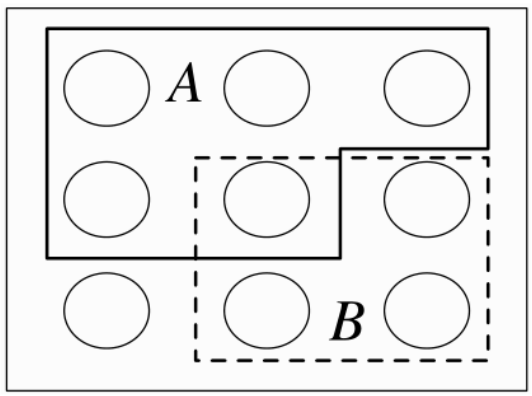
1 Probabilidad

Probabilidad es el lenguaje para cuantificar la incertidumbre.

1.1 Espacio muestral y eventos

El espacio muestral Ω de un experimento es el conjunto de todos los resultados o salidas del experimento. Puntos ω en Ω son llamados elementos, realizaciones o resultados muestrales. Subconjuntos de Ω son llamados **eventos**. El espacio muestral de un experimento puede ser finito, infinito contable o infinito no contable. En particular el conjunto vacio $\emptyset$ y el espacio muestra Ω son eventos. Cuando el espacio muestral es finito, se puede visualizar como la figura siguiente. Cada óvalo representa una salida y un evento es un conjunto de óvalos A o B.



Ejemplo 1.1 Si lanzamos una moneda dos veces, entonces $\Omega = \{SS, SC, CS, CC\}$. El evento que la primera cara es obtenida en los lanzamientos es $A = \{CS, CC\}$.

Ejemplo 1.2 En el experimento aleatorio de lanzar una moneda tres veces y observar los resultados el espacio muestral es el conjunto

$$\Omega = \{CCC, CCS, CSC, CSS, SCC, SCS, SSC, SSS\}.$$

El evento

$$A = \{CCC, CCS, CSS, CSC\} \subset \Omega$$

describe `una cara se obtuvo en el lanzamiento de la primera moneda`. En este caso tanto el espacio muestral ω como el evento A son conjuntos finitos.

Ejemplo 1.3 Sea ω el resultado de la medición de alguna cantidad física, por ejemplo, la temperatura, entonces $\Omega = \mathbf{R} = (-\infty, \infty)$. El evento en que la medida es mayor que 10 pero menos o igual a 23 es $A = (10, 23]$.

Ejemplo 1.4 Si lanzamos una moneda por siempre, entonces el espacio muestral es infinito

$$\Omega = \{\omega = (\omega_1, \omega_2, \ldots,) : \omega_i \in \{C, S\}\}$$

Sea A el evento en que la primera cara aparece en el tercer lanzamiento. Entonces

$$A = \{(\omega_1, \omega_2, \ldots,) : \omega_1 = S, \omega_2 = S, \omega_3 = S, \omega_i \in \{C, S\} \text{ para } i > 3\}.$$

Ejemplo 1.5 Considere un experimento aleatorio de lanzar un dardo en un tablero redondo, sin salirse del tablero. Suponiendo que el radio del tablero es 1, el espacio muestral es el conjunto de todos los vectores bidimensionales dentro del círculo unitario

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

Un evento que describe un golpe en el tablero puede puede ser

$$A = \left\{(x, y) \,:\, x, y \in \mathbb{R}, \sqrt{x^2 + y^2} < 0.1\right\} \subset \Omega.$$

En ambos casos el espacio muestral Ω y el evento A son infinitos contables.

El código R, siguiente, muestra todos los posibles eventos de un experimento con $\Omega = \{a, b, c\}$.

```
> Omega  = set("a", "b", "c")
> # mostramos un conjunto con todos los posibles
> # eventos de un experimento  en un espacio muestral Omega
> 2^Omega

{{}, {"a"}, {"b"}, {"c"}, {"a", "b"}, {"a", "c"}, {"b", "c"}, {"a", "b", "c"}}
```

Para un evento A, el resultado del experimento aleatorio $\omega \in \Omega$ está en A ($\omega \in A$) o no estén $A(w \notin A)$. En el primer caso, decimos que el evento A ocurrió y en el segundo caso decimos que el evento A no ocurrió.

