# Cálculo de probabilidades

## Capítulo 3: Familias paramétricas especiales de distribuciones univariadas

Mg. Sc. J. Eduardo Gamboa U.

# Familia paramétrica

Es una colección de densidades que están indexadas por una o más cantidades llamadas parámetros.

Por ejemplo, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) I_{(-\infty,\infty)}(x)$ es una familia paramétrica indexada con los parámetros $\sigma^2$ y $\mu$. Si $\sigma^2 = 4$ y $\mu = 10$ se tiene el siguiente miembro de la familia paramétrica:

$$f(x) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{1}{2\times 4}(x-10)^2\right) I_{(-\infty,\infty)}(x)$$

# Distribuciones discretas

1. Distribución uniforme discreta
2. Distribución Bernoulli
3. Distribución Binomial
4. Distribución hipergeométrica
5. Distribución Poisson
6. Distribución geométrica
7. Distribución Pascal
8. Distribución zeta

# Distribuciones continuas

1. Distribución uniforme continua
2. Distribución triangular
3. Distribución normal
4. Distribución lognormal
5. Distribución exponencial
6. Distribución gamma
7. Distribución Weibull
8. Distribución beta
9. Distribución Cauchy
10. Distribución Laplace
11. Distribución logística
12. Distribución Pareto

# Distribución uniforme discreta

## Función de probabilidad

Una variable aleatoria discreta $X$ que toma los valores $x_1$, ..., $x_N$ tiene distribución uniforme discreta si su función de probabilidad es:

$$f(x) = \frac{1}{N} I_{\{x_1,\ldots,x_N\}}(x)$$

Notación: $X \sim Unif\ Discreta(x_1, ..., x_N)$
Ejemplo:

- Resultado al lanzar un dado justo
- Resultado al sacar al azar un ticket, numerado del 1 al $n$
- Resultado de seleccionar una carta de corazones

## Valor esperado

$$E(X) = \frac{1}{N}\sum_{i=1}^{N} x_i$$

## Varianza

$$V(X) = \frac{1}{N}\sum_{i=1}^{N} (x_i - E(X))^2$$

## Función generatriz de momentos

$$M_X(t) = \frac{1}{N}\sum_{i=1}^{N} e^{tx_i}$$

## Función característica

$$\phi_X(t) = \frac{1}{N}\sum_{j=1}^{N} e^{itx_j}$$

## Implementación en R

- Paquete: distr
- Funciones:
  - DiscreteDistribution(supp, prob)
  - d(X): probabilidad en un punto
  - p(X): probabilidad acumulada
  - r(X): muestra aleatoria

## Ejemplo 1

La variable aleatoria $X$ es la cara de un dado justo obtenida luego de ser lanzado.

$$f(x) = \frac{1}{6} I_{\{1,2,3,4,5,6\}}(x)$$

$$E(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

$$V(X) = \frac{1}{6}\left((1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (4 - 3.5)^2 + (5 - 3.5)^2 + (6 - 3.5)^2\right)$$

$$V(X) = 2.9167$$

$$M_X(t) = \frac{1}{6}\left(e^t + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t}\right)$$

$$M_X^{(1)}(t) = \frac{1}{6}\left(e^t + 2e^{2t} + 3e^{3t} + 4e^{4t} + 5e^{5t} + 6e^{6t}\right)$$

$$M_X^{(1)}(0) = \frac{1}{6}\left(1 + 2 + 3 + 4 + 5 + 6\right) = 3.5 = E(X)$$

$$M_X^{(2)}(t) = \frac{1}{6}\left(e^t + 4e^{2t} + 9e^{3t} + 16e^{4t} + 25e^{5t} + 36e^{6t}\right)$$

$$M_X^{(2)}(0) = \frac{1}{6}\left(1 + 4 + 9 + 16 + 25 + 36\right) = 15.1667 = E(X^2)$$

$$V(X) = 15.1667 - 3.5^2 = 2.9167$$

$$\phi_X(t) = \frac{1}{6}\left(e^{it} + e^{2it} + e^{3it} + e^{4it} + e^{5it} + e^{6it}\right)$$

$$\phi_X^{(1)}(t) = \frac{1}{6}\left(ie^{it} + 2ie^{2it} + 3ie^{3it} + 4ie^{4it} + 5ie^{5it} + 6ie^{6it}\right)$$

$$\phi_X^{(1)}(0) = \frac{1}{6}\left(i + 2i + 3i + 4i + 5i + 6i\right) = 3.5i \Rightarrow E(X) = \frac{3.5i}{i} = 3.5$$

$$\phi_X^{(2)}(t) = \frac{1}{6}\left(i^2e^{it} + 4i^2e^{2it} + 9i^2e^{3it} + 16i^2e^{4it} + 25i^2e^{5it} + 36i^2e^{6it}\right)$$

$$\phi_X^{(2)}(0) = \frac{1}{6}\left(i^2 + 4i^2 + 9i^2 + 16i^2 + 25i^2 + 36i^2\right) = 15.1667i^2$$

$$E(X^2) = \frac{15.1667i^2}{i^2} = 15.1667$$

$$V(X) = 15.1667 - 3.5^2 = 2.9167$$

```r
library(distr)
library(distrEx)
X <- DiscreteDistribution(supp = 1:6,
                          prob = rep(1/6,6))
X |> E()
```

```
[1] 3.5
```

```r
X |> var()
```

```
[1] 2.916667
```

```r
3 |> d(X)()
```

```
[1] 0.1666667
```

```r
3 |> p(X)()
```

```
[1] 0.5
```

```r
3 |> r(X)()
```

```
[1] 5 1 3
```

## Ejemplo 2

Un laboratorio de cómputo puede ser usado en alguno de los siguientes horarios, con la misma probabilidad: 8, 10, 14 o 17 horas.

$$f(x) = \frac{1}{4} I_{\{8,10,14,17\}}(x)$$

$$E(X) = \frac{1}{4}(8 + 10 + 14 + 17) = 12.25$$

$$V(X) = \frac{1}{4}\left((8 - 12.25)^2 + (10 - 12.25)^2 + (14 - 12.25)^2 + (17 - 12.25)^2\right) = 12.1875$$

$$M_X(t) = \frac{1}{4}\left(e^{8t} + e^{10t} + e^{14t} + e^{17t}\right)$$

$$M_X^{(1)}(t) = \frac{1}{4}\left(8e^{8t} + 10e^{10t} + 14e^{14t} + 17e^{17t}\right)$$

$$M_X^{(1)}(0) = \frac{1}{4}\left(8 + 10 + 14 + 17\right) = 12.25 = E(X)$$

$$M_X^{(2)}(t) = \frac{1}{4}\left(64e^{8t} + 100e^{10t} + 196e^{14t} + 289e^{17t}\right)$$

$$M_X^{(2)}(0) = \frac{1}{4}\left(64 + 100 + 196 + 289\right) = \frac{649}{4} = 162.25 = E(X^2)$$

