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

Sebastian Grynberg

11-13 de marzo 2013

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 general

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

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

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

$$A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots, \tag{1}$$

tal que

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

vale que

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

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 $A^c := \{\omega \in \Omega : \omega \notin A\}$ para designar al evento que no ocurre A. El evento A^c se llama el *complemento* de A.
2. Escribiremos $A \cup B := \{\omega \in \Omega : \omega \in A \text{ o } \omega \in B\}$ para designar al evento que ocurre al menos uno de los eventos A o B. El evento $A \cup B$ se llama la *unión* de A y B.
3. Escribiremos $A \cap B := \{\omega \in \Omega : \omega \in A \text{ y } \omega \in B\}$ para designar al evento ocurren ambos A y 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 $A_1, A_2, \ldots$ 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-IV 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:

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

$\square$

Equiprobabilidad: fórmula de Laplace. Sea Ω un espacio muestral finito. Cuando todos los eventos elementales tienen la misma probabilidad, esto es, cuando para todo $\omega \in \Omega$ vale que $p(\omega) = |\Omega|^{-1}$, se dice que el espacio es *equiprobable*. En ese caso las probabilidades de los eventos $A \subset \Omega$ se calculan usando la *fórmula de Laplace*:

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

En este contexto el problema principal del cálculo de probabilidades consiste determinar la cantidad de eventos elementales favorables a cada evento posible (sin tener que enumerarlo). En otras palabras, la teoría de probabilidades se reduce al *análisis combinatorio*, una importante (y a veces muy díficil) rama de la matemática dedicada a lo que podría llamarse "contar sin contar". En la Sección 3 se desarrollan sus elementos básicos. $\square$

1.2. Relación con los datos experimentales

En el mundo real de los experimentos la teoría de probabilidad se aplica de la siguiente manera:

(1) Consideramos un sistema de condiciones, $\mathcal{S}$, que se pueden repetir 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 1.5 (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). Hay cuatro resultados posibles (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 la misma manera todo evento puede considerarse como un conjunto de eventos elementales. $\square$

(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 siguientes características:

(a) Se puede estar prácticamente seguro de que si el sistema de condiciones $\mathcal{S}$ se repite una gran cantidad 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 I, II, III. 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 $\mathbb{P}(A)$. $\square$

Ejemplo 1.6. 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 $A = \{H\}$

$$0.753;\; 0.757;\; 0.756;\; 0.750;\; 0.746;\; 0.758;\; 0.751;\; 0.748;\; 0.749;\; 0.746,$$

parece razonable asignarle a p el valor 0.75. $\square$

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 tirando un dado equilibrado, entonces en n tiradas del dado la probabilidad de obtener un seis en todas ellas es $(1/6)^n > 0$; en otras palabras, con probabilidad $(1/6)^n$ tendremos una frecuencia relativa igual a *uno* de sacar un seis en todas las tiradas ; y con probabilidad $(5/6)^n$ no saldrá ningún seis, es decir, la frecuencia relativa de sacar seis será igual a *cero*. $\square$

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

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

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

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

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

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

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

1.4. Sobre el axioma de continuidad

Nota Bene 1. Si la familia de eventos $\mathcal{A}$ es finita el axioma de continuidad IV se deduce de los axiomas I-III. En tal caso, en la sucesión (1) solo hay una cantidad finita de eventos diferentes. Si A_k es el menor de ellos, entonces todos los conjuntos A_{k+m}, $m \geq 1$ 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. $\square$

Nota Bene 2. Se puede probar que para espacios muestrales infinitos, el axioma de continuidad IV es independiente de los axiomas I-III. Este 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. $\square$

Ejemplo 1.7. 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 $\lim_{n\to\infty} \mathbb{P}(A_n) = 1$, debido a que $\mathbb{P}(A_n) = 1$ para todo $n \geq 1$. $\square$

Teorema 1.8.

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

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

$\square$

Ejemplo 1.9 (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 $b-a$, para cualquier pareja de números reales a y b tales que $0 < a < b \leq 1$. $\square$

Ejemplo 1.10 (Ternario de Cantor)**.** Se elije al azar un número U del intervalo $(0, 1]$, ¿cuál es la probabilidad de que el 1 no aparezca en el desarrollo en base 3 de U?

Consideramos la representación en base 3 del número U:

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

donde $a_k(U) \in \{0, 1, 2\}$, $k \geq 1$.