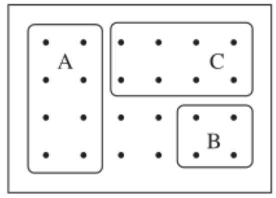
$A \cup B$ es el evento de que A o B ocurren y $A \cap B$ es el evento de que A y B ocurren. El complemento A^c (el complemento en el conjunto universal se toma como $\Omega : A^c = \Omega - A$) representa el evento de que A no ocurrió.

Si los eventos A, B son disjuntos ($A \cap B = \emptyset$), los dos eventos no pueden ocurrir al mismo tiempo, ya que ningún resultado del experimento aleatorio pertenece tanto a A como B. Si $A \subset B$ entonces si B ocurre implica que A también ocurre.

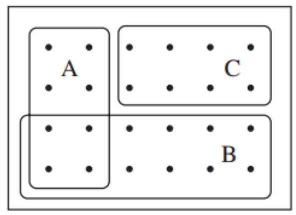
Ejemplo 1.6 Sea el evento A `lanzamiento de un número mayor que 3` y sea el evento B `lanzamiento de un número impar`. El evento A ocurre si un 4,5 o 6 es lanzado y el evento B ocurre si 1, 3 o 5 es lanzado. Así, ambos eventos ocurren si se lanza un 5 (este es el evento que es la intersección de los eventos A y B) y ninguno de los eventos ocurre si se lanza un 2 (este es el evento que es el complemento de la unión de A y B). En este ejemplo se tiene

$$A^c = \{1, 2, 3\}; \quad A \cup B = \{1, 2, 3, 4, 5, 6\}; \quad A \cap B = \{5\}; \quad A - B = \{4, 6\}$$

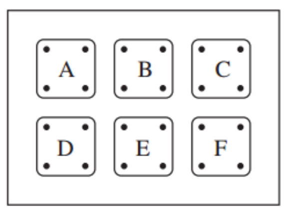
Definición 1.1 Una lista de eventos se dice <u>mutualmente exclusivo</u> si ningún elemento del espacio muestral pertenece a más de un evento. Como lo muestra la siguiente figura



Cuando todos los elementos en un espacio muestral pueden ser encontrados en al menos un evento en una lista de eventos, entonces la lista de eventos se dice colectivamente exhaustivos. En este caso ningún elemento del espacio muestral, es omitido y un único elemento puede pertenecer a más de un evento. Como se muestra en la siguiente figura



Se dice que los eventos que son mutuamente exclusivos y colectivamente exhaustivos, forman una partición del espacio muestral. Como se muestra en la figura.



Cualquier conjunto de eventos mutuamente excluyentes y colectivamente exhaustivos se denomina espacio de eventos.

Ejemplo 1.7 Secuencias de bits se transmiten a través de un canal de comunicación en grupos de cinco. Cada bit puede ser recibido correctamente o puede ser modificado en tránsito, lo que ocasiona un error. Consideremos un experimento que consiste en observar los valores de bits a medida que llegan e identificarlos con la letra c si el bit es correcto y con la letra e si el bit está en error. El espacio muestral consta de 32 resultados de *ccccc* a *eeeee*, de cero bits transmitidos incorrectamente a los cinco bits que están en error.

Sea el evento $A_i, i = 0, 1, \ldots, 5$, consistente de todos los resultados en los cuales los i bits están en error. Así $A_0 = \{ccccc\}, A_1 = \{ecccc, ceccc, ccecc, cccec, cccce\}$ y así sucesivamente hasta $A_5 = \{eeeee\}$. Los sucesos $A_i, i = 0, 1, \ldots 5$, particionan el espacio muestral y por lo tanto constituya un espacio de eventos.