$$V(X) = 162.25 - 12.25^2 = 12.1875$$

$$\phi_X(t) = \frac{1}{4}\left(e^{8it} + e^{10it} + e^{14it} + e^{17it}\right)$$

$$\phi_X^{(1)}(t) = \frac{1}{4}\left(8ie^{8it} + 10ie^{10it} + 14ie^{14it} + 17ie^{17it}\right)$$

$$\phi_X^{(1)}(0) = \frac{1}{4}\left(8i + 10i + 14i + 17i\right) = 12.25i \Rightarrow E(X) = \frac{12.25i}{i} = 12.25$$

$$\phi_X^{(2)}(t) = \frac{1}{4}\left(64i^2e^{8it} + 100i^2e^{10it} + 196i^2e^{14it} + 289i^2e^{17it}\right)$$

$$\phi_X^{(2)}(0) = \frac{1}{4}\left(64i^2 + 100i^2 + 196i^2 + 289i^2\right) = 162.25i^2 \Rightarrow E(X^2) = \frac{162.25i^2}{i^2} = 162.25$$

$$V(X) = 162.25 - 12.25^2 = 12.1875$$

```r
X <- DiscreteDistribution(supp = c(8,10,14,17),
                          prob = rep(1/4,4))
X |> E()
```

```
[1] 12.25
```

```r
X |> var()
```

```
[1] 12.1875
```

```r
14 |> d(X)()
```

```
[1] 0.25
```

```r
14 |> p(X)()
```

```
[1] 0.75
```

```r
5 |> r(X)()
```

```
[1] 14 14  8 10 17
```

# Distribución Binomial

## Experimento Bernoulli

- Consiste en una secuencia de n ensayos (muestreo con reemplazo, o sin reemplazo de una población infinita o muy grande), donde n se fija antes del experimento.
- Los ensayos son idénticos e independientes, y cada uno de ellos solo tiene dos posibles resultados: éxito (E) o fracaso (F), por ello se dice que es dicotómico.
- La probabilidad de éxito es conocida y constante de un ensayo a otro; se denota esta probabilidad por $P(E) = \pi$ y la probabilidad de fracaso es $P(F) = 1 - \pi$.
- La distribución Binomial es la suma de $n$ variables de Bernoulli independientes:

$$X = \sum_{i=1}^{n} B_i \text{ donde } B_i \sim Bern(\pi)$$

## Función de probabilidad

Dado un experimento binomial, entonces se define la V.A.D. X $=$ El número de éxitos en una secuencia de n ensayos Bernoulli independientes. Su función de probabilidad está dada por:

$$f(x) = P(X = x) = \binom{n}{x} \pi^x (1-\pi)^{n-x} I_{\{0,1,2,\ldots,n\}}(x)$$

Notación: $X \sim Bin(n, \pi)$

## Ejemplos

- X $=$ Número de correos promocionales abiertos $\rightarrow X \sim Bin(n = 1000, \pi = 0.20)$
- Y $=$ Número de piezas defectuosas producidas $\rightarrow Y \sim Bin(n = 30, \pi = 0.05)$
- L $=$ Número de respuestas correctas en un examen de opción múltiple (cuando se elige al azar) $\rightarrow L \sim Bin(n = 10, \pi = 0.25)$
- S $=$ Número de semillas que germinan $\rightarrow S \sim Bin(n = 200, \pi = 0.88)$

## Media

$$\mu_X = E(X) = n\pi$$

## Varianza

$$\sigma_X^2 = Var(X) = n\pi(1-\pi)$$

## Asimetría

- Cuando $\pi = 0.5$, la distribución es simétrica
- Cuando $\pi < 0.5$, la distribución es sesgada a la derecha
- Cuando $\pi > 0.5$, la distribución es sesgada a la izquierda

## Función generatriz de probabilidades

$$G_X(s) = E(s^x) = (1 - p + ps)^n$$

## Función generatriz de momentos

$$M_X(t) = E(e^{tX}) = (1 - p + pe^t)^n$$

## Función característica

$$\phi_X(t) = E(e^{itX}) = (1 - p + pe^{it})^n$$

## Implementación en R

- Paquete: stats
- Funciones:
  - dbinom: probabilidad en un punto
  - pbinom: probabilidad acumulada
  - qbinom: cuantil
  - rbinom: muestra aleatoria

## Ejemplo 3

Una tienda de ropa implementó una nueva estrategia de marketing basada en recomendaciones personalizadas. En cada una de las semanas previas, se observó que el 75% de los clientes que recibieron la recomendación terminaron comprando al menos una prenda.

Esta semana, llegaron 20 nuevos clientes, cada uno expuesto a la misma estrategia.

**Experimento Bernoulli:**

- Ensayo: llegada de un cliente a la tienda. Se tienen $n = 20$ ensayos
- Cada cliente que llega es independiente y tiene dos posibilidades (E=éxito o F=fracaso): E $=$ $\{$un cliente compra$\}$ y F $=$ $\{$un cliente no compra$\}$
- La probabilidad de éxito es $\pi = 0.75$, y la de fracaso es $1 - \pi = 0.25$

**Variable aleatoria:**

$X$ : número de clientes que hacen una compra, $X \sim Bin(n = 20, \pi = 0.75)$

**Distribución de probabilidades:**

$$f(x) = P(X = x) = \binom{20}{x} 0.75^{x} \times 0.25^{20-x} I_{\{0,1,2,\ldots,20\}}(x)$$

La cantidad media de clientes que hacen una compra:

$$\mu_X = E(X) = 20 \times 0.75 = 15$$

Su varianza:

$$\sigma_X^2 = V(X) = 20 \times 0.75 \times 0.25 = 3.75$$

Función generatriz de probabilidades:

$$G_X(s) = E(s^x) = (0.25 + 0.75s)^{20}$$

Función generatriz de momentos:

$$M_X(t) = E(e^{tX}) = (0.25 + 0.75e^t)^{20}$$

Función característica

$$\phi_X(t) = E(e^{itX}) = (0.25 + 0.75e^{it})^{20}$$

La probabilidad de que el número de clientes que hacen una compra sea mayor que 18:

$$P(X > 18) = P(X = 19) + P(X = 20) = \binom{20}{19} 0.75^{19} 0.25 + \binom{20}{20} 0.75^{20} 0.25^{0} = 0.0243$$

```
(px19 <- dbinom(x = 19, size = 20, prob = 0.75))
```

```
[1] 0.02114141
```

```
(px20 <- dbinom(x = 20, size = 20, prob = 0.75))
```

```
[1] 0.003171212
```

```
(px   <- px19 + px20)
```

```
[1] 0.02431262
```

La probabilidad de que el número de clientes que hacen una compra sea como máximo 14 es:

$$P(X \leq 14) = \sum_{i=1}^{14} P(X = i) = 0.383$$

```r
pbinom(q = 14, size = 20, prob = 0.75)
```

```
[1] 0.3828273
```

La cantidad mediana de clientes que hacen una compra es:

```
qbinom(p = 0.5, size = 20, prob = 0.75)
```

```
[1] 15
```

Esto significa que en al menos la mitad de las semanas, 15 o menos clientes realizan una compra.

