de r bolas en $n = 365$ cajas (suponiendo que el año tiene 365 días).

2. *Accidentes.* Clasificar r accidentes de acuerdo con el día de la semana en que ocurrieron es equivalente a poner r bolas en $n = 7$ cajas.

3. *Muestreo.* Un grupo de personas se clasifica de acuerdo con, digamos, edad o profesión. Las clases juegan el rol de las cajas y las personas el de las bolas.

4. *Dados.* Los posibles resultados de una tirada de r dados corresponde a poner r bolas en $n = 6$ cajas. Si en lugar de dados se lanzan monedas tenemos solamente $n = 2$ cajas.

5. *Dígitos aleatorios.* Los posibles ordenamientos de una sucesión de r dígitos corresponden a las distribuciones de r bolas (= lugares) en diez cajas llamadas $0, 1, \dots, 9$.

6. *Coleccionando figuritas.* Los diferentes tipos de figuritas representan las cajas, las figuritas coleccionadas representan las bolas.

□

6.2. Muestras ordenadas

Se considera una "población" de n elementos $a_1, a_2, \dots, a_n$. Cualquier *secuencia ordenada* $a_{j_1}, a_{j_2}, \dots, a_{j_k}$ de k símbolos se llama una *muestra ordenada de volumen* k tomada de la población. (Intuitivamente los elementos se pueden elegir uno por uno). Hay dos procedimientos posibles.

(a) Muestreo con reposición. Cada elección se hace entre toda la población, por lo que cada elemento puede escogerse más de una vez. Cada uno de los k elementos puede elegirse en n formas: la cantidad de muestras posibles es, por lo tanto, n^k, lo que resulta de la regla del producto con $n_1 = n_2 = \cdots = n_k = n$.

(b) Muestreo sin reposición. Una vez elegido, el elemento se quita de la población, de modo que las muestras son arreglos sin repeticiones. El volumen de la muestra k no puede exceder el tamaño de la población total n.

Tenemos n posibles elecciones para el primer elemento, pero sólo $n-1$ para el segundo, $n-2$ para el tercero, etcétera. Usando la regla del producto se obtiene un total de

$$(n)_k := n(n-1)(n-2)\cdots(n-k+1) \tag{25}$$

elecciones posibles.

Teorema 6.4. *Para una población de n elementos y un tamaño de muestra prefijado k, existen n^k diferentes muestras con reposición y $(n)_k$ muestras sin reposición.*

Ejemplo 6.5. Consideramos una urna con 8 bolas numeradas $1, 2, \dots, 8$

(a) **Extracción con reposición.** Extraemos 3 bolas *con reposición*, esto es, después de extraer una bola, anotamos su número y la ponemos de nuevo en la urna. El espacio muestral Ω_1 correspondiente a este experimento consiste de todas las secuencias de longitud 3 que pueden formarse con los símbolos $1, 2, \dots 8$. De acuerdo con el Teorema 6.4, Ω_1 tiene $8^3 = 512$ elementos. Bajo la hipótesis que todos los elementos tienen la misma probabilidad, la probabilidad de observar la secuencia $(4, 4, 8)$ es $1/512$.

(b) **Extracción de una colección ordenada sin reposición.** Extraemos 3 bolas *sin reposición*, esto es, cada bola elegida *no* se vuelve a poner en la urna. Anotamos los números de las bolas elegidas, en orden. El espacio muestral Ω_2 correspondiente a este experimento es el conjunto de todas las secuencias de longitud 3 que pueden formarse con los símbolos $1, 2 \dots, 8$ donde cada símbolo puede aparecer a los sumo una vez. De acuerdo con el Teorema 6.4, Ω_2 tiene $(8)_3 = 8 \times 7 \times 6 = 336$ elementos. Bajo la hipótesis que todos los elementos tienen la misma probabilidad, la probabilidad de observar la secuencia $(3, 7, 1)$ (en ese orden) es $1/336$.

$\square$

Ejemplo 6.6. Una urna contiene 6 bolas rojas y 4 bolas negras. Se extraen 2 bolas con reposición. Para fijar ideas supongamos que las bolas están numeradas de la siguiente manera: las primeras 6 son las rojas y las últimas 4 son las negras. El espacio muestral asociado es $\Omega = \{1, \dots, 10\}^2$ y su cantidad de elementos $|\Omega| = 10^2$.

(a) *¿Cuál es la probabilidad de que todas sean rojas?* Sea R el evento "todas rojas", $R = \{1, \dots, 6\}^2$ y $|R| = 6^2$. Por lo tanto, $\mathbb{P}(R) = 6^2/10^2 = 0.36$.

(b) *¿Cuál es la probabilidad de que todas sean del mismo color?* Sea N el evento "todas negras", $N = \{7, \dots, 10\}^2$ y $|N| = 4^2$, entonces $\mathbb{P}(N) = 4^2/10^2 = 0.16$. Por lo tanto, $\mathbb{P}(R \cup N) = \mathbb{P}(R) + \mathbb{P}(N) = 0.52$.

(c) *¿Cuál es la probabilidad de que al menos una sea roja?* El evento "al menos una roja" es el complemento de "todas negras". Por lo tanto, $\mathbb{P}(N^c) = 1 - \mathbb{P}(N) = 0.84$.

Si se consideran extracciones sin reposición, deben reemplazarse las cantidades $(10)^2$, 6^2 y 4^2 por las correspondientes $(10)_2$, $(6)_2$ y $(4)_2$. $\square$

Caso especial $k = n$. En muestreo sin reposición una muestra de tamaño n incluye a toda la población y representa una *permutación* de sus elementos. En consecuencia, n elementos $a_1, a_2, \dots, a_n$ pueden ordenarse en $(n)_n = n \cdot (n-1) \cdots 2 \cdot 1$ formas distintas. Usualmente el número $(n)_n$ se denota $n!$ y se llama el *factorial de n*.