Puede ser mucho más fácil trabajar en este pequeño espacio de eventos que en el espacio muestral más grande, especialmente si nuestro único interés es saber el número de bits transmitidos por error. Además, cuando los bits se transmiten en grupos más grandes, la diferencia es aun más importante. Con 16 bits por grupo en lugar de cinco, el espacio de eventos contiene ahora 17 eventos, mientras que el espacio muestral contiene 216 resultados.

1.2 Función probabilidad

Asignamos un número $\mathbb{P}(A)$ para cada evento A, llamada probabilidad de A. También llamamos a $\mathbb{P}$ distribución de probabilidad o medida de probabilidad. Las probabilidades son números reales en el intervalo cerrado $[0,1]$. Cuanto mayor sea el valor de la probabilidad, más probable es que ocurra el evento. Si un evento tiene probabilidad cero, ese evento no puede ocurrir; si tiene una probabilidad, entonces es seguro que ocurrirá.

Definición 1.2 Sea un espacio muestral Ω asociado a un experimento aleatorio. Una función probabilidad $\mathbb{P}$ es una función que asigna un número real a cada evento $A \subset \Omega$ y que satisface los siguientes tres axiomas

1. $\mathbb{P}(A) \geq 0, \quad \text{para todo} \quad A.$
2. $\mathbb{P}(\Omega) = 1.$
3. Sean $A_n, n \in \mathbb{P}$ es una secuencia de eventos disjuntos por pares $(A_i \cap A_j = \emptyset)$ siempre que $i \neq j$, entonces

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Podemos derivar algunas propiedades de $\mathbb{P}$ desde los anteriores axiomas.

1. $\mathbb{P}(\emptyset) = 0.$
2. $A \subset B \to \mathbb{P}(A) \leq \mathbb{P}(B).$
3. $0 \leq \mathbb{P}(A) \leq 1.$
4. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$
5. $A \cap B = \emptyset \to P(A \cup B) = P(A) + P(B).$

Podemos usar R, para demostrar una propiedad de la Probabilidad, definida en $\Omega = \{1,2,3,4\}$ usando $p_1 = 1/2, p_2 = 1/4, p_3 = p_4 = 1/8$.

```
> # Espacio muestral
> Omega = c(1, 2, 3, 4)
> # probabilidades de  4 eventos
> p = c(1/2, 1/4, 1/8, 1/8)
> # ellos suman 1
> sum(p)

[1] 1
```

Más propiedades de la probabilidad, pueden ser encontradas usando el paquete **sets** de R: `http://cran.r-project.org/web/packages/sets/index.html`.

Ejemplo 1.8 En el experimento de lanzamiento de una moneda, la probabilidad de obtener una cara en un solo lanzamiento es 0.5, ya que tenemos la misma probabilidad de obtener una cara como lo es para obtener un sello. Esto se escribe como

$$\mathbb{P}(C) = 0.5. \text{ o } \mathbb{P}(A_1) = 0.5$$

donde A_1 es el evento $\{C\}$.

Del mismo modo, la probabilidad de lanzar un 6 con un dado es $1/6$ y la probabilidad de elegir la reina de corazones es $1/52$. En estos casos, los eventos de cada espacio muestral tienen la misma probabilidad de ser el resultado en cualquier experimento dado. Estos eventos se llaman equiprobables y se dice que el resultado del experimento es aleatorio, ya que cada evento tiene la misma probabilidad de ocurrir.

En un espacio muestral que contiene n resultados igualmente probables, la probabilidad de que se produzca cualquier resultado particular es $1/n$. Naturalmente, podemos asignar probabilidades a eventos distintos de eventos elementales.

Ejemplo 1.9 Encuentre la probabilidad que debería estar asociada con el evento $A_2 = \{1, 2, 3\}$, es decir, lanzando un número menor que 4 usando un dado. Este evento ocurre si cualquiera de los números 1, 2 ó 3 es el resultado del lanzamiento. Puesto que cada uno tiene una probabilidad de $1/6$ y hay tres de ellos, la probabilidad del evento A_2 es la suma de las probabilidades de estos tres eventos elementales y por lo tanto es igual a 0.5.