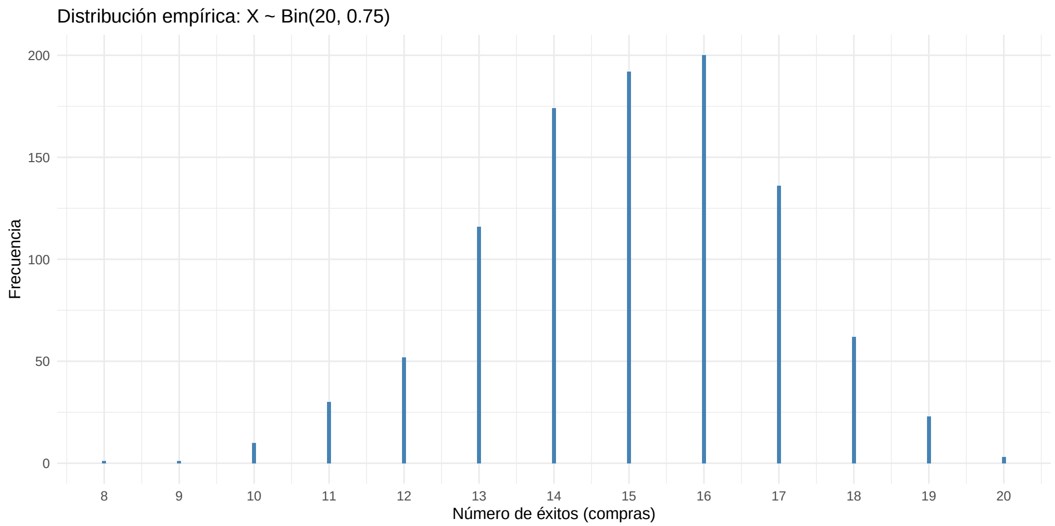
Simulación de una muestra aleatoria de la cantidad de clientes que hace una compra:

```r
set.seed(159)
rbinom(n = 10, size = 20, prob = 0.75)
```

```
 [1] 17 15 14 17 19 18 17 13 16 17
```

Se ha simulado 10 semanas, y en cada una de ellas la cantidad de clientes que realizaron una compra fue 17, 15, 14, etc.

```r
library(ggplot2)
set.seed(123)
x <- rbinom(n = 1000, size = 20, prob = 0.75)
df <- as.data.frame(table(x))
colnames(df) <- c("Éxitos", "Frecuencia")
df$Éxitos <- as.numeric(as.character(df$Éxitos))
ggplot(df, aes(x = Éxitos, y = Frecuencia)) +
  geom_bar(stat = "identity", fill = "steelblue", width = .05) +
  scale_x_continuous(breaks=8:20)+
  labs(title = "Distribución empírica: X ~ Bin(20, 0.75)",
       x = "Número de éxitos (compras)",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Bin(20, 0.75)
Frecuencia
200
150
100
50
0
8
9
10
11
12
13
14
15
16
17
18
19
20
Número de éxitos (compras)

# Distribución Hipergeométrica

Los supuestos que se consideran para una distribución hipergeométrica son:

- Se tiene una población de N elementos, individuos u objetos (una población finita)
- Cada elemento tiene dos posibles resultados: éxito (E) o fracaso (F). Existen A éxitos y (N-A) fracasos en la población.
- Se selecciona una muestra aleatoria sin reemplazo de tamaño $n$.

Al no reponer elementos, la probabilidad de selección varía en cada ensayo. Se aplica en poblaciones pequeñas, como en la prueba exacta de Fisher y en muestreos de aceptación por atributos.

## Función de probabilidad

Se define la V.A. Discreta X=Número de éxitos en la muestra de tamaño n. Su distribución de probabilidad es:

$$f(x) = P(X = x) = \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}} I_{\{\textbf{máx}(0,n+A-N),\ldots,\textbf{mín}(n,A)\}}(x)$$

Notación: $X \sim Hiper(N, n, A)$

## Ejemplos

- X = Número de focos defectuosos en una muestra
  $X \sim Hiper(N = 100, n = 15, A = 10)$
- Z = Número de expedientes con errores detectados en una auditoría
  $Z \sim Hiper(N = 200, n = 20, A = 30)$
- N = Número de estudiantes que recibieron una beca
  $N \sim Hiper(N = 80, n = 10, A = 25)$

## Media

$$\mu_X = E(X) = n\frac{A}{N}$$

## Varianza

$$\sigma_X^2 = Var(X) = n\frac{A}{N}\left(1 - \frac{A}{N}\right)\left(\frac{N-n}{N-1}\right)$$

## Asimetría

- Cuando $\frac{A}{N} \approx 0.5$ y $n \approx \frac{N}{2}$, la distribución es simétrica
- Cuando $\frac{A}{N}$ es pequeña, la distribución es sesgada a la izquierda.
- Cuando $\frac{A}{N}$ es grande, la distribución es sesgada a la derecha.

## Implementación en R

- Paquete: stats
- Funciones:
  - dhyper: probabilidad en un punto
  - phyper: probabilidad acumulada
  - qhyper: cuantil
  - rhyper: muestra aleatoria

## Diferencia entre Distribuciones Hipergeométrica y Binomial

- D. Binomial: La probabilidad del evento es constante en cada ensayo.
- D. Hipergeométrica: La probabilidad cambia en cada ensayo porque no hay reemplazo.
- Ejemplo (población de 5 personas, 3 con sangre O+):
  - $P(1^a$ persona O+$) = \frac{3}{5} = 0.6$
  - Si la primera persona es O+, $P(2^a$ persona O+$) = \frac{2}{4} = 0.5$
  - En la distribución hipergeométrica, cada selección altera las probabilidades, especialmente en poblaciones pequeñas.
- Ejemplo (población de 500 personas, 300 con sangre O+):
  - $P(1^a$ persona O+$) = \frac{300}{500} = 0.6$
  - Si la primera persona es O+, $P(2^a$ persona O+$) = \frac{299}{499} = 0.5992 \approx 0.6$
  - La diferencia es imperceptible, por el tamaño grande de la población. Podría considerarse como un caso de distribución Binomial con prob. de éxito $\pi = 0.6$.

## Ejemplo 4

Un embarque internacional de sustancias químicas ha llegado al puerto en 15 contenedores sellados. Por protocolo de bioseguridad, se debe realizar un muestreo aleatorio sin reemplazo para evaluar la pureza del producto antes de autorizar su ingreso al país.
Tras un informe preliminar, se sospecha que 2 de los 15 contenedores no cumplen con los requisitos de pureza. Para verificar la situación, se seleccionan aleatoriamente 3 contenedores para análisis de laboratorio.

**Variable aleatoria:**

Éxito=Contenedor que no cumplen con los estándares de pureza.

La V.A. Discreta: X=Número de contenedores que no cumplen con los estándares de pureza

$$X \sim Hiper(N = 15, n = 3, A = 2)$$

**Distribución de probabilidades:**

$$f(x) = P(X = x) = \frac{\binom{2}{x}\binom{13}{3-x}}{\binom{15}{3}} I_{\{\textbf{máx}(0,-10),\ldots,\textbf{mín}(3,2)\}}(x)$$

La cantidad media de contenedores que no cumplen con los estándares de pureza:

$$\mu_X = E(X) = 3 \times \frac{2}{15} = 0.4$$

Su varianza:

$$\sigma_X^2 = V(X) = 3 \times \frac{2}{15} \times \left(1 - \frac{2}{15}\right) \times \left(\frac{15-3}{15-1}\right) = 0.0248$$

La probabilidad de encontrar 1 contenedor que no cumpla con los estándares de pureza:

$$P(X = 1) = \frac{\binom{2}{1}\binom{13}{2}}{\binom{15}{3}} = 0.3428571$$

