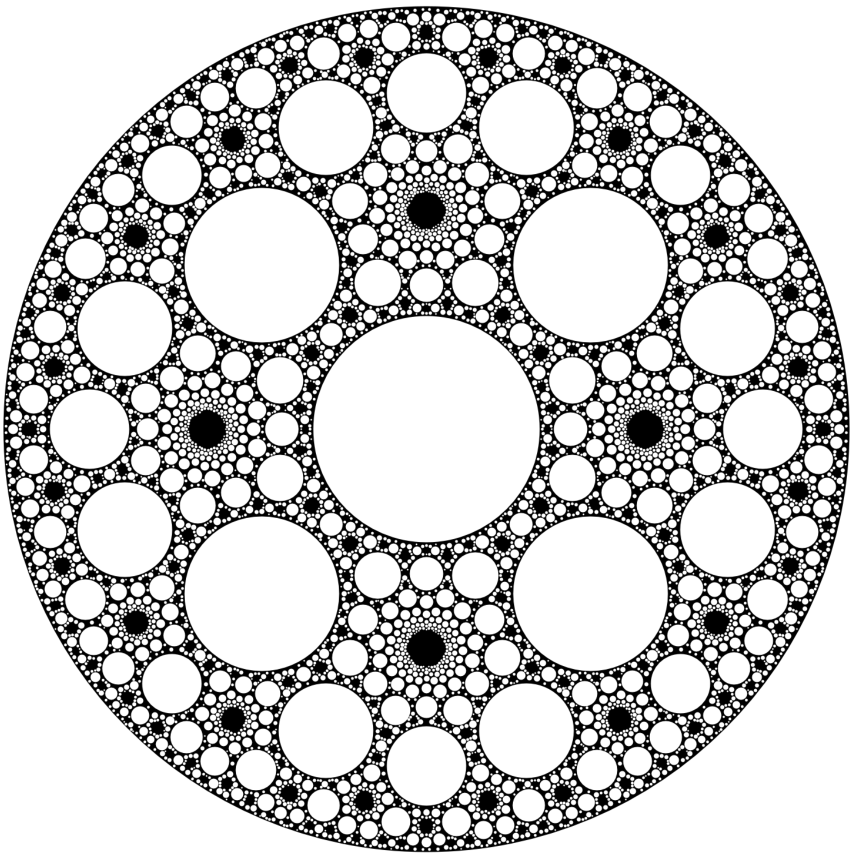
Probabilidad Condicional, Independencia Estocástica
Algunos modelos probabilísticos
(Borradores, Curso 23)

Sebastian Grynberg

18-20 de marzo 2013



"No importa lo que yo piense.
Es lo que tú piensas lo que es relevante."

(Dr. House)

Índice

1. Probabilidad Condicional

1.1. Probabilidad Condicional

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

Definición 1.1 (Probabilidad condicional)**.** Sea $A \subset \Omega$ un evento de probabilidad positiva. Para cada evento B definimos

$$\mathbb{P}(B|A) := \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}. \tag{1}$$

La cantidad definida en (1) se llama la *probabilidad condicional de B dado que ocurrió A.*

Nota Bene (La probabilidad condicional induce una medida de probabilidad sobre los eventos aleatorios). 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
$$\mathbb{P}(B \cup C|A) = \mathbb{P}(B|A) + \mathbb{P}(C|A).$$
4. Para cada sucesión decreciente de eventos $B_1 \supset B_2 \supset \cdots$ tal que $\bigcap_{n=1}^{\infty} B_n = \emptyset$ vale que $\lim_{n\to\infty} \mathbb{P}(B_n|A) = 0$.

Comparando las propiedades 1-4 con los axiomas I-IV, se concluye 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 1.2. Se lanza un dado equilibrado. Sabiendo que el resultado del dado no superó al 4, cuál es la probabilidad condicional de haber obtenido un 3? Denotando mediante A al evento "*el resultado no supera al* 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

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A). \tag{2}$$

denominada *regla del producto.*

El siguiente Teorema generaliza la regla del producto (2) y se obtiene por inducción.