Asignar probabilidades a los eventos es una parte importante del desarrollo de modelos de probabilidad. En algunos casos, sabemos de antemano las probabilidades de asociarse con eventos elementales, mientras que en otros casos deben ser estimadas. Si asumimos que una moneda y un dado son imparciales y la baraja de cartas completamente mezclada, entonces es fácil asociar las probabilidades con los elementos del espacio muestral y posteriormente con los eventos descritos en estos espacios muestrales. En otros casos, las probabilidades deben ser estimadas.

Se han desarrollado dos enfoques para definir las probabilidades: el enfoque de frecuencia relativa y el enfoque axiomático. El primero, como su nombre lo indica, consiste en realizar el experimento de probabilidad muchas veces, digamos N y contar el número de veces que ocurre un cierto evento, digamos n. Entonces se puede obtener una estimación de la probabilidad del evento como la frecuencia relativa n/N con la que se produce el evento, ya que se espera que, en el límite (límite en un sentido probabilístico) cuando $N \to \infty$, la relación n/N tiende a la probabilidad correcta del evento.

En términos matemáticos, esto se afirma de la siguiente manera: Dado que la probabilidad de un evento es p, entonces

$$\lim_{N\to\infty} \mathbb{P}\left\{ \left| \frac{n}{N} - p \right| > \epsilon \right\} = 0$$

Para cualquier $\epsilon > 0$. En otras palabras, no importa lo pequeño que elegimos ϵ, la probabilidad de que la diferencia entre n/N y p sea mayor que ϵ tiende a cero cuando $N \to \infty$. El uso de frecuencias relativas como estimaciones de probabilidad puede justificarse matemáticamente

El enfoque axiomático establece un pequeõ número de leyes o axiomas sobre los cuales se basa toda la teoria de la probabilidad. Fundamental a este concepto es el hecho de que es posible manipular probabilidades usando la misma álgebra lógica con el cual los eventos mismos son manipulados, como se muestra en los axiomas mostrados anteriormente.

1.3 Propiedades importantes de la probabilidad

Proposición 1.1 (Aditividad finita de la probabilidad) Para cada secuencia $A_1, \ldots A_n$ de eventos disjuntos dos a dos (por pares) $(A_i \cap A_j) = \emptyset$ siempre que $i \neq j$), entonces

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \cdots \mathbb{P}(A_n).$$

Proposición 1.2 (Principio de inclusión-exclusión) Para dos eventos A y B,

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

En general el principio de inclusión-exclusión se puede escribir como

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

El signo que precede a una suma con la intersección de m conjuntos es $(-1)^{m+1}$. La razón para sumar sobre índices crecientes es evitar el doble conteo.)

Se debe notar que si los conjuntos son disjuntos dos a dos, las intersecciones anteriores son todas vacías y por lo tanto tienen probabilidad cero y esto se reduce a la aditividad finita.

Proposición 1.3 (Continuidad de la probabilidad) Si $A_n \to A$ entonces

$$\mathbb{P}(A_n) \to \mathbb{P}(A) \qquad \text{si } n \to \infty$$

Ejemplo 1.10 Dos lanzamiento de moneda. Sea H_1 el evento que sale cara en el primer lanzamiento y sea H_2 el evento que ocurra cara en el segundo lanzamiento. Si todas los lanzamientos son igual de probables, entonces $\mathbb{P}(H_1 \cup H_2) = \mathbb{P}(H_1) + P(H_2) - \mathbb{P}(H_1 \cap H_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = 3/4$.

Ejemplo 1.11 Un sistema de una computadora usa passwords que son 5 caracteres y cada caracter es una de las 26 letras (a-z) o 10 enteros (0-9). El primer caracter tiene que ser una letra. Determina la proporción de passwords que

1. inician con una consonante.
2. terminan con un número par $(0, 2, 4, 6, 8)$.
3. inicia con una consonante o termina con un número par.