```r
dhyper(x = 1, m = 2, n = 13, k = 3) # m = A, n = N-A, k = n
```

```
[1] 0.3428571
```

La probabilidad de encontrar a lo más 1 contenedor que no cumpla con los estándares de pureza:

$$P(X \leq 1) = \frac{\binom{2}{0}\binom{13}{3}}{\binom{15}{3}} + \frac{\binom{2}{1}\binom{13}{2}}{\binom{15}{3}} = 0.9714285$$

```
dhyper(x = 0, m = 2, n = 13, k = 3) + dhyper(x = 1, m = 2, n = 13, k = 3)
```

```
[1] 0.9714286
```

```
phyper(q = 1, m = 2, n = 13, k = 3)
```

```
[1] 0.9714286
```

El percentil 65 de la cantidad de contenedores que no cumple con los estándares de pureza es:

```r
qhyper(p = 0.65, m = 2, n = 13, k = 3)
```

```
[1] 1
```

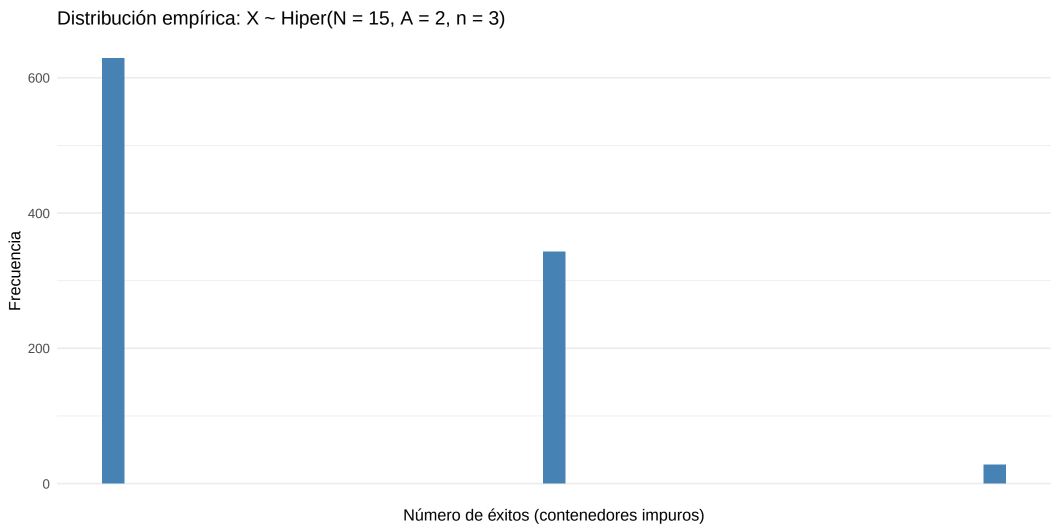
Es decir en al menos el 65% de las inspecciones se encontrará como máximo un contenedor que no cumple con los estándares de pureza.

Simulación de una muestra aleatoria de la cantidad de contenedores que no cumple con los estándares de pureza:

```
set.seed(159)
rhyper(nn = 15, m = 2, n = 13, k = 3)
```

```
 [1] 0 0 0 0 0 0 0 1 0 0 1 1 0 0 0
```

```r
library(ggplot2)
set.seed(123)
x <- rhyper(nn = 1000, m = 2,n = 13, k = 3)
df <- as.data.frame(table(x))
colnames(df) <- c("Éxitos", "Frecuencia")
df$Éxitos <- as.numeric(as.character(df$Éxitos))
ggplot(df, aes(x = Éxitos, y = Frecuencia)) +
  geom_bar(stat = "identity", fill = "steelblue", width = .05) +
  scale_x_continuous(breaks=8:20)+
  labs(title = "Distribución empírica: X ~ Hiper(N = 15, A = 2, n = 3)",
       x = "Número de éxitos (contenedores impuros)",
       y = "Frecuencia") +
  theme_minimal()
```



Distribución empírica: X ~ Hiper(N = 15, A = 2, n = 3)
Frecuencia
600
400
200
0
Número de éxitos (contenedores impuros)

# Distribución Poisson

- Distribución discreta propuesta por Siméon-Denis Poisson (1838)
- Modela la cantidad de eventos en un intervalo de tiempo, espacio o volumen, dada una frecuencia media.
- Se usa cuando los sucesos son independientes y tienen baja probabilidad individual.

## Proceso de Poisson

Es un experimento aleatorio en el que ocurren sucesos en un intervalo de longitud t:

- Los sucesos son de la misma clase u homogéneos.
- Los sucesos en un intervalo son independientes de los sucesos en otros intervalos no superpuestos.
- La probabilidad de más de un suceso en un intervalo muy pequeño es despreciable.
- El promedio de sucesos $(v)$ por unidad de intervalo $(t)$, es denotado por $\lambda = v \times t$

## Función de probabilidad

Se define la V.A.D. X=El número de sucesos que ocurren en intervalos de tamaño t. Su función de probabilidades es:

$$f(x) = P(X = x) = \begin{cases} \frac{e^{-vt}(vt)^x}{x!}, & x \in \{0, 1, 2, ...\} \\ 0, & \text{de otro modo} \end{cases}$$

$v$ = promedio de sucesos por unidad de intervalo.
$t$ = tamaño del intervalo (ejemplo: t = 2.3, t = 5.8, etc.).
$vt$ = promedio de sucesos por intervalo de tamaño t (tasa de ocurrencia)
Notación: $X \sim Pois(vt)$
También se puede expresar:

$$f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x \in \{0, 1, 2, ...\}$$

donde $\lambda = vt$, entonces $X \sim Pois(\lambda)$

## Media

$$\mu_X = E(X) = \lambda$$

## Varianza

$$\sigma_X^2 = Var(X) = \lambda$$

La igualdad de media y varianza es un rasgo característico que permite identificar si una variable puede ajustarse a esta distribución.

## Función generatriz de probabilidades

$$G_X(s) = e^{\lambda(s-1)}$$

## Función generatriz de momentos

$$M_X(t) = e^{\lambda(e^t-1)}$$

## Función característica

$$\phi_X(t) = e^{\lambda(e^{it}-1)}$$

### Implementación en R

- Paquete: stats
- Funciones:
  - dpois: probabilidad en un punto
  - ppois: probabilidad acumulada
  - qpois: cuantil
  - rpois: muestra aleatoria

## Ejemplos

- X $=$ Número de llamadas que recibe una central de emergencias en 1 minuto
  $X \sim Pois(4)$
- R $=$ Número de picaduras en una persona durante una noche en una zona tropical
  $R \sim Pois(2)$
- E $=$ Número de errores en una página de un libro impreso $E \sim Pois(0.3)$
- A $=$ Cantidad de pacientes que llegan a emergencias en una hora $A \sim Pois(6)$

## Ejemplo 5

Se cree que el número promedio de individuos por cada 2 $km^2$ de cierta especie de mamífero que habita en las alturas de cierta región es de 1.2.

**Variable aleatoria:**

V.A. X=Número de individuos en 2 $km^2$, $X \sim Pois(1.2)$

**Distribución de probabilidades:**

$$f(x) = P(X = x) = \frac{e^{-1.2} 1.2^x}{x!} \quad x \in \{0, 1, 2, ...\}$$

La cantidad media de individuos en un área de $3km^2$ es:

$$\mu_X = E(X) = 1.8$$

Su varianza:

$$\sigma_X^2 = V(X) = 1.8$$

Función generatriz de probabilidades:

$$G_X(s) = e^{1.8(s-1)}$$

Función generatriz de momentos:

$$M_X(t) = e^{1.8(e^t-1)}$$

Función característica:

$$\phi_X(t) = e^{1.8(e^{it}-1)}$$

Si se observa un área de 3 $km^2$ en dicha región, ¿cuál es la probabilidad de que se encuentren 3 individuos de esta especie?

Para 3 $km^2$: $\lambda = 3 \times 1.2/2 = 1.8$

X = Número de individuos en 3 $km^2$, $X \sim Pois(1.8)$

$$P(X = 3) = \frac{e^{-1.8}(1.8)^3}{3!} = 0.1606705$$