Corolario 6.7. *La cantidad de formas distintas en que pueden ordenarse n elementos es*

$$n! = 1 \cdot 2 \cdots n. \tag{26}$$

Observación 6.8. Las muestras ordenadas de tamaño k, sin reposición, de una población de n elementos, se llaman *variaciones* de n elementos tomados de a k. Su número total $(n)_k$ puede calcularse del siguiente modo

$$(n)_k = \frac{n!}{(n-k)!} \tag{27}$$

Nota Bene sobre muestreo aleatorio. Cuando hablemos de *"muestras aleatorias de tamaño k"*, el adjetivo aleatorio indica que todas las muestras posibles tienen la misma probabilidad, a saber: $1/n^k$ en muestreo con reposición y $1/(n)_k$ en muestreo sin reposición. En ambos casos, n es el tamaño de la población de la que se extraen las muestras.

Si n es grande y k es relativamente pequeño, el cociente $(n)_k/n^k$ está cerca de la unidad. En otras palabras, para grandes poblaciones y muestras relativamente pequeñas, las dos formas de muestrear son prácticamente equivalentes.

Ejemplos

Consideramos muestras aleatorias de volumen k (*con reposición*) tomadas de una población de n elementos $a_1, \dots, a_n$. Nos interesa el evento que en una muestra no se repita ningún elemento. En total existen n^k muestras diferentes, de las cuales $(n)_k$ satisfacen la condición estipulada. Por lo tanto, *la probabilidad de ninguna repetición en nuestra muestra es*

$$p = \frac{(n)_k}{n^k} = \frac{n(n-1)\cdots(n-k+1)}{n^k} \tag{28}$$

Las interpretaciones concretas de la fórmula (28) revelan aspectos sorprendentes.

Muestras aleatorias de números. La población consiste de los diez dígitos $0, 1, \dots, 9$. Toda sucesión de cinco dígitos representa una muestra de tamaño $k = 5$, y supondremos que cada uno de esos arreglos tiene probabilidad 10^{-5}. *La probabilidad de que* 5 *dígitos aleatorios sean todos distintos es* $p = (10)_5 10^{-5} = 0.3024$.

Bolas y cajas. *Si n bolas se ubican aleatoriamente en n cajas, la probabilidad de que cada caja este ocupada es*

$$p = \frac{n!}{n^n}.$$

Interpretaciones:

(a) Para $n = 7$, $p = 0.00612....$ Esto significa que *si en una ciudad ocurren* 7 *accidentes por semana, entonces (suponiendo que todas las distribuciones posibles son igualmente probables) prácticamente todas las semanas contienen días con dos o más accidentes, y en promedio solo una semana de* 164 *mostrará una distribución uniforme de un accidente por día.*

(b) Para $n = 6$ la probabilidad p es igual a 0.01543... Esto muestra cuan extremadamente improbable es que en seis tiradas de un dado perfecto aparezcan todas las caras.

Cumpleaños. Los cumpleaños de k personas constituyen una muestra de tamaño k de la población formada por todos los días del año.

De acuerdo con la ecuación (28) *la probabilidad, p_k, de que todos los k cumpleaños sean diferentes es*

$$p_k = \frac{(365)_k}{365^k} = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\cdots\left(1 - \frac{k-1}{365}\right).$$

Una fórmula aparentemente abominable. Si $k = 23$ tenemos $p_k < 1/2$. En palabras, *para 23 personas la probabilidad que al menos dos personas tengan un cumpleaños común excede 1/2.*

Aproximaciones numéricas de p_k. Si k es chico, tomando logaritmos y usando que para x pequeño y positivo $\log(1 - x) \sim -x$, se obtiene

$$\log p_k \sim -\frac{1 + 2 + \cdots + (k-1)}{365} = -\frac{k(k-1)}{730}.$$

Ejercicios adicionales

7. Hallar la probabilidad p_k de que en una muestra de k dígitos aleatorios no haya dos iguales. Estimar el valor numérico de p_{10} usando la *fórmula de Stirling (1730)*: $n! \sim e^{-n} n^{n+\frac{1}{2}}\sqrt{2\pi}$.

8. Considerar los primeros 10000 decimales del número π. Hay 2000 grupos de cinco dígitos. Contar la cantidad de grupos en los que los 5 dígitos son diferentes e indicar la frecuencia relativa del evento considerado. Comparar el resultado obtenido con la probabilidad de que en una muestra de 5 dígitos aleatorios no haya dos iguales.

6.3. Subpoblaciones

En lo que sigue, utilizaremos el término *población de tamaño n* para designar una colección de n elementos *sin considerar su orden*. Dos poblaciones se consideran diferentes si una de ellas contiene algún elemento que no está contenido en la otra.

Uno de los problemas más importantes del cálculo combinatorio es *determinar la cantidad $C_{n,k}$ de subpoblaciones distintas de tamaño k que tiene una población de tamaño n.* Cuando n y k son pequeños, el problema se puede resolver por enumeración directa. Por ejemplo, hay seis formas distintas elegir dos letras entre cuatro letras A, B, C, D, a saber:

AB, AC, AD, BC, BD, CD. Así, $C_{4,2} = 6$. Cuando la cantidad de elementos de la colección es grande la enumeración directa es impracticable. El problema general se resuelve razonando de la siguiente manera:

Consideramos una subpoblación de tamaño k de una población de n elementos. Cada numeración arbitraria de los elementos de la subpoblación la convierte en una muestra ordenada de tamaño k. Todas las muestras ordenadas de tamaño k se pueden obtener de esta forma. Debido a que k elementos pueden ordenarse en $k!$ formas diferentes, resulta que $k!$ veces la cantidad de subpoblaciones de tamaño k coincide con la cantidad de muestras ordenadas de dicho tamaño. En otros términos, $C_{n,k} \cdot k! = (n)_k$. Por lo tanto,

$$C_{n,k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}. \tag{29}$$

Los números definidos en (29) se llaman *coeficientes binomiales* o *números combinatorios* y la notación clásica para ellos es $\binom{n}{k}$.