1. Suponiendo que el sistema de la computadora usa passwords que no distinguen las mayúsculas y las minúsculas. Sea A el evento en el cuál el password empieza con una consonante. Hay 21 consonantes, entonces $\mathbb{P}(A) = 21/26$. Aquí el denominador es 26, ya que todas las letras son posibles en el passwords.
2. Sea B ele evento que los passwords terminan en un número par $(0, 2, 4, 6, 8)$. Entonces $\mathbb{P}(B) = 5/36$. Aquí el denominador es 36, ya que todas las letras del alfabeto son 26 y 10 enteros son posibles en el passwords.
3. La propiedad de inclusión e exclusión, dice que la probabilidad de A o B es

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

   Para calcular $\mathbb{P}(A \cap B)$, la probabilidad que un password, inicie con una consonante y termine en un número par, calculamos el número posible de passwords. El primer carácter puede ser cualquiera de las 26 letras del alfabeto, y cada uno de los siguientes caracteres puede ser cualquiera de las 26 letras, así como cualquiera de los 10 números enteros, esto es 36 posibilidades en total. Así que el número posible de passwords es $(26)(36^4)$.

Ahora hay 21 maneras de empezar con una consonante y 5 maneras de terminar en un número par, así que el número posible de passwords que empiezan en una vocal y terminan en un número impar es $(21)(36^3)(5)$.

Por tanto

$$\mathbb{P}(A \cap B) = \frac{(21)(36^3)(5)}{(26)(36^4} = \left(\frac{21}{26}\right)\left(\frac{5}{36}\right).$$

Usando el principio de inclusión e inclusión que empieza con una vocal o termina con un número par

$$\mathbb{P}(A \cup B) = \frac{21}{26} + \frac{5}{36} - \left(\frac{21}{26}\right)\left(\frac{5}{36}\right).$$

En R

```
> 21/26 + 5/36 - (21/26)*(5/36)

[1] 0.8344017
```

esto significa que hay un 80% de posibilidad de seleccionar un passwords iniciando con una consonante y terminando en un número par.

Definición 1.3 Para un espacio muestral finito Ω un evento conteniendo un único elemento $A = \{\omega\}, \omega \in \Omega$ es llamado un evento elemental.

Si el espacio de muestra es finito $\omega = \{\omega_1, \ldots, \omega_n\}$ es relativamente sencillo definir la función de probabilidad definiendo n probabilidades de los eventos elementales. Más específicamente, para un espacio muestral con n elementos, supongamos que se nos da un conjunto de n números no negativos $\{p_\omega : \omega \in \Omega\}$ que suman a uno, entonces existe una función de probabilidad única $\mathbb{P}$ sobre esos eventos tales que $\mathbb{P}(\{\omega\}) = p_\omega$.

Esta probabilidad se define para eventos arbitrarios a través de la propiedad de aditividad finita

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{\omega \in A} p_\omega.$$

Un similar se cumple para espacios muestrales que son finitos contables.

En la clásica interpretación de probabilidad sobre espacios muestrales finitos, las probabilidades de todos los eventos elementales $\{\omega\}, \omega \in \Omega$ son iguales. Desde que la función de probabilidad debe satisfacer $\mathbb{P}(\Omega) = 1$, tenemos

$$\mathbb{P}(\{\omega\}) = |\Omega|^{-1}, \qquad \text{para todo} \quad \omega \in \Omega.$$

Esto implica que bajo el modelo clásico sobre un finito Ω, tenemos

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

Ejemplo 1.12 Consideremos el experimento de lanzar dos dados distintos y observar dos caras en orden. El espacio muestral es

$$\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\} = \{(x, y) : x, y \in \{1, 2, \ldots, 6\}\}$$

Desde que Ω tiene 36 elementos, la probabilidad del evento elemental $A = \{(4, 4\}$ es $\mathbb{P}(A) = 1/|\Omega| = 1/36$. La probabilidad de conseguir una suma de 9 en ambos dados es

$$\begin{aligned}
\mathbb{P}(\text{suma} = 9) &= \mathbb{P}(\{(6,3),(3,6),(4,5),(5,4)\}) \\
&= \frac{|\{(6,3),(3,6),(4,5),(5,4)\}|}{36} \\
&= \frac{4}{36}.
\end{aligned}$$

El modelo clásico en este caso es razonable, suponiendo que los dados son lanzados de forma independiente y son imparciales.