Lo que queremos calcular es la probabilidad del evento $A = \{a_k(U) \neq 1, \forall\, k \geq 1\}$. Primero observamos 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 $A_1 \supset A_2 \supset \cdots$. Usando el inciso (a) del **Teorema 1.8** 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 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, $\mathbb{P}(A) = \lim_{n\to\infty} (2/3)^n = 0$. □

Teorema 1.11 (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

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) \tag{2}$$

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

$$\lim_{n\to\infty} \mathbb{P}(R_n) = 0 \tag{3}$$

y por el teorema de aditividad tenemos que

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{k=1}^{n} \mathbb{P}(A_k) + \mathbb{P}(R_n). \tag{4}$$

De (4) y (3) se obtiene (2). □

Corolario 1.12 (Teorema de cubrimiento)**.** Si $B, A_1, A_2, \ldots$ es una sucesión de eventos tal que $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ y $B \subset A$, entonces

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

Demostración. Una cuenta. Descomponemos B en una unión disjunta de eventos

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

y aplicamos el teorema de σ-aditividad

$$\mathbb{P}(B) \quad = \quad \sum_{n=1}^{\infty} \mathbb{P} \left(B \cap \left(A_n \setminus \bigcup_{k=1}^{n-1} (A_n \cap A_k) \right) \right) \le \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

□

Ejercicios adicionales

4. Sean Ω un conjunto no vacío y $\mathcal{A}$ un álgebra de eventos. Sea $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ una función tal que

I. Para cada $A \in \mathcal{A}$, $\mathbb{P}(A) \ge 0$,
II. $\mathbb{P}(\Omega) = 1$.
III. Si los eventos A y B no tienen elementos en común, entonces $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
IV'. Si $(A_n)_{n \ge 1}$ es una sucesión de eventos disjuntos dos a dos y $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, entonces

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

Mostrar que bajo esas condiciones la función $\mathbb{P}$ satisface el axioma de continuidad.

1.5. σ-álgebras y teorema de extensión

El álgebra $\mathcal{A}$ se llama una σ*-álgebra*, si toda unión numerable $\bigcup_{n=1}^{\infty} A_n$ 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 1.13 (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 1.14 (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 $\sigma(\mathcal{A})$. $\square$

2. Simulación de experimentos aleatorios con espacio muestral finito

2.1. Números aleatorios.

Toda computadora tiene instalado un algoritmo para simular números aleatorios que se pueden obtener mediante una instrucción del tipo "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))

Cómo usar los números aleatorios. La idea principal se puede presentar 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

$$X := \mathbf{1}\{1 - p < U \leq 1\}. \tag{5}$$

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 se puede implementar casi inmediatamente 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 equilibrada

```
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.

2.2. 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, para simular los resultados del experimento aleatorio considerado. Definimos

$$X = \sum_{k=1}^{m} k \mathbf{1}\left\{L_{k-1} < U \leq L_k\right\}, \tag{6}$$

donde

$$L_0 := 0 \quad \text{y} \quad L_k := \sum_{i=1}^{k} p_i, \quad (1 \leq k \leq m)$$

e identificamos cada evento elemental $\omega_k \in \Omega$ con su correspondiente subíndice k.

En efecto, de la definición (6) se deduce que para cada $k = 1, \dots, m$ vale que

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

$\square$

Nota Bene. El mecanismo aleatorio definido en (6) se puede construir "gráficamente" de la siguiente manera:

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

2. Sortear un número aleatorio, U, y observar en qué intervalo de la partición cae.

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

Ejemplo 2.1 (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 (6), se construye de la siguiente manera:

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,

$$X = \sum_{k=1}^{6} k\mathbf{1}\left\{\frac{k-1}{6} < U \leq \frac{k}{6}\right\}. \tag{7}$$

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
```

2.3. 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 2.2. 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 2.1 para simular el resultado del lanzamiento de una moneda equilibrada se pueden simular $n = 10000$ realizaciones del experimento que consiste en lanzar 5 monedas equilibradas. Veamos como hacerlo. Usamos la construcción (5) 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

(8)