```r
dpois(x = 3, lambda = 1.8)
```

```
[1] 0.1606705
```

Si se observa un área de 3 $km^2$ en dicha región, ¿cuál es la probabilidad de que se encuentren como máximo 3 individuos de esta especie?

Para 3 $km^2$: $\lambda = 3 \times 1.2/2 = 1.8$

X = Número de individuos en 3 $km^2$, $X \sim Pois(1.8)$

$$P(X \leq 3) = \sum_{x=0}^{3} \frac{e^{-1.8}(1.8)^x}{x!} = \frac{e^{-1.8}(1.8)^0}{0!} + \frac{e^{-1.8}(1.8)^1}{1!} + \frac{e^{-1.8}(1.8)^2}{2!} + \frac{e^{-1.8}(1.8)^3}{3!} = 0.8913$$

```r
ppois(q = 3, lambda = 1.8)
```

```
[1] 0.8912916
```

Si se observa un área de 3 $km^2$ en dicha región, ¿cuál es la probabilidad de que se encuentren más de 3 individuos de esta especie?

Para 3 $km^2$: $\lambda = 3 \times 1.2/2 = 1.8$

X = Número de individuos en 3 $km^2$, $X \sim Pois(1.8)$

$$P(X > 3) = 1 - P(X \leq 3) = 1 - \sum_{x=0}^{3} \frac{e^{-1.8}(1.8)^x}{x!} = 1 - 0.8912 = 0.1087$$

```r
1 - ppois(q = 3, lambda = 1.8)
```

```
[1] 0.1087084
```

```r
ppois(q = 3, lambda = 1.8, lower.tail = FALSE)
```

```
[1] 0.1087084
```

El percentil 18 de la cantidad de individuos en un área de 3 $km^2$ es:

```r
qpois(p = 0.18, lambda = 3)
```

```
[1] 1
```

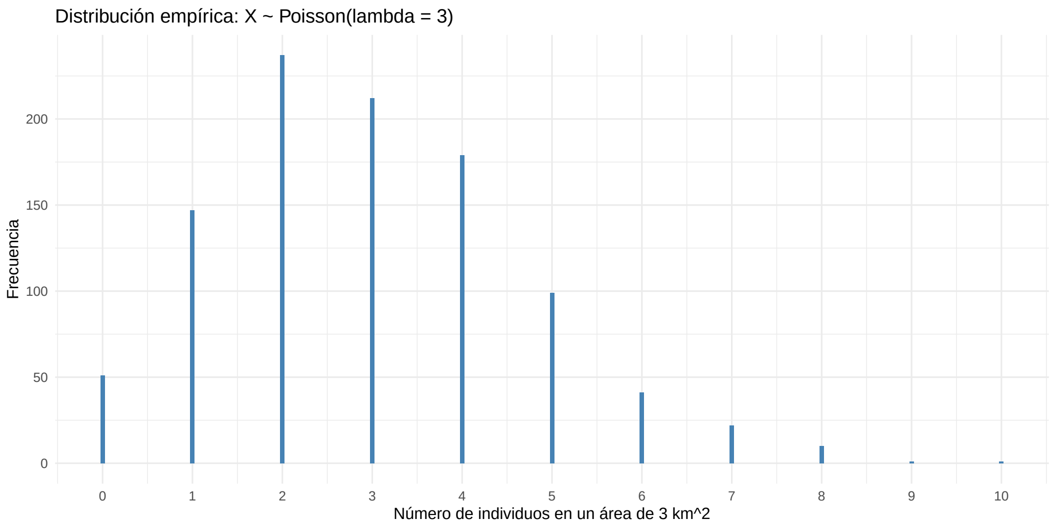
Es decir en al menos el 18% de las áreas de 3 $km^2$ se ecuentra como máximo un individuo.

Simulación de una muestra aleatoria de la cantidad de individuos en áreas de 3 $km^2$ es:

```r
set.seed(555)
rpois(n = 12, lambda = 3)
```

```
 [1] 2 5 4 3 3 0 7 2 0 5 4 2
```

```r
library(ggplot2)
set.seed(123)
x <- rpois(n = 1000, lambda = 3)
df <- as.data.frame(table(x))
colnames(df) <- c("Sucesos", "Frecuencia")
df$Sucesos <- as.numeric(as.character(df$Sucesos))
ggplot(df, aes(x = Sucesos, y = Frecuencia)) +
  geom_bar(stat = "identity", fill = "steelblue", width = .05) +
  scale_x_continuous(breaks=0:10)+
  labs(title = "Distribución empírica: X ~ Poisson(lambda = 3)",
       x = "Número de individuos en un área de 3 km^2",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Poisson(lambda = 3)
Frecuencia
200
150
100
50
0
0
1
2
3
4
5
6
7
8
9
10
Número de individuos en un área de 3 km^2

La distibución Poisson surge como límite de una binomial cuando $n \to \infty$, $\pi \to 0$ y $\lambda = n\pi$ es constante.