El siguiente código R demuestra el modelo clásico y las probabilidades resultantes en un Ω pequeño.

```
> Omega = set(1, 2, 3)
> # todos los posibles eventos
> 2^Omega

{{}, {1}, {2}, {3}, {1, 2}, {1, 3}, {2, 3}, {1, 2, 3}}
```

```
> # tam de todos los posibles eventos
> sapply(2^Omega, length)

[1] 0 1 1 1 2 2 2 3
```

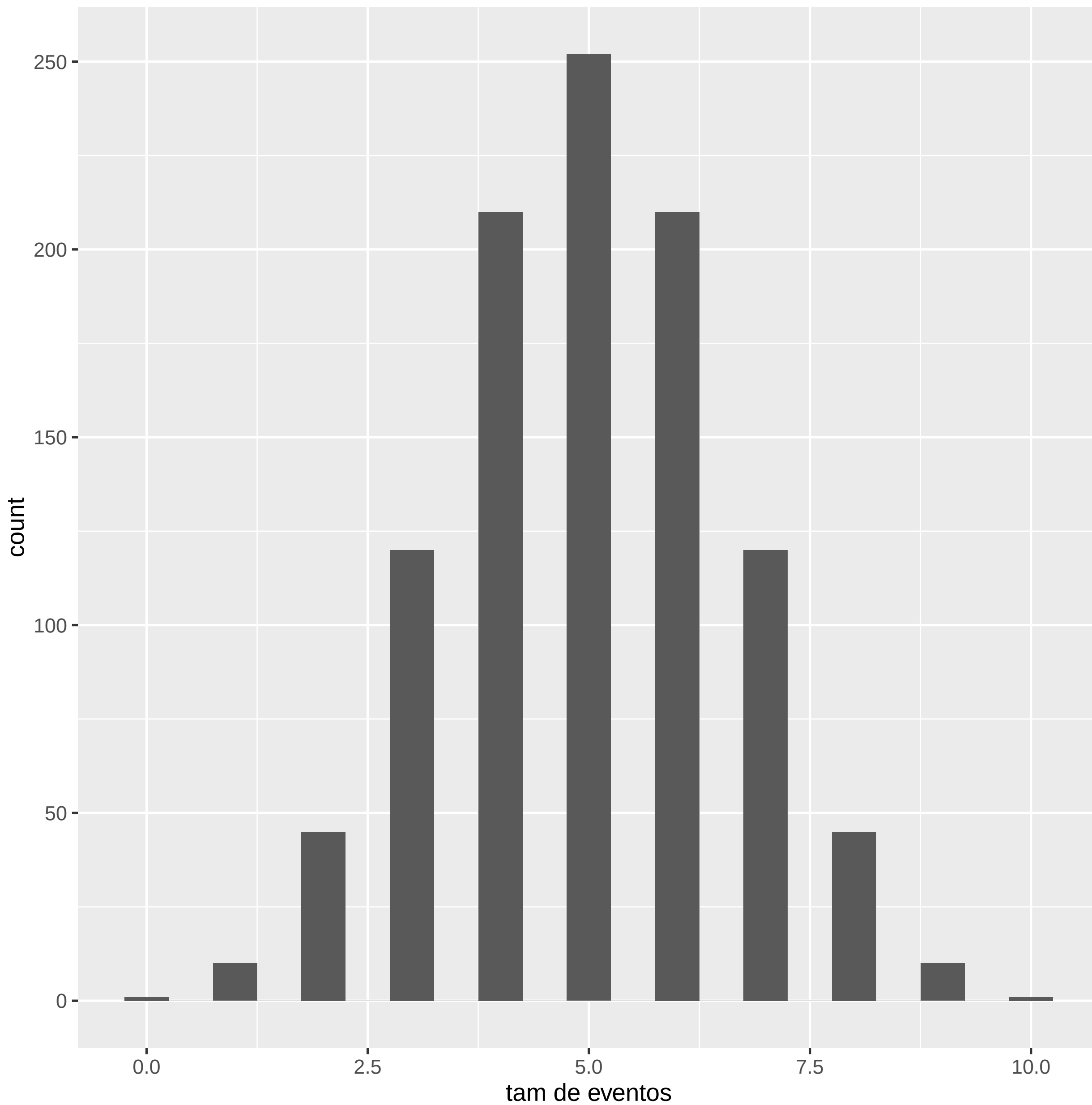
```
> # probabilidad de todos los elementos bajo el modelo clasico
> sapply(2^Omega, length)/length(Omega)

[1] 0.0000000 0.3333333 0.3333333 0.3333333 0.6666667 0.6666667 0.6666667 1.0000000
```