Teorema 6.9. *Una población de n elementos tiene*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{30}$$

diferentes subpoblaciones de tamaño $k \leq n$.

Ejemplo 6.10. Consideramos una urna con 8 bolas numeradas $1, 2, \ldots, 8$. Extraemos 3 bolas simultáneamente, de modo que el orden es irrelevante. El espacio muestral Ω_3 correspondiente a este experimento consiste de todos los subconjuntos de tamaño 3 de un conjunto de 8 elementos. De acuerdo con el Teorema 6.9, Ω_3 tiene $\binom{8}{3} = 56$ elementos. Bajo la hipótesis que todos los elementos tienen la misma probabilidad, la probabilidad de seleccionar $\{3, 7, 1\}$ es $1/56$. □

Dada una población de tamaño n podemos elegir una subpoblación de tamaño k de $\binom{n}{k}$ maneras distintas. Ahora bien, elegir los k elementos que vamos a quitar de una población dada es lo mismo que elegir los $n-k$ elementos que vamos a dejar dentro. Por lo tanto, es claro que para cada $k \leq n$ debe valer

$$\binom{n}{k} = \binom{n}{n-k}. \tag{31}$$

La ecuación (31) se deduce inmediatamente de la identidad (30). El lado izquierdo de la ecuación (31) no está definido para $k = 0$, pero el lado derecho si lo está. Para que la ecuación (31) sea valida para todo entero k tal que $0 \leq k \leq n$, *se definen*

$$\binom{n}{0} := 1, \qquad 0! := 1, \qquad \text{y} \qquad (n)_0 := 1.$$

Triángulo de Pascal. Las ecuaciones en diferencias

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}, \tag{32}$$

junto con el conocimiento de los datos de borde

$$\binom{n}{0} = \binom{n}{n} = 1, \tag{33}$$

determinan completamente los números combinatorios $\binom{n}{k}$, $0 \leq k \leq n$, $n = 0, 1, \ldots$. Usando dichas relaciones se construye el famoso *"triángulo de Pascal"*, que muestra todos los números combinatorios en la forma de un triángulo

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 \\
\ldots & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots & &
\end{array}$$

La n-ésima fila de este triángulo contiene los coeficientes $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$. Las condiciones de borde (33) indican que el primero y el último de esos números son 1. Los números restantes se determinan por la ecuación en diferencias (32). Vale decir, para cada $0 < k < n$, el k-ésimo coeficiente de la n-ésima fila del "triángulo de Pascal" se obtiene *sumando* los dos coeficientes inmediatamente superiores a izquierda y derecha. Por ejemplo, $\binom{5}{2} = 4+6 = 10$.

Control de calidad. Una planta de ensamblaje recibe una partida de 50 piezas de precisión que incluye 4 defectuosas. La división de control de calidad elige 10 al azar para controlarlas y rechaza la partida si encuentra 1 o más defectuosas. ¿Cuál es la probabilidad de que la partida pase la inspección? Hay $\binom{50}{10}$ formas de elegir la muestra para controlar y $\binom{46}{10}$ de elegir todas las piezas sin defectos. Por lo tanto, la probabilidad es

$$\binom{46}{10}\binom{50}{10}^{-1} = \frac{46!}{10!36!}\frac{10!40!}{50!} = \frac{40 \cdot 39 \cdot 38 \cdot 37}{50 \cdot 49 \cdot 48 \cdot 47} = 0,3968....$$

Usando cálculos casi idénticos una compañía puede decidir sobre qué cantidad de piezas defectuosas admite en una partida y diseñar un programa de control con una probabilidad dada de éxito. $\square$

Ejercicios adicionales

9. Considerar el siguiente juego: el jugador I tira 4 veces una moneda honesta y el jugador II lo hace 3 veces. Calcular la la probabilidad de que el jugador I obtenga más caras que el jugador II.

Ejemplicio sobre el juego de poker

Para calcular las probabilidades de las manos de poker comenzamos observando que hay $\binom{52}{5} = 2598960$ formas distintas de elegir 5 naipes de un mazo de 52. El cálculo de probabilidades se reduce a calcular la cantidad de formas distintas en que puede ocurrir cada mano. Calcularemos algunas para ilustrar las ideas principales.

- *un par:* $13 \cdot \binom{4}{2} \cdot \binom{12}{3} \cdot 4^3 = 1098240$. Primero elegimos el valor para el par (13 formas), después les asignamos los palos ($\binom{4}{2}$ formas), luego elegimos tres valores para las otras cartas ($\binom{12}{3}$ formas) y finalmente les asignamos los palos (4^3 formas).
- *escalera:* $10 \cdot 4^5 = 10240$. Una escalera puede empezar con una carta mayor o igual que 5, hay 10 posibilidades. Una vez que los valores fueron determinados, hay 4^5 formas de asignar los palos. Esta forma de contar considera a la escalera de color como escalera. Si se quieren excluir las escaleras de color, los palos pueden asignarse en $4^5 - 4$ maneras.
- *full:* $13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2} = 3744$. Primero elegimos el valor para la terna (que puede hacerse de 13 formas), después les asignamos palos ($\binom{4}{3}$ formas), luego elegimos el valor para el par (12 formas), finalmente asignamos les asignamos palos ($\binom{4}{2}$ formas).
- *Manos perdedoras.* Cinco valores diferentes que no forman color ni escalera. La cantidad de manos perdedoras es $\left(\binom{13}{5} - 10\right) \cdot (4^5 - 4)$.