En tal caso diremos que se obtuvieron las siguientes estimaciones

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

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 (8)

```
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 análisis combinatorio (ver sección 3) 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,

$$\begin{array}{lll} \mathbb{P}(N = 0) = 0.03125, & \mathbb{P}(N = 1) = 0.15625, & \mathbb{P}(N = 2) = 0.31250, \\ \mathbb{P}(N = 3) = 0.31250, & \mathbb{P}(N = 4) = 0.15625, & \mathbb{P}(N = 5) = 0.03125. \end{array}$$

□

Ejemplo 2.3 (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.

1. La construcción (7) 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.

2. La construcción (7) permite simular 24 tiros de dos dados usando 48 números aleatorios independientes $U_1, U_2, \dots, U_{47}, U_{48}$.

La cantidad de veces que se obtiene un doble as en los 24 tiros de dos dados 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 dados.

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 doble as en 24 tiros de dos dados 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.

3. 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.

3.1. 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 $|A \times B| = |A| \cdot |B|$.

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

	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.

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. $\square$

Teorema 3.1 (Regla del producto)**.** Sean $A_1, A_2, \dots, A_n$, n conjuntos cualesquiera. El producto cartesiano de los n conjuntos $A_1, A_2, \dots, A_n$ se define por

$$A_1 \times A_2 \times \cdots \times A_n = \{(x_1, x_2, \dots, x_n) : \ x_i \in A_i, \ 1 \leq i \leq n\}.$$

Si los conjuntos $A_1, A_2, \dots, A_n$ son finitos, entonces

$$|A_1 \times A_2 \times \cdots \times A_n| = \prod_{i=1}^{n} |A_i|.$$

Demostración. Si $n = 2$ ya lo demostramos. Si $n = 3$, tomamos los pares (x_1, x_2) como elementos de un nuevo tipo. Hay $|A_1| \cdot |A_2|$ elementos de ese tipo y $|A_3|$ elementos x_3. Cada terna (x_1, x_2, x_3) es un par formado por un elemento (x_1, x_2) y un elemento x_3; por lo tanto, la cantidad de ternas es $|A_1| \cdot |A_2| \cdot |A_3|$. Etcétera. $\square$

Nota Bene. Muchas aplicaciones se basan en la siguiente reformulación de la regla del producto: *r decisiones sucesivas con exactamente n_k elecciones posibles en el k-ésimo paso pueden producir un total de $n_1 \cdot n_2 \cdots n_r$ resultados diferentes.* $\square$

Ejemplo 3.2 (Ubicar r bolas en n urnas)**.** Los resultados posibles del experimento se pueden representar 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 se ubicó en la urna j"*. Cada bola puede ubicarse en una de las n urnas posibles. Con r bolas tenemos r elecciones sucesivas con exactamente n elecciones posibles en cada paso. En consecuencia, *r bolas pueden ubicarse en n urnas de n^r formas distintas.*

Usamos el lenguaje figurado de bolas y urnas, pero el mismo espacio muestral admite muchas interpretaciones distintas. Para ilustrar el asunto *listaremos una cantidad de situaciones en las cuales aunque el contenido intuitivo varía son todas abstractamente equivalentes al esquema de ubicar r bolas en n urnas, en el sentido de 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 $n = 365$ urnas (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 $n = 7$ urnas.

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

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

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

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

□

3.2. 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 tamaño 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 se puede elegir más de una vez. Cada uno de los k elementos se puede elegir 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 elecciones posibles 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