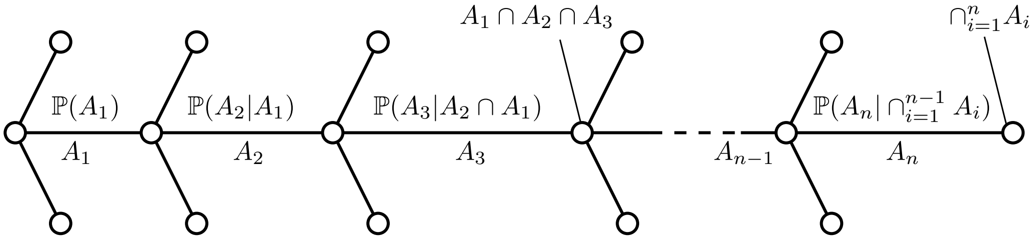


Figura 1: **Ilustración de la regla del producto.** El evento $\cap_{i=1}^n A_i$ tiene asociada una única trayectoria sobre un árbol que describe la historia de un experimento aleatorio realizado por etapas sucesivas. Las aristas de esta trayectoria corresponden a la ocurrencia sucesiva de los eventos $A_1, A_2, \ldots, A_n$ 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 $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)$.

Teorema 1.3 (Regla del producto)**.** Suponiendo que todos los eventos condicionantes tienen probabilidad positiva, tenemos que

$$\mathbb{P}\left(\cap_{i=1}^n A_i\right) = \mathbb{P}\left(A_n \left|\cap_{i=1}^{n-1} A_i\right.\right) \cdots \mathbb{P}\left(A_3 | A_1 \cap A_2\right) \mathbb{P}(A_2|A_1)\mathbb{P}(A_1). \tag{3}$$

Ejemplo 1.4. 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

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

□

1.2. Fórmula de probabilidad total

Teorema 1.5 (Fórmula de probabilidad total)**.** Sea $A_1, A_2, \ldots$ una sucesión de eventos disjuntos dos a dos tal que $\bigcup_{n\geq 1} A_n = \Omega$. Para cada $B \in \mathcal{A}$ vale la siguiente fórmula

$$\mathbb{P}(B) = \sum_{n\geq 1} \mathbb{P}(B|A_n)\mathbb{P}(A_n), \tag{4}$$

denominada *fórmula de probabilidad total*[1].

[1]Rigurosamente, $\mathbb{P}(B|A_n)$ está definida cuando $\mathbb{P}(A_n) > 0$, por lo cual en la fórmula (4) interpretaremos que $\mathbb{P}(B|A_n)\mathbb{P}(A_n) = 0$ cuando $\mathbb{P}(A_n) = 0$.

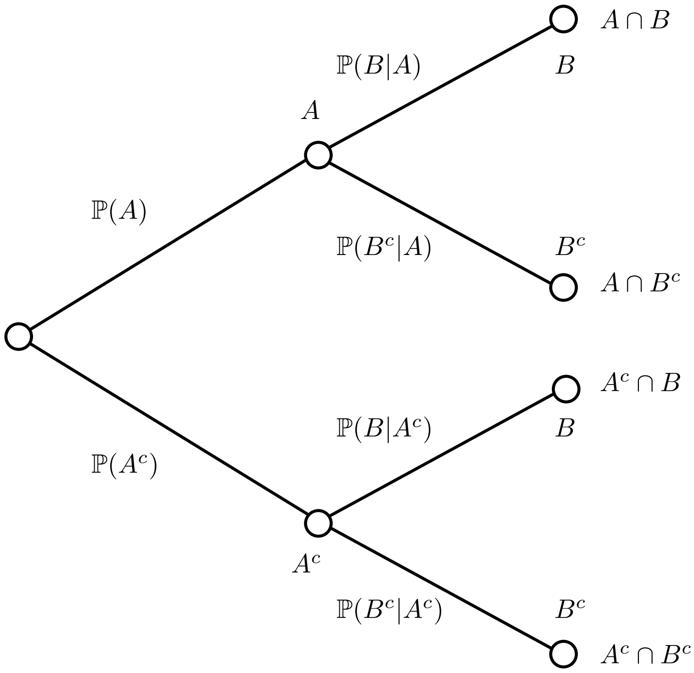


Figura 2: **Ilustración de la fórmula de probabilidad total.** *Un experimento de dos etapas binarias y su correspondiente diagrama de árbol.* La primera ramificación (de izquierda a derecha) se basa en el resultado de la primer etapa del experimento (A o A^c) y la segunda en su resultado final (B o 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: $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)$.