(a) *un par*	0.422569
(b) *dos pares*	0.047539
(c) *terna*	0.021128
(d) *escalera*	0.003940
(e) *color*	0.001981
(f) *full*	0.001441
(g) *poker*	0.000240
(h) *escalera de color*	0.000015
(i) *nada*	0.501177

Cuadro 2: Valores numéricos de todas las manos del poker. Observar que (d) incluye a (h).

□

6.4. Particiones

Teorema 6.11. *Sean $r_1, \dots, r_k$ enteros tales que*

$$r_1 + r_2 + \cdots + r_k = n, \quad r_i \geq 0. \tag{34}$$

El número de formas en que una población de n elementos puede dividirse en k partes ordenadas (particionarse en k subpoblaciones) tales que la primera contenga r_1 elementos, la segunda r_2, etc, es

$$\frac{n!}{r_1! r_2! \cdots r_k!}. \tag{35}$$

Los números (35) se llaman coeficientes multinomiales.

Demostración. Un uso repetido de (30) muestra que el número (35) puede reescribirse en la forma

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\binom{n-r_1-r_2}{r_3}\cdots\binom{n-r_1-\cdots-r_{k-2}}{r_{k-1}} \tag{36}$$

Por otro lado, para efectuar la partición deseada, tenemos primero que seleccionar r_1 elementos de los n; de los restantes $n - r_1$ elementos seleccionamos un segundo grupo de tamaño r_2, etc. Después de formar el grupo $(k-1)$ quedan $n - r_1 - r_2 - \cdots - r_{k-1} = r_k$ elementos, y esos forman el último grupo. Concluimos que (36) representa el número de formas en que la operación puede realizarse. □

Ejercicios adicionales

10. ¿Cuántas palabras distintas pueden formarse permutando las letras de la palabra *"manzana"* y cuántas permutando las letras de la palabra *"aiaiiaiiiaiiii"*?

6.5. Distribución Hipergeométrica

Muchos problemas combinatorios pueden reducirse a la siguiente forma. En una urna hay n_1 bolas rojas y n_2 bolas negras. Se elige al azar un grupo de r bolas. Se quiere calcular la probabilidad p_k de que en el grupo elegido, haya exactamente k bolas rojas, $0 \leq k \leq \mathrm{mín}(n_1, r)$.

Para calcular p_k, observamos que el grupo elegido debe contener k bolas rojas y $r-k$ negras. Las rojas pueden elegirse de $\binom{n_1}{k}$ formas distintas y la blancas de $\binom{n_2}{r-k}$ formas distintas. Como cada elección de las k bolas rojas debe combinarse con cada elección de las $r-k$ negras, se obtiene

$$p_k = \binom{n_1}{k}\binom{n_2}{r-k}\binom{n_1+n_2}{r}^{-1} \tag{37}$$

Control de calidad.

En control de calidad industrial, se someten a inspección lotes de n unidades. Las unidades defectuosas juegan el rol de las bolas rojas y su cantidad n_1 es desconocida. Se toma una muestra de tamaño r y se determina la cantidad k de unidades defectuosas. La fórmula (37) permite hacer inferencias sobre la cantidad desconocida n_1; se trata de problema típico de estimación estadística que será analizado más adelante. $\square$

Ejemplo 6.12. Una planta de ensamblaje recibe una partida de 100 piezas de precisión que incluye exactamente 8 defectuosas. La división control de calidad elige 10 piezas al azar para controlarlas y rechaza la partida si encuentra al menos 2 defectuosas. ¿Cuál es la probabilidad de que la partida pase la inspección?

El criterio de decisión adoptado indica que la partida pasa la inspección si (y sólo si) en la muestra no se encuentran piezas defectuosas o si se encuentra exactamente una pieza defectuosa. Hay $\binom{100}{10}$ formas de elegir la muestra para controlar, $\binom{92}{10}\binom{8}{0}$ formas de elegir muestras sin piezas defectos y $\binom{92}{9}\binom{8}{1}$ formas de elegir muestras con exactamente una pieza defectuosa. En consecuencia la probabilidad de que la partida pase la inspección es

$$\binom{92}{10}\binom{8}{0}\binom{100}{10}^{-1} + \binom{92}{9}\binom{8}{1}\binom{100}{10}^{-1} \approx 0.818.$$

$\square$

Ejemplo 6.13. Una planta de ensamblaje recibe una partida de 100 piezas de precisión que incluye exactamente k defectuosas. La división control de calidad elige 10 piezas al azar para controlarlas y rechaza la partida si encuentra al menos 2 defectuosas. ¿Con ese criterio de decisión, cómo se comporta la probabilidad $p(k)$ de que la partida pase la inspección?.

Con este criterio de decisión una partida pasará la inspección si (y sólo si) al extraer una muestra de control la cantidad de piezas defectuosas encontradas es 0 o 1. Hay $\binom{100}{10}$ formas de elegir la muestra para controlar. Para cada $k = 1, \dots, 90$ hay $\binom{100-k}{10-k}\binom{k}{0}$ formas de elegir muestras sin piezas defectos y $\binom{100-k}{9}\binom{k}{1}$ formas de elegir muestras con exactamente una pieza defectuosa. En consecuencia la probabilidad $p(k)$ de que la partida pase la inspección es

$$p(k) = \binom{100-k}{10}\binom{k}{0}\binom{100}{10}^{-1} + \binom{100-k}{9}\binom{k}{1}\binom{100}{10}^{-1}.$$

Una cuenta sencilla muestra que para todo $k = 1, \dots, 90$ el cociente $\frac{p(k)}{p(k-1)}$ es menor que 1. Esto significa que a medida que aumenta la cantidad de piezas defectuosas en la partida, la probabilidad de aceptarla disminuye.