Por ejemplo, si $X \sim Bin(n = 2000, \pi = 0.001)$, entonces $X \to Y$, donde $Y \sim Pois(2)$

$$P(X = 3) = P(Y = 3) = 0.18$$

```r
dbinom(x = 3, size = 2000, prob = 0.001)
```

```
[1] 0.1805373
```

```r
dpois(x = 3, lambda = 2)
```

```
[1] 0.180447
```

# Distribución Geométrica

- Distribución discreta que modela el número de intentos hasta obtener el primer éxito en una secuencia de ensayos de Bernoulli independientes.
- Fue propuesta en el contexto de los primeros estudios de probabilidad con dados y juegos de azar, y es fundamental para modelar tiempos de espera discretos.
- Se aplica cuando los ensayos son independientes, con dos posibles resultados (éxito o fracaso) y una probabilidad constante de éxito.

## Función de probabilidad

Se define la V.A.D. X = el número de ensayos necesarios antes del primer éxito (es decir el número de fracasos). Su función de probabilidad es:

$$f(x) = P(X = x) = \pi(1 - \pi)^x I_{\{0,1,2,3,\ldots\}}(x)$$

Notación: $X \sim Geom(\pi)$

## Propiedad de falta de memoria

$$P(X > s + t | X > s) = P(X > t)$$

## Media

$$\mu_X = E(X) = \frac{1-\pi}{\pi}$$

## Varianza

$$\sigma_X^2 = V(X) = \frac{1-\pi}{\pi^2}$$

## Asimetría

La distribución geométrica no es simétrica, y se sesga a la derecha, especialmente cuando $\pi$ es pequeña.

## Función generatriz de probabilidades

$$G_X(s) = \frac{\pi s}{1 - (1 - \pi)s}, \quad |s| < \frac{1}{1 - \pi}$$

## Función generatriz de momentos

$$M_X(t) = \frac{\pi e^t}{1 - (1 - \pi)e^t}$$

## Función característica

$$\phi_X(t) = \frac{\pi e^{it}}{1 - (1 - \pi)e^{it}}$$

## Implementación en R

- Paquete: stats
- Funciones:
  - dgeom: probabilidad en un punto
  - pgeom: probabilidad acumulada
  - qgeom: cuantil
  - rgeom: muestra aleatoria

## Ejemplos

- X $=$ Número de intentos antes de que un cliente realice su primera compra en una tienda online $\rightarrow X \sim Geom(0.2)$
- D $=$ Número de inspecciones antes de encontrar el primer producto con algún defecto $\rightarrow D \sim Geom(0.05)$
- E $=$ Cantidad de intentos antes de que un estudiante resuelva correctamente un ejercicio sin ayuda $\rightarrow E \sim Geom(0.1)$

## Ejemplo 6

Se estima que el 15% de los correos enviados por una empresa reciben una respuesta.

**Variable aleatoria**

Éxito = Recibir una respuesta

La V.A. Discreta: X= Número de correos enviados antes de recibir la primera respuesta

$$X \sim Geom(\pi = 0.15)$$

**Distribución de probabilidades**

$$f(x) = P(X = x) = 0.85^x 0.15 \quad x \in \{1, 2, 3, ...\}$$

La cantidad media de correos enviados antes de recibir la primera respuesta es:

$$\mu_X = E(X) = \frac{1-0.15}{0.15} = 5.67$$

Su varianza:

$$\sigma_X^2 = V(X) = \frac{1-0.15}{0.15^2} = 37.778$$

Función generatriz de probabilidades:

$$G_X(s) = \frac{0.15s}{1-0.85s}, \quad |s| < \frac{1}{0.15}$$

Función generatriz de momentos:

$$M_X(t) = \frac{0.15e^t}{1-0.85e^t}$$

Función característica:

$$\phi_X(t) = \frac{0.15e^{it}}{1-0.85e^{it}}$$

Probabilidad de que se reciba la primera respuesta en el tercer correo enviado:

$$P(X = 2) = 0.85^2 \times 0.15 = 0.108375$$

```
dgeom(x = 2, prob = 0.15)
```

```
[1] 0.108375
```

Probabilidad de que se reciba la primera respuesta a más tardar el tercer correo enviado:

$$P(X \leq 2) = 0.85^0 \times 0.15 + 0.85^1 \times 0.15 + 0.85^2 \times 0.15 = 0.108375$$

```r
dgeom(x = 0, prob = 0.15) + dgeom(x = 1, prob = 0.15) +
  dgeom(x = 2, prob = 0.15)
```

```
[1] 0.385875
```

```r
pgeom(q = 2, prob = 0.15)
```

```
[1] 0.385875
```

Probabilidad de que se requieran más de 5 intentos para lograr obtener una respuesta:

$$P(X \geq 5) = \sum_{i=5}^{\infty} P(X = i) = P(X = 5) + P(X = 6) + ... = 0.4437$$

```r
pgeom(q = 4, prob = 0.15, lower.tail=FALSE)
```

```
[1] 0.4437053
```

```r
1-pgeom(q = 4, prob = 0.15)
```

```
[1] 0.4437053
```

Primer cuartil de la cantidad de intentos antes de recibir la primera respuesta

```
qgeom(p = 0.25, prob = 0.15)
```

```
[1] 1
```

Esto significa que en al menos el 25% de los casos se tiene como máximo un intento fallido antes de recibir la respuesta por correo.

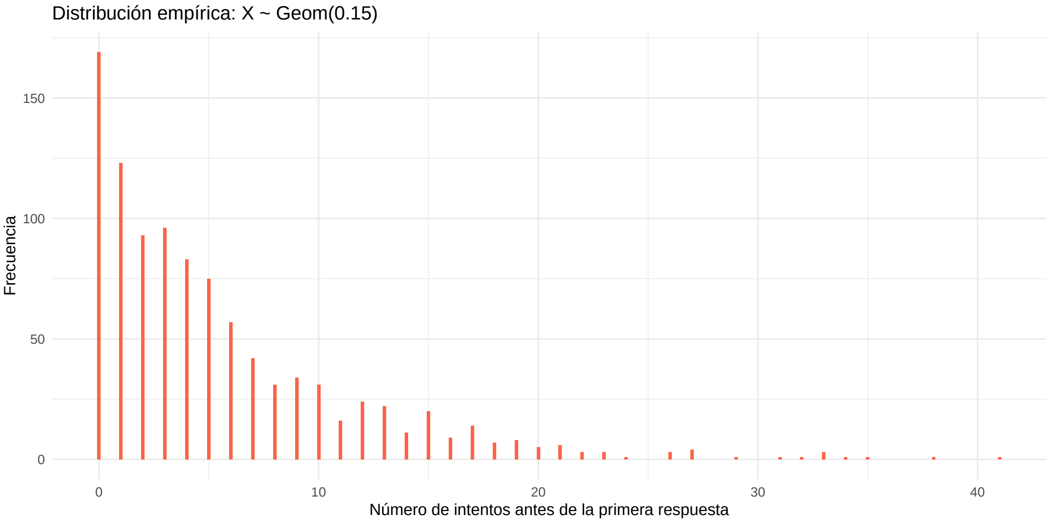
Simulación de una muestra aleatoria de la cantidad de correos enviados antes de recibir la respuesta.

```r
set.seed(2025)
rgeom(n = 10, prob = 0.15)
```