Demostración de la fórmula de probabilidad total. De la identidad de conjuntos

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

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 $\mathbb{P}(B \cap A_n) = \mathbb{P}(B|A_n)\mathbb{P}(A_n)$. $\square$

Nota Bene: Cálculo mediante condicionales. Si se dispone de una colección de eventos $A_1, A_2, \dots$ de los cuales uno y solamente uno debe ocurrir, la fórmula de probabilidad total (4) permite calcular la probabilidad de cualquier evento B condicionando a saber cuál de los eventos A_i ocurrió. Más precisamente, la fórmula (4) 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 1.6 (Experimentos de dos etapas)**.** La primera etapa del experimento produce una partición $A_1, A_2, \ldots$ del espacio muestral Ω. La segunda etapa produce el evento B. La fórmula (4) se utiliza para calcular la probabilidad de B.

Ejemplo 1.7. Una urna contiene 5 bolas rojas y 10 bolas negras. Se extraen dos bolas sin reposición. ¿Cuál es la probabilidad de 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.

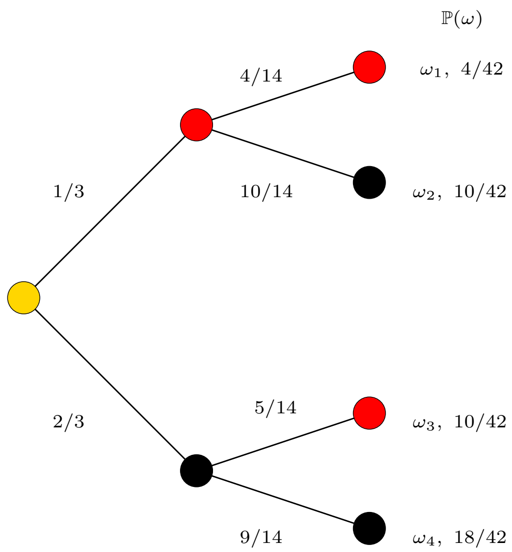


Figura 3: Observando el árbol se deduce que la probabilidad de que la segunda bola sea negra es: $\frac{1}{3} \cdot \frac{10}{14} + \frac{2}{3} \cdot \frac{9}{14} = \frac{2}{3}$.

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

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

$\square$

1.3. Regla de Bayes

Primera versión de la regla de Bayes. Sean A y B dos eventos de probabilidad positiva. De la regla del producto (2) y su análoga $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$ se obtiene la siguiente fórmula importante

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}, \tag{5}$$

que contiene lo esencial del Teorema de Bayes.

Ejemplo 1.8. 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. □

Teorema 1.9 (Bayes)**.** Sean $A_1, A_2, \dots,$ eventos disjuntos dos a dos y tales que $\bigcup_{n \geq 1} A_n = \Omega$. Sea B un evento de probabilidad positiva. Entonces,

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(B|A_n)\mathbb{P}(A_n)}{\sum_{k \geq 1} \mathbb{P}(B|A_k)\mathbb{P}(A_k)}, \qquad n \geq 1. \tag{6}$$

Si los eventos $A_1, A_2, \dots$ se llaman "hipótesis", la fórmula (6) se considera como la probabilidad de ocurrencia de la hipótesis A_n sabiendo que ocurrió el evento B. En tal caso, $\mathbb{P}(A_n)$ es la probabilidad *a priori* de la hipótesis A_n y la fórmula (6) para $\mathbb{P}(A_n|B)$ se llama la *regla de Bayes para la probabilidad a posteriori* de la hipótesis A_n.

Nota Bene. Advertimos al lector que no trate de memorizar la fórmula (6). Matemáticamente, solo se trata de una forma especial de escribir la fórmula (5) y de nada más. $\square$

Ejemplo 1.10 (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 se puede representar en la forma de un diagrama de árbol tal como se indicó en la sección 1.2.