¿Cuál es la máxima probabilidad de aceptar una partida de 100 que contenga más de 20 piezas defectuosas? Debido a que la función $p(k)$ es decreciente, dicha probabilidad es $p(20) \approx 0.3630$. $\square$

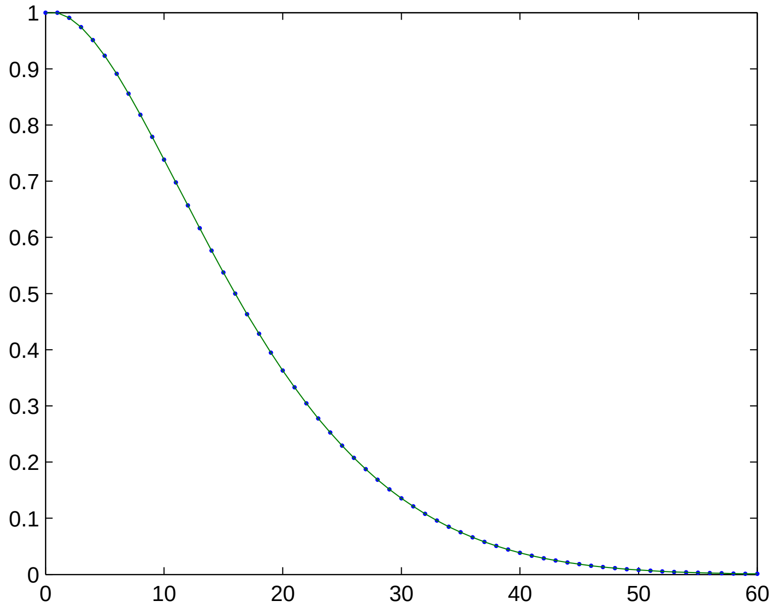


Figura 9: Gráfico de función $p(k)$.

Ejemplo 6.14. Una planta de ensamblaje recibe un lote de $n = 100$ piezas de precisión, de las cuales una cantidad desconocida n_1 son defectuosas. Para controlar el lote se elige una muestra (sin reposición) de $r = 10$ piezas. Examinadas estas, resultan $k = 2$ defectuosas. ¿Qué se puede decir sobre la cantidad de piezas defectuosas en el lote?

Sabemos que de 10 piezas examinadas 2 son defectuosas y 8 no lo son. Por lo tanto, $2 \leq n_1 \leq 92$. Esto es todo lo que podemos decir con absoluta certeza. Podría suponerse que el lote contiene 92 piezas defectuosas. Partiendo de esa hipótesis, llegamos a la conclusión de que ha ocurrido un evento de probabilidad

$$\binom{8}{8}\binom{92}{2}\binom{100}{10}^{-1} = O(10^{-10}).$$

En el otro extremo, podría suponerse que el lote contiene exactamente 2 piezas defectuosas, en ese caso llegamos a la conclusión de que ha ocurrido un evento de probabilidad

$$\binom{98}{8}\binom{2}{2}\binom{100}{10}^{-1} = \frac{1}{110}.$$

Las consideraciones anteriores conducen a buscar el valor de n_1 que maximice la probabilidad

$$p(n_1) := \binom{100 - n_1}{8}\binom{n_1}{2}\binom{100}{10}^{-1},$$

puesto que para ese valor de n_1 nuestra observación tendría la mayor probabilidad de ocurrir. Para encontrar ese valor consideramos el cociente $\frac{p(n_1)}{p(n_1-1)}$. Simplificando los factoriales, obtenemos

$$\begin{aligned}
\frac{p(n_1)}{p(n_1 - 1)} \quad &= \quad \frac{n_1(93 - n_1)}{(n_1 - 2)(101 - n_1)} > 1 \\
&\iff \quad n_1(93 - n_1) > (n_1 - 2)(101 - n_1) \\
&\iff \quad n_1 < 20.2 \iff n_1 \leq 20.
\end{aligned}$$

Esto significa que cuando n_1 crece la sucesión $p(n_1)$ primero crece y después decrece; alcanza su máximo cuando $n_1 = 20$. Suponiendo que $n_1 = 20$, la probabilidad de que en una muestra de 10 piezas extraídas de un lote de 100 se observen 2 defectuosas es:

$$p(20) = \binom{80}{8}\binom{20}{2}\binom{100}{10}^{-1} \approx 0.318.$$

Aunque el verdadero valor de n_1 puede ser mayor o menor que 20, si se supone que $n_1 = 20$ se obtiene un resultado consistente con el sentido común que indicaría que los eventos observables deben tener "alta probabilidad".