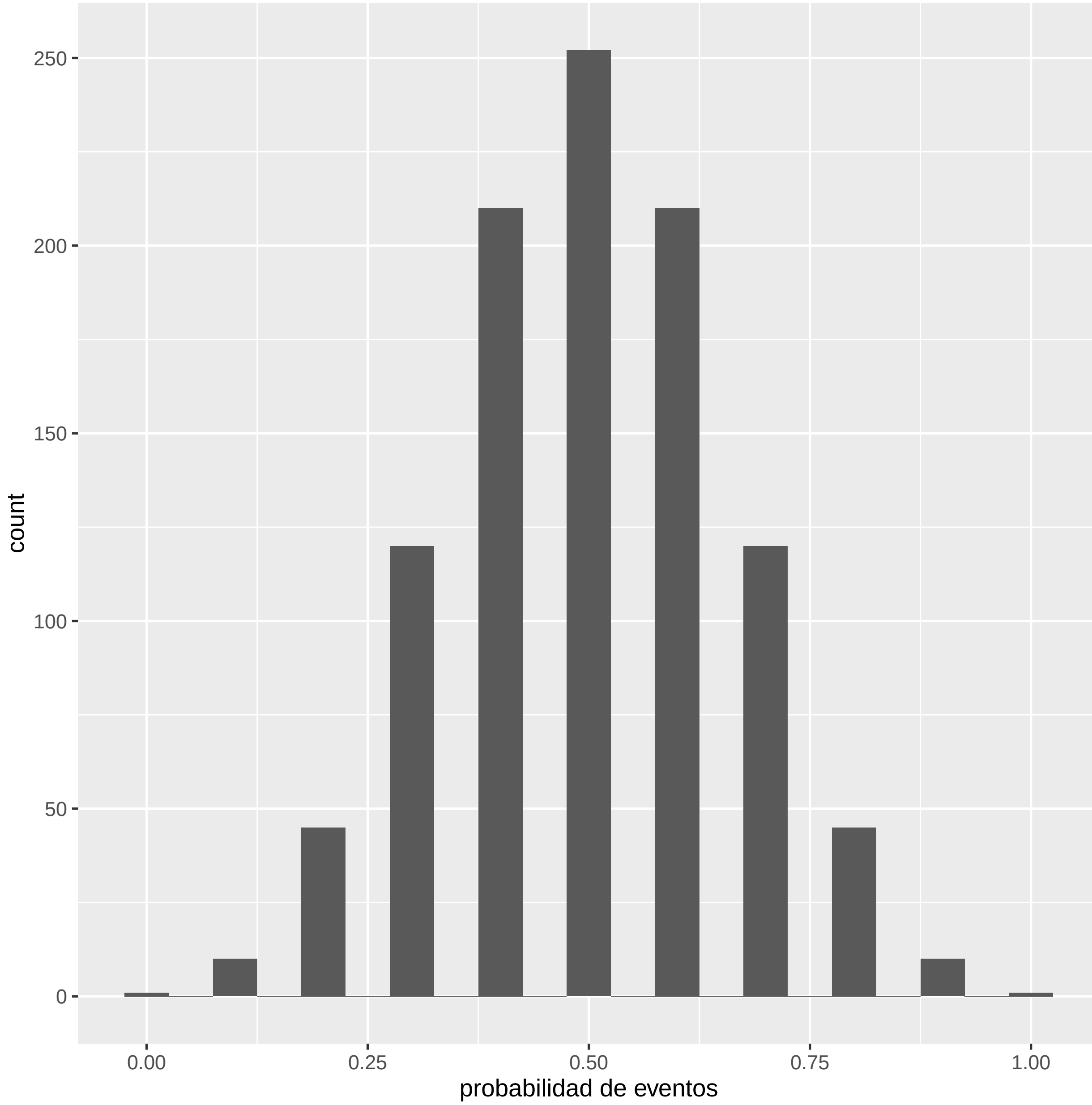
Note que la secuencia de probabilidades anterior no suma uno, ya que contiene las probabilidades de eventos no disjuntos.