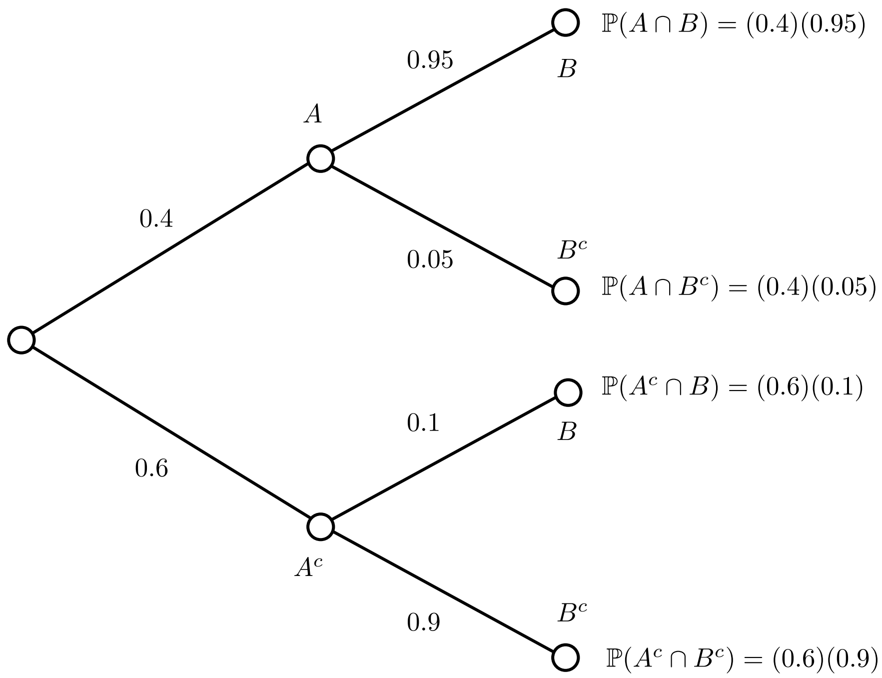


Figura 4: Observando el árbol se deduce que la probabilidad de recibir un 1 es $\mathbb{P}(B) = (0.4)(0.95) + (0.6)(0.1) = 0.44$. También se deduce que la probabilidad de que haya sido transmitido un 1 dado que se recibió un 1 es $\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} =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

Ejercicios adicionales

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

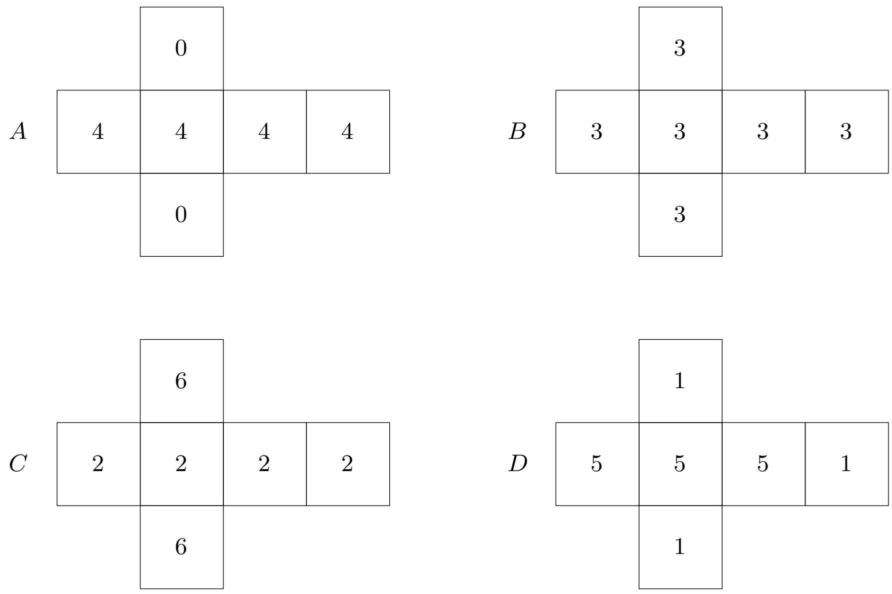


Figura 5: Dados de Efron

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?