$$(n)_k := n(n-1)(n-2)\cdots(n-k+1) \tag{9}$$

elecciones posibles.

Teorema 3.3. *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 3.4. Consideramos una urna con 8 bolas numeradas $1, 2, \dots, 8$

(a) **Extracción con reposición.** Extraemos 3 bolas *con reposición*: 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 3.3, Ω_1 tiene $8^3 = 512$ elementos. Bajo la hipótesis de que todos los elementos tienen la misma probabilidad, la probabilidad de observar la secuencia $(3, 7, 1)$ es $1/512$.

(b) **Extracción de una colección ordenada sin reposición.** Extraemos 3 bolas *sin reposición*: cada bola elegida *no* se vuelve a poner en la urna. Anotamos los números de las bolas en el orden en que fueron extraídas de la urna. 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 3.3, Ω_2 tiene $(8)_3 = 8 \cdot 7 \cdot 6 = 336$ 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 3.5. 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 las dos sean rojas?* Sea R el evento "las dos son 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 las dos sean del mismo color?* Sea N el evento "las dos son 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 de las dos sea roja?* El evento "al menos una de las dos es roja" es el complemento de "las dos son 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 $a_1, a_2, \ldots, a_n$ se pueden ordenar de $(n)_n = n \cdot (n-1) \cdots 2 \cdot 1$ formas distintas. Usualmente el número $(n)_n$ se denota $n!$ y se llama el *factorial de* n.

Corolario 3.6. *La cantidad de formas distintas en que se pueden ordenar n elementos es*

$$n! = 1 \cdot 2 \cdots n. \tag{10}$$

Observación 3.7. 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 $(n)_k$ se puede calcular del siguiente modo

$$(n)_k = \frac{n!}{(n-k)!} \tag{11}$$

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 $a_1, \ldots, a_n$. 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*

$$p = \frac{(n)_k}{n^k} = \frac{n(n-1)\cdots(n-k+1)}{n^k} \tag{12}$$

Las interpretaciones concretas de la fórmula (12) revelan aspectos sorprendentes.

Muestras aleatorias de números. La población consiste de los diez dígitos $0, 1, \ldots, 9$. 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 urnas. *Si n bolas se ubican aleatoriamente en n urnas, la probabilidad de que cada urna esté 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 ubicaciones 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 lo extremadamente improbable 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 (12) *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

5. 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)*: $n! \sim e^{-n} n^{n+\frac{1}{2}} \sqrt{2\pi}$.

6. 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.

3.3. 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 se pueden ordenar de $k!$ 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,

$$C_{n,k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}. \tag{13}$$

Los números definidos en (13) se llaman *coeficientes binomiales* o *números combinatorios* y la notación clásica para ellos es $\binom{n}{k}$.

Teorema 3.8. *Una población de n elementos tiene*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{14}$$

diferentes subpoblaciones de tamaño $k \leq n$.

Ejemplo 3.9. 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 del conjunto $\{1, 2, \ldots, 8\}$. Por el Teorema 3.8 Ω_3 tiene $\binom{8}{3} = 56$ elementos. Bajo la hipótesis de 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 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

$$\binom{n}{k} = \binom{n}{n-k}. \tag{15}$$

La ecuación (15) se deduce inmediatamente de la identidad (14). El lado izquierdo de la ecuación (15) no está definido para $k = 0$, pero el lado derecho si lo está. Para que la ecuación (15) 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

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}, \tag{16}$$