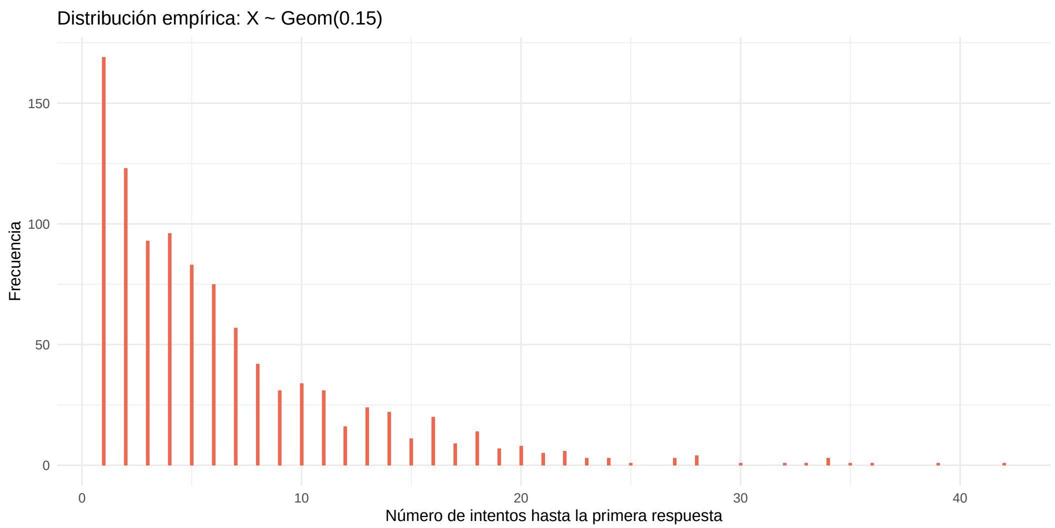
```
 [1]  2  0  3  0  2  4  1  8 19  2
```

El primer correo se tuvo que enviar dos veces y se obtuvo respuesta en el tercer envío, el segundo correo obtuvo respuesta en el envío original, el tercer correo obtuvo respuesta en el cuarto correo enviado (3 correos fallidos previos), etc.

```r
library(ggplot2)
set.seed(2025)
x <- rgeom(1000, prob = 0.15)
df <- as.data.frame(table(x))
df$x <- as.numeric(as.character(df$x))

ggplot(df, aes(x = x, y = Freq)) +
  geom_bar(stat = "identity", fill = "tomato", width = 0.15) +
  labs(title = "Distribución empírica: X ~ Geom(0.15)",
       x = "Número de intentos antes de la primera respuesta",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Geom(0.15)
Frecuencia
150
100
50
0
0
10
20
30
40
Número de intentos antes de la primera respuesta

```r
library(ggplot2)
set.seed(2025)
x <- rgeom(1000, prob = 0.15) + 1
df <- as.data.frame(table(x))
df$x <- as.numeric(as.character(df$x))

ggplot(df, aes(x = x, y = Freq)) +
  geom_bar(stat = "identity", fill = "tomato", width = 0.15) +
  labs(title = "Distribución empírica: X ~ Geom(0.15)",
       x = "Número de intentos hasta la primera respuesta",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Geom(0.15)
Frecuencia
150
100
50
0
0
10
20
30
40
Número de intentos hasta la primera respuesta

# Distribución Binomial negativa

- Distribución discreta que modela el número de fracasos antes de alcanzar un número fijo de éxitos en una secuencia de ensayos de Bernoulli independientes.
- Se aplica cuando se desea conocer cuántos intentos fallidos ocurren antes de lograr el r-ésimo éxito, con una probabilidad constante de éxito $\pi$
- Es una generalización de la distribución geométrica (que corresponde al caso particular donde $r = 1$).
- También es conocida como distribución de Pascal.

## Función de probabilidad

Se define la V.A.D. X = número de fracasos antes del r-ésimo éxito. Su función de probabilidad es:

$$f(x) = P(X = x) = \binom{x+r-1}{x}(1-\pi)^x \pi^r I_{\{0,1,2,3,\ldots\}}(x)$$

Notación: $X \sim BinNeg(r, \pi)$
Si $r = 1$, se trata de una $Geom(\pi)$

## Media

$$\mu_X = E(X) = r \times \frac{1-\pi}{\pi}$$

## Varianza

$$\sigma^2_X = V(X) = r \times \frac{1-\pi}{\pi^2}$$

## Función generatriz de probabilidades

$$G_X(s) = \left(\frac{\pi s}{1-(1-\pi)s}\right)^r, \quad |s| < \frac{1}{1-\pi}$$

## Función generatriz de momentos

$$M_X(t) = \left(\frac{\pi e^t}{1-(1-\pi)e^t}\right)^r$$

## Función característica

$$\phi_X(t) = \left(\frac{\pi e^{it}}{1-(1-\pi)e^{it}}\right)^r$$

## Implementación en R

- Paquete: stats
- Funciones:
  - dnbinom: probabilidad en un punto
  - pnbinom: probabilidad acumulada
  - qnbinom: cuantil
  - rnbinom: muestra aleatoria

## Ejemplos

- X = Número de intentos fallidos antes de que un estudiante acierte 3 respuestas correctas en una trivia $\rightarrow X \sim BinNeg(r = 3, \pi = 0.25)$
- V = Número de ventas fallidas antes de conseguir 5 ventas exitosas $\rightarrow V \sim BinNeg(r = 5, \pi = 0.18)$
- R = Número de fallas que ocurren antes de que un robot logre completar correctamente 6 ensamblajes exitosos $R \sim BinNeg(r = 6, \pi = 0.3)$

## Ejemplo 7

Un agente de ventas logra cerrar un trato con probabilidad 0.2. Se desea saber cuántos intentos fallidos ocurren, en promedio, antes de cerrar 3 ventas exitosas.

**Variable aleatoria**

Éxito = Cerrar una venta

V.A. X = Número de intentos fallidos antes de lograr 3 ventas

$$X \sim BinNeg(r = 3, \pi = 0.2)$$

**Distribución de probabilidad**

$$f(x) = P(X = x) = \binom{x+2}{x} 0.8^x 0.2^3 \qquad x \in \{0, 1, 2, 3, ...\}$$

La cantidad media de intentos fallidos antes de lograr las 3 ventas es

$$\mu_X = E(X) = 3 \times \frac{1 - 0.2}{0.2} = 12$$

Su varianza:

$$\sigma_X^2 = V(X) = 3 \times \frac{1 - 0.2}{0.2^2} = 60$$

Función generatriz de probabilidades:

$$G_X(s) = \left(\frac{0.2s}{1-0.8s}\right)^3, \quad |s| < \frac{1}{0.8}$$

Función generatriz de momentos:

$$M_X(t) = \left(\frac{0.2e^t}{1-0.8e^t}\right)^3$$

Función característica:

$$\phi_X(t) = \left(\frac{0.2e^{it}}{1-0.8e^{it}}\right)^3$$

Probabilidad de que ocurran exactamente 5 intentos fallidos antes de cerrar exitosamente 3 ventas:

$$
exisP(X = 5) = \binom{5+2}{5} \times 0.8^5 \times 0.2^3 = 0.05505024
$$

```r
dnbinom(x = 5, size = 3, prob = 0.2)
```

```
[1] 0.05505024
```