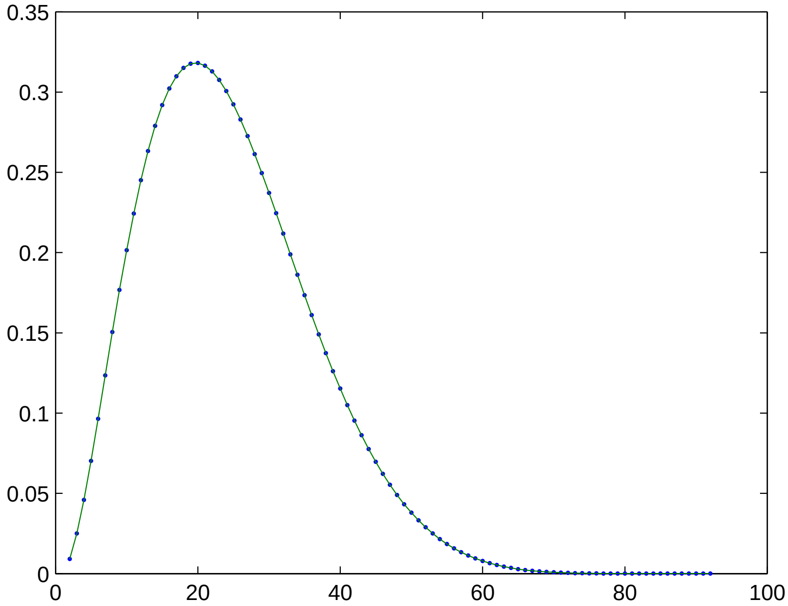


Figura 10: Gráfico de función $p(n_1)$. Observar que $\arg\max\{p(n_1) : 2 \leq n_1 \leq 92\} = 20$. $\square$

Estimación por captura y recaptura.

Para estimar la cantidad n de peces en un lago se puede realizar el siguiente procedimiento. En el primer paso se capturan n_1 peces, que luego de marcarlos se los deja en libertad. En el segundo paso se capturan r peces y se determina la cantidad k de peces marcados. La fórmula (37) permite hacer inferencias sobre la cantidad desconocida n.

Ejemplo 6.15 (Experimentos de captura y recaptura)**.** Se capturan 1000 peces en un lago, se marcan con manchas rojas y se los deja en libertad. Después de un tiempo se hace una nueva captura de 1000 peces, y se encuentra que 100 tienen manchas rojas. ¿Qué conclusiones pueden hacerse sobre la cantidad de peces en el lago?

Suponemos que las dos capturas pueden considerarse como muestras aleatorias de la población total de peces en el lago. También vamos a suponer que la cantidad de peces en el lago no cambió entre las dos capturas.

Generalizamos el problema admitiendo tamaños muestrales arbitrarios. Sean

- $n =$ el número (desconocido) de peces en el lago.
- $n_1 =$ el número de peces en la primera captura. Estos peces juegan el rol de las bolas rojas.
- $r =$ el número de peces en la segunda captura.

- $k =$ el número de peces rojos en la segunda captura.
- $p_k(n) =$ la probabilidad de que la segunda captura contenga exactamente k peces rojos.

Con este planteo la probabilidad $p_k(n)$ se obtiene poniendo $n_2 = n - n_1$ en la fórmula (37):

$$p_k(n) = \binom{n_1}{k}\binom{n-n_1}{r-k}\binom{n}{r}^{-1}. \tag{38}$$

En la práctica n_1, r, y k pueden observarse, pero n es desconocido.

Notar que n es un número fijo que no depende del azar. Resultaría insensato preguntar por la probabilidad que n sea mayor que, digamos, 6000.

Sabemos que fueron capturados n_1+r-k peces diferentes, y por lo tanto $n \geq n_1+r-k$. Esto es todo lo que podemos decir con absoluta certeza. En nuestro ejemplo tenemos $n_1 = r = 1000$ y $k = 100$, y podría suponerse que el lago contiene solamente 1900 peces. Sin embargo, partiendo de esa hipótesis, llegamos a la conclusión de que ha ocurrido un evento de probabilidad fantásticamente pequeña. En efecto, si se supone que hay un total de 1900 peces, la fórmula (38) muestra que la probabilidad de que las dos muestras de tamaño 1000 agoten toda la población es ,

$$\binom{1000}{100}\binom{900}{900}\binom{1900}{1000}^{-1} = \frac{(1000!)^2}{100!1900!}$$

La fórmula de Stirling muestra que esta probabilidad es del orden de magnitud de 10^{-430}, y en esta situación el sentido común indica rechazar la hipótesis como irrazonable. Un razonamiento similar nos induce a rechazar la hipótesis de que n es muy grande, digamos, un millón.

Las consideraciones anteriores nos conducen a buscar el valor de n que maximice la probabilidad $p_k(n)$, puesto que para ese n nuestra observación tendría la mayor probabilidad de ocurrir. Para cualquier conjunto de observaciones n_1, r, k, el valor de n que maximiza la probabilidad $p_k(n)$ se denota por $\hat{n}_{mv}$ y se llama el *estimador de máxima verosimilitud* de n. Para encontrar $\hat{n}_{mv}$ consideramos la proporción

$$\begin{aligned}
\frac{p_k(n)}{p_k(n-1)} \quad &= \quad \frac{(n-n_1)(n-r)}{(n-n_1-r+k)n} > 1 \\
&\iff \quad (n-n_1)(n-r) > (n-n_1-r+k)n \\
&\iff \quad n^2 - nn_1 - nr + n_1 r > n^2 - nn_1 - nr + nk \\
&\iff \quad n < \frac{n_1 r}{k}.
\end{aligned}$$

Esto significa que cuando n crece la sucesión $p_k(n)$ primero crece y después decrece; alcanza su máximo cuando n es el mayor entero menor que $\frac{n_1 r}{k}$, así que $\hat{n}_{mv}$ es aproximadamente igual a $\frac{n_1 r}{k}$. En nuestro ejemplo particular el estimador de máxima verosimilitud del número de peces en el lago es $\hat{n}_{mv} = 10000$.