junto con el conocimiento de los datos de borde

$$\binom{n}{0} = \binom{n}{n} = 1, \tag{17}$$

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 (17) indican que el primero y el último de esos números son 1. Los números restantes se determinan por la ecuación en diferencias (16). 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 piezas 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. □

Ejercicios adicionales

7. 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.

3.4. Particiones

Teorema 3.10. *Sean $r_1, \ldots, r_k$ enteros tales que*

$$r_1 + r_2 + \cdots + r_k = n, \quad r_i \geq 0. \tag{18}$$

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

$$\frac{n!}{r_1!r_2!\cdots r_k!}. \tag{19}$$

Los números (19) se llaman coeficientes multinomiales.

Demostración. Un uso repetido de (14) muestra que el número (19) se puede reescribir en la forma

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\binom{n-r_1-r_2}{r_3}\cdots\binom{n-r_1-\cdots-r_{k-2}}{r_{k-1}} \tag{20}$$

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 (20) representa el número de formas en que se puede realizar la partición. □

Ejemplo 3.11 (Accidentes)**.** En una semana ocurrieron 7 accidentes. Cuál es la probabilidad de que en dos días de esa semana hayan ocurrido dos accidentes cada día y de que en otros tres días hayan ocurrido un accidente cada día?

Primero particionamos los 7 días en 3 subpoblaciones: dos días con dos accidentes en cada uno, tres días con un accidente en cada uno y dos días sin accidentes.. Esa partición en tres grupos de tamaños $2, 3, 2$ se puede hacer de $7!/(2!3!2!)$ formas distintas y por cada una de ellas hay $7!/(2!2!1!1!1!0!0!) = 7!/(2!2!)$ formas diferentes de ubicar los 7 accidentes en los 7 días. Por lo tanto, el valor de la probabilidad requerido es igual a

$$\frac{7!}{2!3!2!}\times\frac{7!}{2!2!}\frac{1}{7^7}=0.3212...$$

□

Ejercicios adicionales

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

9. Se ubicarán 6 bolas distinguibles en 8 urnas numeradas $1, 2, \dots, 8$. Suponiendo que todas las configuraciones distintas son equiprobables calcular la probabilidad de que resulten tres urnas ocupadas con una bola cada una y que otra urna contenga las tres bolas restantes.

3.5. Distribución Hipergeométrica

Muchos problemas combinatorios se pueden reducir 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, $0 \leq k \leq \text{mín}(n_1, r)$.

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 negras 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

$$p_k = \binom{n_1}{k}\binom{n_2}{r-k}\binom{n_1+n_2}{r}^{-1} \tag{21}$$

El sistema de probabilidades obtenido se llama la *distribución hipergeométrica.*

3.5.1. 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 (21) permite hacer inferencias sobre la cantidad desconocida n_1; se trata de problema típico de estimación estadística que será analizado más adelante. □

Ejemplo 3.12. 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 defectuosas 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

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

□

Ejemplo 3.13. 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?.

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.

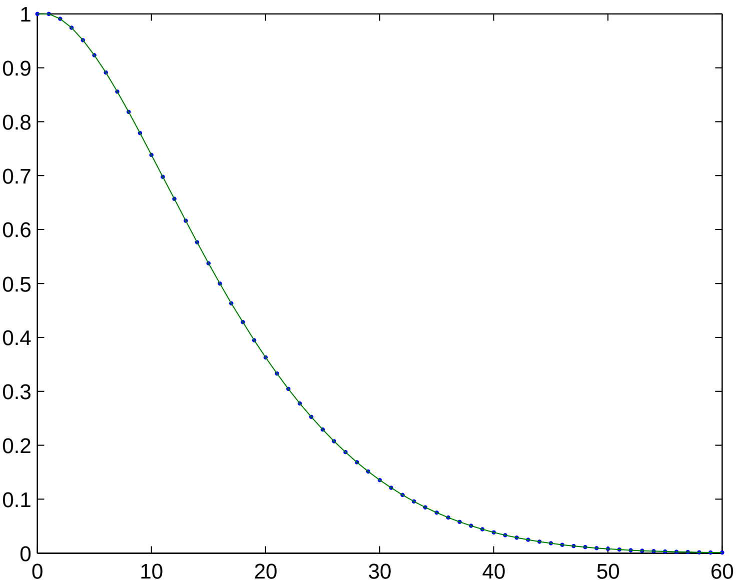

Figura 1: Gráfico de función $p(k)$.

¿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 $p(20) \approx 0.3630$. $\square$

Ejemplo 3.14. 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

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

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".

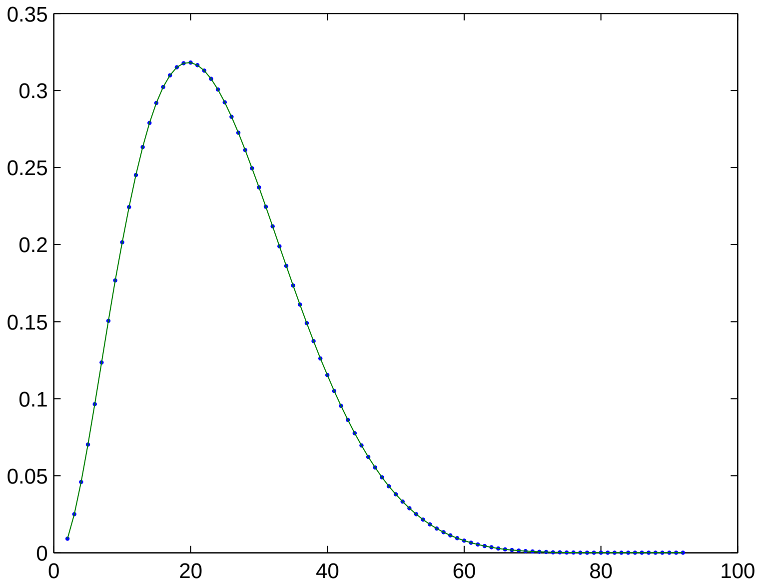

Figura 2: Gráfico de función $p(n_1)$. Observar que $\arg\max\{p(n_1) : 2 \leq n_1 \leq 92\} = 20$. $\square$

3.5.2. 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 (21) permite hacer inferencias sobre la cantidad desconocida n.

Ejemplo 3.15 (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 (21):