El código R a continuación demuestra esto para un conjunto más grande usando un histograma de tamaños y probabilidades

```
> library(ggplot2)
> Omega = set(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
> # histograma tam vs probabilidades
> qplot(sapply(2^Omega, length), xlab = "tam de eventos", binwidth = 0.5, size = I(1/2))
```


250
200
150
100
50
0
count
0.0
2.5
5.0
7.5
10.0
tam de eventos

```
> # histograma tam vs probabilidades
> qplot(sapply(2^Omega, length)/length(Omega), xlab = "probabilidad de eventos", binwidth = 0.05)
```



Definición 1.4 Para un espacio muestral continuo de dimensión n (por ejemplo $\Omega = \mathbb{R}^n$), definimos la función de probabilidad clásica como

$$\mathbb{P}(A) = \frac{\mathrm{vol}_n(A)}{\mathrm{vol}_n(\Omega)},$$

donde $\mathrm{vol}_n(S)$ es el volumen n-dimensional del conjunto S. El volumen 1-dimensional de un conjunto $S \subset \mathbb{R}$ es su longitud. El volumen 2-dimensional de un conjunto $S \subset \mathbb{R}^2$ es su area. El volumen 3-dimensional de un conjunto $S \subset \mathbb{R}^3$ es su volumen. En general el volumen n-dimensional de A es la n-ésima integral de la función constante 1 sobre el conjunto A.

Ejemplo 1.13 En un experimento que mide el peso de los residentes en una región geográfica determinada, el espacio muestral es $\Omega = (0, 1000) \subset \mathbb{R}^1$ (suponiendo que nuestras unidades de medida son libras y las personas pesan menos de 1000 libras). La probabilidad de obtener una medida entre 150 y 250 (en el modelo clásico) es la relación de de los volúmenes o longitudes 1-dimensionales

$$\mathbb{P}((150, 250)) = \frac{|250 - 150|}{|1000 - 0|} = 0.1.$$

El modelo clásico en este caso es muy impreciso y no es probable que sea útil.

Ejemplo 1.14 Suponiendo que el modelo clásico en el espacio muestral del ejemplo 1.5, la probabilidad de golpear en el tablero es

$$\mathbb{P}\left(\left\{(x, y) \,:\, \sqrt{x^2 + y^2} < 0.1\right\}\right) = \frac{\pi\, 0.1^2}{\pi\, 1^2} = 0.01$$

(Ya que el área de un círculo de radio r es πr^2). El modelo clásico en este caso supone que la persona que lanza no hace ningún intento de golpear el centro. Para la mayoría de los lanzadores este modelo es inexacto.

- Para que el modelo clásico se aplique, el espacio de muestra Ω debe ser finito o ser continuo con un volumen finito no nulo.
- El modelo clásico (en espacios finitos y continuos) satisface los tres axiomas que definen una función de probabilidad.
- Una consecuencia del modelo clásico en espacios continuos es que la probabilidad de un evento elemental es cero (el volumen de un solo elemento es 0).

2 Referencias

1. Lecture 1: KC Border Introduction Introduction to Probability and Statistics Winter 2016.
2. Lecture 2: Probability Spaces; Random Variables; Conditional Probability; Independence KC Border Introduction to Probability and Statistics Winter 2016.
3. Chapter 1 Probability Probability, Markov Chains, Queues and Simulation Willian J.Stewart Princeton University Press 2009.