El verdadero valor de n puede ser mayor o menor, y podemos preguntar por los límites entre los que resulta razonable esperar que se encuentre n. Para esto testeamos la hipótesis que n sea menos que 8500. Sustituimos en (38) $n = 8500, n_1 = r = 1000$, y calculamos la probabilidad que la segunda muestra contenga 100 o menos peces rojos. Esta probabilidad es $p = p_0 + p_1 + \cdots + p_{100}$. Usando una computadora encontramos que $p \approx 0.04$. Similarmente, si $n = 12.000$, la probabilidad que la segunda muestra contenga 100 o más peces rojos esta cerca de 0.03. Esos resultados justificarían la apuesta de que el verdadero número n de peces se encuentra en algún lugar entre 8500 y 12.000.

$\square$

Ejercicios adicionales

11. Un estudiante de ecología va a una laguna y captura 60 escarabajos de agua, marca cada uno con un punto de pintura y los deja en libertad. A los pocos días vuelve y captura otra muestra de 50, encontrando 12 escarabajos marcados. ¿Cuál sería su mejor apuesta sobre el tamaño de la población de escarabajos de agua en la laguna?

---

7. Un problema de Mecánica Estadística

El espacio se divide en una gran cantidad, n, de pequeñas regiones llamadas celdas. Se considera un sistema mecánico compuesto por r partículas que se distribuyen al azar entre las n celdas. ¿Cuál es la distribución de las partículas en las celdas? La respuesta depende de lo que se considere un evento elemental.

1. *Estadística de Maxwell-Boltzmann.* Suponemos que todas las partículas son distintas y que todas las ubicaciones de las partículas son igualmente posibles. Un evento elemental está determinado por la r-upla $(x_1, x_2, \ldots, x_r)$, donde x_i es el número de la celda en la que cayó la partícula i. Puesto que cada x_i puede tomar n valores distintos, el número de tales r-uplas es n^r. La probabilidad de un evento elemental es $1/n^r$.

2. *Estadística de Bose-Einstein.* Las partículas son indistinguibles. De nuevo, todas las ubicaciones son igualmente posibles. Un evento elemental está determinado por la n-upla $(r_1, \ldots, r_n)$, donde $r_1 + \cdots + r_n = r$ y r_i es la cantidad de partículas en la i-ésima celda, $1 \le i \le n$. La cantidad de tales n-uplas se puede calcular del siguiente modo: a cada n- upla $(r_1, r_2, \ldots, r_n)$ la identificamos con una sucesión de unos y ceros $s_1, \ldots, s_{r+n-1}$ con ceros en las posiciones numeradas $r_1 + 1, r_1 + r_2 + 2, \ldots, r_1 + r_2 + \cdots + r_{n-1} + n - 1$ (hay $n - 1$ de ellas) y unos las restantes posiciones. La cantidad de tales sucesiones es igual al número de combinaciones de $r + n - 1$ cosas tomadas de a $n - 1$ por vez. La probabilidad de un evento elemental es $1/\binom{r+n-1}{n-1}$.

3. *Estadística de Fermi-Dirac*. En este caso $r < n$ y cada celda contiene a lo sumo una partícula. La cantidad de eventos elementales es $\binom{n}{r}$. La probabilidad de un evento elemental es $1/\binom{n}{r}$.

A modo de ejemplo calcularemos, para cada una de las tres estadísticas mencionadas, la probabilidad de que una celda dada (digamos, la número 1) no tenga partícula. En cada uno de los tres casos la cantidad de eventos elementales favorables es igual a la cantidad de ubicaciones de las partículas en $n-1$ celdas. Por lo tanto, designando por p_1, p_2, p_3 las probabilidades del evento especificado para cada una de las estadísticas (siguiendo el orden de exposición), tenemos que

$$\begin{aligned} p_1 &= \frac{(n-1)^r}{n^r} = \left(1-\frac{1}{n}\right)^r, \\ p_2 &= \binom{r+n-2}{n-2}\binom{r+n-1}{n-1}^{-1} = \frac{n-1}{N+n-1}, \\ p_3 &= \binom{n-1}{r}\binom{n}{r}^{-1} = 1-\frac{r}{n}. \end{aligned}$$

Si $r/n = \alpha$ y $n \to \infty$, entonces

$$p_1 = e^{-\alpha}, \qquad p_2 = \frac{1}{1+\alpha}, \qquad p_3 = 1-\alpha.$$

Si α es pequeño, esas probabilidades coinciden hasta $O(\alpha^2)$. El número α caracteriza la *"densidad promedio"* de las partículas. □

Ejemplo 7.1. Se distribuyen 5 partículas en 10 celdas numeradas $1, 2, \dots, 10$. Calcular, para cada una de las tres estadísticas, la probabilidad de que las celdas 8, 9 y 10 no tengan partículas y que la celdas 6 y 7 tengan exactamente una partícula cada una.

1. *Maxwell-Boltzmann*. Las bolas son distinguibles y todas las configuraciones diferentes son equiprobables. La probabilidad de cada configuración $(x_1, \dots, x_5) \in \{1, \dots, 10\}^5$, donde x_i indica la celda en que se encuentra la partícula i, es $1/10^5$.

   ¿De qué forma podemos obtener las configuraciones deseadas? Primero elegimos (en orden) las 2 bolas que van a ocupar la celdas 6 y 7 (hay 5×4 formas diferentes de hacerlo) y luego elegimos entre las celdas 1, 2, 3, 4, 5 las ubicaciones de las 3 bolas restantes (hay 5^3 formas diferentes de hacerlo). Por lo tanto, su cantidad es $5 \times 4 \times 5^3$ y la probabilidad de observarlas es

   $$p_1 = \frac{5 \times 4 \times 5^3}{10^5} = \frac{1}{5 \times 2^3} = \frac{1}{40} = 0.025.$$

2. *Bose-Einstein*. Las partículas son indistinguibles y todas las configuraciones distintas son equiprobables. La probabilidad de cada configuración $(r_1, \dots, r_{10})$, donde $r_1 + \cdots + r_{10} = 5$ y r_i es la cantidad de partículas en la i-ésima celda, es $1/\binom{14}{9}$.