$$p_k(n) = \binom{n_1}{k}\binom{n-n_1}{r-k}\binom{n}{r}^{-1}. \tag{22}$$

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 (22) 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

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

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 (22) $n = 8500, n_1 = r = 1000$, 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

10. 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?

---

4. 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, $1 \leq i \leq n$. 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 unos 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 ceros en 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 $1/\binom{n}{r}$.

Ejemplo 4.1. Se distribuyen 5 partículas en 10 celdas numeradas $1, 2, \ldots, 10$. 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 $(x_1, \ldots, x_5) \in \{1, \ldots, 10\}^5$, 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

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

2. *Bose-Einstein*. Las partículas son indistinguibles y todas las configuraciones distintas son equiprobables. La probabilidad de cada configuración $(r_1, \ldots, r_{10})$, 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 $(r_1, \ldots, r_5, 1, 1, 0, 0, 0)$, donde $r_1 + \cdots + r_5 = 3$, su cantidad es igual a la cantidad de configuraciones distintas que pueden formarse usando 3 ceros y 4 unos. Por lo tanto, su cantidad es $\binom{7}{3}$ y la probabilidad de observarlas es

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

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

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

□

Ejemplo 4.2. Calcular para cada una de las tres estadísticas mencionadas, la probabilidad de que una celda determinada (p.ej., la número 1) no contenga 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_{MB}, p_{BE}, p_{FD} las probabilidades del evento especificado para cada una de las estadísticas (siguiendo el orden de exposición), tenemos que

$$\begin{aligned}
p_{MB} &= \frac{(n-1)^r}{n^r} = \left(1-\frac{1}{n}\right)^r, \\
p_{BE} &= \binom{r+n-2}{n-2}\binom{r+n-1}{n-1}^{-1} = \frac{n-1}{N+n-1}, \\
p_{FD} &= \binom{n-1}{r}\binom{n}{r}^{-1} = 1-\frac{r}{n}.
\end{aligned}$$

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

$$p_{MB} = e^{-\lambda}, \qquad p_{BE} = \frac{1}{1+\lambda}, \qquad p_{FD} = 1-\lambda.$$

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

Ejercicios adicionales

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

4.1. Algunas distribuciones relacionadas con la estadística de Maxwell-Boltzmann

Se distribuyen r partículas en n celdas y cada una de las n^r configuraciones tiene probabilidad n^{-r}.

4.1.1. Cantidad de partículas por celda: la distribución binomial

Cantidad de partículas en una celda específica. Para calcular la probabilidad, $p_{MB}(k)$, de que una celda específica contenga exactamente k partículas ($k = 0, 1, \ldots, r$) notamos que las k partículas pueden elegirse de $\binom{r}{k}$ formas, y las restantes $r-k$ partículas pueden ubicarse en las restantes $n-1$ celdas de $(n-1)^{r-k}$ formas. Resulta que

$$p_{MB}(k) = \binom{r}{k}(n-1)^{r-k}\frac{1}{n^r}$$

Dicho en palabras, en la estadística de Maxwell-Boltzmann la probabilidad de que una celda dada contenga exactamente k partículas está dada por la distribución Binomial $(r, \frac{1}{n})$ definida por

$$p(k) := \binom{r}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{r-k}, \qquad 0 \leq k \leq r. \tag{23}$$

$\square$

Cantidad de partículas más probable en una celda específica. La cantidad más probable de partículas en una celda específica es el entero ν tal que

$$\frac{(r-n+1)}{n} < \nu \leq \frac{(r+1)}{n}. \tag{24}$$

Para ser más precisos:

$$p_{MB}(0) < p_{MB}(1) < \cdots < p_{MB}(\nu - 1) \leq p_{MB}(\nu) > p_{MB}(\nu + 1) > \cdots > p_{MB}(r).$$

Demostración. (Ejercicio.) $\square$

4.1.2. Forma límite: la distribución de Poisson