2. Independencia estocástica

Definición 2.1 (Independencia estocástica)**.** Los eventos $A_1, A_2, \ldots, A_n$ son *mutuamente independientes* si satisfacen las siguientes $2^n - n - 1$ ecuaciones:

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_m}), \tag{7}$$

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 (7) se reduce a una condición: dos eventos A_1 y A_2 son independientes si satisfacen la ecuación

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2). \tag{8}$$

Ejemplo 2.2.

(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. □

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 2.3. 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:

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

□

Independencia y probabilidades condicionales. Para introducir el concepto de independencia no utilizamos probabilidades condicionales. 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 positivas. 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 2.4. Sean $A_1, A_2, \dots A_n$ eventos tales que todas las probabilidades $\mathbb{P}(A_i)$ son positivas. 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

$$\mathbb{P}(A_i|A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_i) \tag{9}$$

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

Ejercicios adicionales

2. 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.

3. Andrés, Francisco, Jemina e Ignacio fueron amigos en la escuela primaria. Se reencontraron en el curso 23 (PyE 61.09) de la FIUBA y se reunieron de a parejas a charlar. Como resultado de esas charlas, cada pareja renovó su amistad con probabilidad 1/2 y no lo hizo con probabilidad 1/2, independientemente de las demás. Posteriormente, Andrés recibió un rumor y lo transmitió a todas sus amistades. Suponiendo que cada uno de los que reciba un rumor lo transmitirá a todas sus amistades, cuál es la probabilidad de que Ignacio haya recibido el rumor transmitido por Andrés?.

---

3. 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 3.1. *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:

$$\mathbb{P}(A) := \sum_{\omega \in A} p(\omega). \tag{10}$$

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 (10) se llama *la medida de probabilidad inducida por p.*

Nota Bene 1. De la definición (10) 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 3.2 (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 3.3. 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 3.2) 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 3.4. 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 3.5 (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. □

4. Modelos continuos

4.1. 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 $A_1, \dots, A_k \in \mathcal{A}_0$ 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"*.

4.2. 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 4.1 (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

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

$\square$

"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,

$$\mathbb{P}(A) := \frac{\text{área}(A)}{\text{área}(\Lambda)}. \tag{11}$$

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 (11) resulta consistente con la noción de independencia.

Ejemplo 4.2. 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 $L/2$.

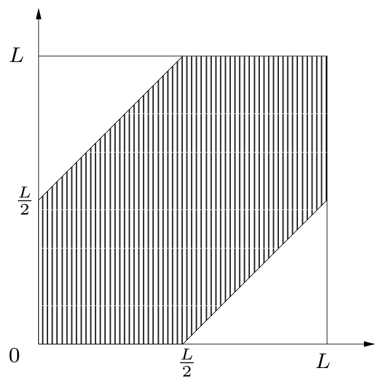


Figura 6: La región sombreada corresponde al evento A="la longitud del segmento limitado por los puntos x_1 y x_2 resulte menor que $L/2$".

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

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 6 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 4.3 (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

$$\mathbb{P}(A) = \frac{2l}{\pi a}. \tag{12}$$

La fórmula (12) 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:

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

□

4.3. 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. $\square$

4.4. 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 4.4. 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 4.5 (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,

$$\begin{aligned}
\iint_{\mathbb{R}^2} 2\pi\rho(x, y) dxdy &= \iint_{\mathbb{R}^2} \exp\left(-\frac{x^2 + y^2}{2}\right) dxdy \\
&= 2 \iint_{\mathbb{R}^2} \exp\left(-(x^2 + y^2)\right) dxdy \\
&= 2 \int_0^{2\pi} \left(\int_0^{\infty} e^{-\rho^2} \rho\, d\rho\right) d\theta \\
&= \int_0^{2\pi} \left(\int_0^{\infty} e^{-\rho^2} 2\rho\, d\rho\right) d\theta \\
&= 2\pi. \qquad (13)
\end{aligned}$$

Nota Bene. Observando con cuidado las identidades (13) 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 $\mathbb{R}$. □

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

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

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

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

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

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