Las configuraciones deseadas son de la forma $(r_1, \dots, r_5, 1, 1, 0, 0, 0)$, donde $r_1 + \cdots + r_5 = 3$, su cantidad es igual a la cantidad de configuraciones distintas que pueden formarse usando 4 ceros y 3 unos. Por lo tanto, su cantidad es $\binom{7}{3}$ y la probabilidad de observarlas es

$$p_2 = \binom{7}{3}\binom{14}{9}^{-1} = \frac{35}{2002} \approx 0.0174....$$

3. *Fermi-Dirac*. Las partículas son indistinguibles, ninguna celda puede contener más de una partícula y todas las configuraciones distintas son equiprobables. La probabilidad de cada configuración es $1/\binom{10}{5}$.

   Las configuraciones deseadas se obtienen eligiendo tres de las las cinco celdas 1, 2, 3, 4, 5 para ubicar las tres partículas que no están en las celdas 6 y 7. Por lo tanto, su cantidad es $\binom{5}{3}$ y la probabilidad de observarlas es

$$p_2 = \binom{5}{3}\binom{10}{5}^{-1} = \frac{10}{252} \approx 0.0396....$$

□

Ejercicios adicionales

12. Utilizando la estadística de Maxwell-Boltzmann construya un mecanismo aleatorio para estimar el número e.

7.1. De la estadística de Maxwell-Boltzmann a la distribución de Poisson

El espacio se divide en una gran cantidad, n, de pequeñas regiones llamadas celdas y se considera un sistema mecánico compuesto por r partículas que se distribuyen al azar entre las n celdas. Se supone que todas las partículas son distintas y que todas las ubicaciones de las partículas son igualmente posibles.

Hemos visto que cuando $n \to \infty$ y $r \to \infty$ de modo tal que $r/n \to \lambda$, la probabilidad p_0 de que una celda dada no tenga partículas tiende a $e^{-\lambda}$. En general, si p_k designa la probabilidad de que una celda dada contenga exactamente k partículas, tenemos que

$$p_k = \frac{1}{n^r}\binom{r}{k}(n-1)^{r-k} = \binom{r}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{r-k},$$

y usando la fórmula de Stirling: $n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}$, puede verse que

$$p_k \to e^{-\lambda}\frac{\lambda^k}{k!}.$$

Dicho en palabras, la *forma límite* de la estadística de Maxwell-Boltzmann es la *distribución de Poisson de intensidad* λ definida por

$$p(k) := e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \dots$$

$\square$

7.2. De la estadística de Bose-Einstein a la distribución geométrica

En este caso las partículas son indistinguibles y todas las ubicaciones son igualmente posibles.

Hemos visto que cuando $n \to \infty$ y $r \to \infty$ de modo tal que $r/n \to \lambda$, la probabilidad p_0 de que una celda dada no tenga partículas tiende a $\frac{1}{1+\lambda}$. En general, si p_k designa la probabilidad de que una celda dada contenga exactamente k partículas, tenemos que

$$p_k = \binom{r-k+n-2}{n-2}\binom{n+r-1}{n-1}^{-1}$$

y usando la fórmula de Stirling se puede ver que:

$$p_k \to \frac{\lambda^k}{(1+\lambda)^{k+1}}.$$

Dicho en palabras, la *forma límite* de la estadística de Bose-Einstein es la *distribución geométrica de parámetro* $\frac{1}{1+\lambda}$ definida por

$$p(k) := \left(1 - \frac{1}{1+\lambda}\right)^k \frac{1}{1+\lambda}, \qquad k = 0, 1, 2, \dots$$

$\square$

8. Bibliografía consultada

Para redactar estas notas se consultaron los siguientes libros:

1. Bertsekas, D. P., Tsitsiklis, J. N.: Introduction to Probability. M.I.T. Lecture Notes. (2000)

2. Brémaud, P.: An Introduction to Probabilistic Modeling. Springer, New York. (1997)

3. Durrett, R. Elementary Probability for Applications. Cambridge University Press, New York. (2009)

4. Feller, W.: An introduction to Probability Theory and Its Applications. Vol. 1. John Wiley & Sons, New York. (1957)

5. Ferrari, P.: Passeios aleatórios e redes eletricas. Instituto de Matemática Pura e Aplicada. Rio de Janeiro. (1987)

6. Grinstead, C. M. & Snell, J. L. Introduction to Probability. American Mathematical Society. (1997)

7. Kolmogorov, A. N.: Foundations of the Theory of Probability. Chelsea Publishing Co., New York. (1956)

8. Kolmogorov, A. N.: The Theory of Probability. Mathematics. Its Content, Methods, and Meaning. Vol 2. The M.I.T. Press, Massachusetts. (1963) pp. 229-264.

9. Meester, R.: A Natural Introduction to Probability Theory. Birkhauser, Berlin. (2008)

10. Meyer, P. L.: Introductory Probability and Statistical Applications. Addison-Wesley, Massachusetts. (1972)

11. Ross, S. M: Introduction to Probability and Statistics foe Engineers and Scientists. Elsevier Academic Press, San Diego. (2004)

12. Skorokhod, A. V.: Basic Principles and Applications of Probability Theory. Springer-Verlag, Berlin. (2005)

13. Soong, T. T.: Fundamentals of Probability and Statistics for Engineers. John Wiley & Sons Ltd. (2004)

14. Stoyanov, J.: Counterexamples in Probability. John Wiley & Sons. (1997)