Forma límite. Si $n \to \infty$ y $r \to \infty$ de modo que la cantidad promedio $\lambda = r/n$ de partículas por celda se mantiene constante, entonces

$$p_{MB}(k) \to e^{-\lambda} \frac{\lambda^k}{k!}.$$

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

$$p(k) := e^{-\lambda} \frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots \tag{25}$$

Demostración. Primero observamos que:

$$\begin{aligned}
\binom{r}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{r-k} &= \frac{r!}{k!(r-k)!} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{r-k} \\
&= \frac{1}{k!} \left(\frac{1}{n}\right)^k \left(\frac{n-1}{n}\right)^{-k} \frac{r!}{(r-k)!} \left(1 - \frac{1}{n}\right)^r \\
&= \frac{1}{k!} \frac{1}{(n-1)^k} \frac{r!}{(r-k)!} \left(1 - \frac{1}{n}\right)^r.
\end{aligned} \tag{26}$$

Reemplazando en (26) $r = \lambda n$ obtenemos:

$$\begin{aligned}
\binom{\lambda n}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{\lambda n - k} &= \frac{1}{k!} \frac{1}{(n-1)^k} \frac{(\lambda n)!}{(\lambda n - k)!} \left(1 - \frac{1}{n}\right)^{\lambda n} \\
&= \left[\left(1 - \frac{1}{n}\right)^n\right]^\lambda \frac{1}{k!} \frac{1}{(n-1)^k} \frac{(\lambda n)!}{(\lambda n - k)!} \\
&\sim e^{-\lambda} \frac{1}{k!} \left(\frac{1}{(n-1)^k} \frac{(\lambda n)!}{(\lambda n - k)!}\right).
\end{aligned} \tag{27}$$

Para estimar el último factor del lado derecho de (27) utilizamos la fórmula de Stirling $n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}$:

$$\begin{aligned}
\frac{1}{(n-1)^k}\frac{(\lambda n)!}{(\lambda n-k)!} &\sim \frac{1}{(n-1)^k}\frac{\sqrt{2\pi}\,(\lambda n)^{\lambda n+\frac{1}{2}}e^{-\lambda n}}{\sqrt{2\pi}\,(\lambda n-k)^{(\lambda n-k)+\frac{1}{2}}e^{-(\lambda n-k)}} \\
&= \frac{1}{(n-1)^k}\frac{(\lambda n)^{\lambda n+\frac{1}{2}}e^{-k}}{(\lambda n-k)^{(\lambda n-k)+\frac{1}{2}}} \\
&= \left(\frac{\lambda n-k}{n-1}\right)^k \left(\frac{\lambda n}{\lambda n-k}\right)^{\lambda n+\frac{1}{2}} e^{-k} \\
&\sim \lambda^k e^{-k}\left[\left(1-\frac{k}{\lambda n}\right)^{\lambda n+\frac{1}{2}}\right]^{-1} \\
&\sim \lambda^k. \qquad (28)
\end{aligned}$$

De (26), (27) y (28) resulta que

$$\binom{r}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{r-k} \sim e^{-\lambda}\frac{\lambda^k}{k!}. \qquad (29)$$

□

4.2. Algunas distribuciones relacionadas con la estadística de Bose-Einstein

Se distribuyen r partículas indistinguibles en n celdas y cada una de las $\binom{r+n-1}{n-1}$ configuraciones tiene probabilidad $1/\binom{r+n-1}{n-1}$.

4.2.1. Cantidad de partículas por celda

Cantidad de partículas en una celda específica. Para calcular la probabilidad, $p_{BE}(k)$, de que una celda específica contenga exactamente k partículas ($k = 0, 1, \dots, r$) fijamos k de los r ceros y 1 de los $n-1$ unos para representar que hay k partículas en la urna específica. La cantidad de configuraciones distintas que pueden formarse con los restantes $r-k$ ceros y $n-2$ unos es $\binom{r-k+n-2}{n-2}$. Resulta que

$$p_{BE}(k) = \binom{r-k+n-2}{n-2}\binom{r+n-1}{n-1}^{-1}. \qquad (30)$$

Cantidad de partículas más probable en una celda específica. Cuando $n > 2$ la cantidad más probable de partículas en una celda específica es 0 o más precisamente $p_{BE}(0) > p_{BE}(1) > \cdots$.

Demostración. (Ejercicio.) □

4.2.2. Forma límite: la distribución de Geométrica

Forma límite. Si $n \to \infty$ y $r \to \infty$ de modo que la cantidad promedio $\lambda = r/n$ de partículas por celda se mantiene constante, entonces

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

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

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

Demostración. Primero observamos que:

$$\begin{aligned} \binom{r-k+n-2}{n-2}\binom{r+n-1}{n-1}^{-1} &= \frac{(r-k+n-2)!}{(n-2)!(r-k)!}\frac{(n-1)!r!}{(r+n-1)!} \\ &= \frac{(n-1)!}{(n-2)!}\frac{r!}{(r-k)!}\frac{(r-k+n-2)!}{(r+n-1)!}. \qquad (31) \end{aligned}$$

Reemplazando en el lado derecho de (31) $r = \lambda n$ obtenemos:

$$\frac{(n-1)!}{(n-2)!}\frac{(\lambda n)!}{(\lambda n-k)!}\frac{(\lambda n-k+n-2)!}{(\lambda n+n-1)!} \qquad (32)$$

Para estimar los factores que intervienen en (32) utilizamos la fórmula de Stirling $n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}$:

$$\begin{aligned} \frac{(n-1)^{n-1+\frac{1}{2}}e^{-n+1}}{(n-2)^{n-2+\frac{1}{2}}e^{-n+2}} &\sim (n-2)e^{-1}\left[\left(1-\frac{1}{n-1}\right)^{n-1}\right]^{-1} \\ &\sim n-2 \sim n, \qquad (33) \\ \frac{(\lambda n)^{\lambda n+\frac{1}{2}}e^{-\lambda n}}{(\lambda n-k)^{\lambda n-k+\frac{1}{2}}e^{-\lambda n+k}} &\sim (\lambda n-k)^k e^{-k}\left[\left(1-\frac{k}{\lambda n}\right)^{\lambda n}\right]^{-1} \\ &\sim (\lambda n-k)^k \sim \lambda^k n^k, \qquad (34) \\ \frac{(\lambda n-k+n-2)^{\lambda n-k+n-2+\frac{1}{2}}e^{-\lambda n+k-n+2}}{(\lambda n+n-1)^{\lambda n+n-1+\frac{1}{2}}e^{-\lambda n-n+1}} &\sim (\lambda n-k+n-2)^{-k-1}e^{k+1} \\ &\quad \times \left(1-\frac{k+1}{\lambda n+n-1}\right)^{\lambda n+n-1} \\ &\sim (\lambda n-k+n-2)^{-k-1} \\ &\sim \frac{1}{(1+\lambda)^{k+1}n^{k+1}}. \qquad (35) \end{aligned}$$

De (31), (32), (33), (34) y (35) resulta que

$$\binom{r-k+n-2}{n-2}\binom{r+n-1}{n-1}^{-1} \sim \frac{\lambda^k}{(1+\lambda)^k}. \qquad (36)$$

□

Ejercicios adicionales

12. Considerando la estadística de Maxwell-Boltzmann para la distribución aleatoria de r partículas en n celdas demostrar que la cantidad de de partículas más probable en una celda determinada es la parte entera de $\frac{r+1}{n}$.

13. Considerando la estadística de Bose-Einstein para la distribución aleatoria de r partículas (indistinguibles) en $n > 2$ celdas demostrar que la cantidad de de partículas más probable en una celda determinada es 0.

4.3. Tiempos de espera

Consideramos una vez más el experimento conceptual de ubicar aleatoriamente partículas (distinguibles) en n celdas. Solo que ahora no fijamos la cantidad r de partículas y ubicamos las partículas una por una hasta que ocurra alguna situación prescrita. Analizaremos dos situaciones:

(i) Ubicar partículas hasta que alguna se ubique en una celda ocupada previamente.

(ii) Fijada una celda, ubicar partículas hasta que alguna ocupe la celda.

Situación (i). Usamos símbolos de la forma $(j_1, j_2, \dots, j_r)$ para indicar que la primera, la segunda,... y la r-ésima partícula están ubicadas en las celdas $j_1, j_2, \dots, j_r$ y que el proceso culmina en el paso r. Esto significa que las j_i son enteros entre 1 y n; que las $j_1, j_2, \dots, j_{r-1}$ son todas diferentes y que j_r es igual a una de ellas. Toda configuración de ese tipo representa un punto muestral. Los posibles valores de r son $2, 3, \dots, n + 1$.

Para un r fijo el conjunto de todos los puntos muestrales $(j_1, j_2, \dots, j_r)$ representa el evento que el proceso termina en el r-ésimo paso. Los números $j_1, j_2, \dots, j_{r-1}$ pueden elegirse de $(n)_{r-1}$ formas diferentes; j_r podemos elegir uno de los $r - 1$ números $j_1, j_2, \dots, j_{r-1}$. Por lo tanto la probabilidad de que el proceso termine en el r-ésimo paso es

$$p_r = \frac{(n)_{r-1}(r-1)}{n^r}. \tag{37}$$

□

Situación (ii). Usamos símbolos de la forma $(j_1, j_2, \dots, j_r)$ para indicar que la primera, la segunda,... y la r-ésima partícula están ubicadas en las celdas $j_1, j_2, \dots, j_r$ y que el proceso culmina en el paso r. Las r-uplas $(j_1, j_2, \dots, j_r)$ están sujetas a la condición de que los números $j_1, j_2, \dots, j_{r-1}$ son diferentes de un número prescrito $a \leq n$, y $j_r = a$.

Para un r fijo el conjunto de todos los puntos muestrales $(j_1, j_2, \dots, j_r)$ representa el evento que el proceso termina en el r-ésimo paso. Los números $j_1, j_2, \dots, j_{r-1}$ pueden elegirse de $(n-1)^{r-1}$ formas diferentes; j_r debe ser a. Por lo tanto la probabilidad de que el proceso termine en el r-ésimo paso es

$$p_r = \frac{(n-1)^{r-1}}{n^r}. \tag{38}$$

□

5. 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)