Probabilidad de que ocurran como máximo 5 intentos fallidos antes de cerrar exitosamente 3 ventas:

$$P(X \leq 5) = \sum_{x=0}^{5} \binom{x+2}{x} \times 0.8^{x} \times 0.2^{3} = 0.20308$$

```r
dnbinom(x=0, size=3, prob=0.2) + dnbinom(x=1, size=3, prob=0.2) +
  dnbinom(x=2, size=3, prob=0.2) + dnbinom(x=3, size=3, prob=0.2) +
  dnbinom(x=4, size=3, prob=0.2) + dnbinom(x=5, size=3, prob=0.2)
```

```
[1] 0.2030822
```

```r
pnbinom(q = 5, size = 3, prob = 0.2)
```

```
[1] 0.2030822
```

Tercer cuartil de la cantidad de fracasos antes de lograr 3 ventas

```r
qnbinom(p = 0.75, size = 3, prob = 0.2)
```

```
[1] 16
```

En al menos el 75% de los casos, el vendedor fallará como máximo 16 veces antes de cerrar existosamente 3 ventas.

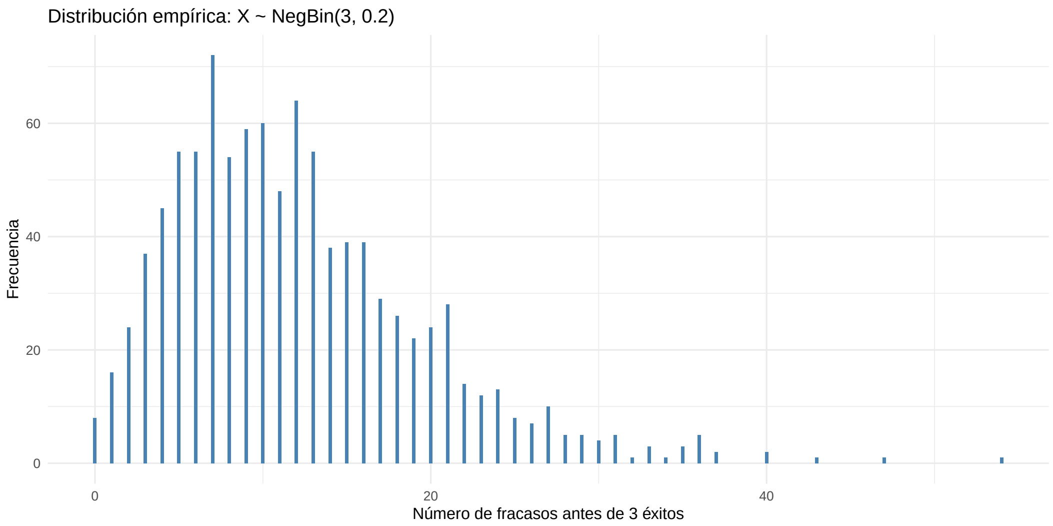
Simulación de una muestra aleatoria de 10 casos de vendedores:

```r
set.seed(75321)
rnbinom(n = 10, size = 3, prob = 0.2)
```

```
 [1] 16 13 12 17 31  2 17 24  9 14
```

El primer vendedor falla 16 veces antes de conseguir su tercera venta exitosa, el segundo falla 13 veces antes de las 3 ventas exitosas, etc.

```r
library(ggplot2)
set.seed(75321)
x <- rnbinom(1000, size = 3, prob = 0.2)
df <- as.data.frame(table(x))
df$x <- as.numeric(as.character(df$x))

ggplot(df, aes(x = x, y = Freq)) +
  geom_bar(stat = "identity", fill = "steelblue", width = 0.2) +
  labs(title = "Distribución empírica: X ~ NegBin(3, 0.2)",
       x = "Número de fracasos antes de 3 éxitos",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ NegBin(3, 0.2)
Frecuencia
0
20
40
60
0
20
40
Número de fracasos antes de 3 éxitos

# Distribución zeta

Una variable aleatoria discreta $X$ tiene distribución Zeta si toma valores enteros positivos y su función de probabilidad decrece según una ley de potencia. Es una distribución de cola pesada.

## Función de probabilidad

$$f(x) = P(X = x) = \frac{1}{\zeta(s+1)x^{s+1}} I_{\{1,2,3,\ldots\}}(x), \quad s > 0$$

donde:

- $s$ es el parámetro de forma
- $\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s}$ es la función zeta de Riemann.

Notación: $X \sim Zeta(s)$

## Media

$$E(X) = \frac{\zeta(s)}{\zeta(s+1)} \qquad s > 3$$

Si $0 < s \leq 1$, $E(X) = \infty$

## Varianza

$$V(X) = \frac{\zeta(s-1)}{\zeta(s+1)} - \left(\frac{\zeta(s)}{\zeta(s+1)}\right)^2 \qquad s > 2$$

Si $0 < s \leq 2$, $V(X) = \infty$

## Implementación en R

- Paquete: VGAM
- Funciones:
  - zeta: Función zeta
  - dzeta: probabilidad en un punto
  - pzeta: probabilidad acumulada
  - qzeta: cuantil

## Ejemplos

- En el estudio de frecuencia de palabras en un texto, $X$ es el rango de frecuencia para una palabra elegida al azar, entonces $X \sim Zeta(s = 1.1)$
- En una plataforma de videos como YouTube, muchos videos tienen pocas vistas, y pocos acumula muchísimas visualizaciones. $X$ es el rango de popularidad de un video elegido al azar, entonces $X \sim Zeta(s = 2)$

## Ejemplo 8

En un servidor web, hay muchos archivos pequeños y pocos archivos muy grandes.

**Variable aleatoria**

V.A. X = Tamaño ordenado por rango

$$X \sim Zeta(s = 4.1)$$

**Distribución de probabilidad**

$$f(x) = P(X = x) = \frac{1}{\zeta(5.1)x^{5.1}} I_{\{1,2,3,\ldots\}}(x)$$

El rango medio:

$$\mu_X = \frac{\zeta(4.1)}{\zeta(5.1)}$$

```
library(VGAM)
zeta(4.1)/zeta(5.1)
```

```
[1] 1.040186
```

```
x  <- 1:100
fx <- dzeta(1:100, shape = 4.1)
sum(x*fx)
```

```
[1] 1.040186
```

Su varianza:

$$V(X) = \frac{\zeta(3.1)}{\zeta(5.1)} - \left(\frac{\zeta(4.1)}{\zeta(5.1)}\right)^2$$

```
zeta(3.1)/zeta(5.1) - (zeta(4.1)/zeta(5.1))**2
```

```
[1] 0.06227857
```

```
x  <- 1:100
fx <- dzeta(1:100, shape = 4.1)
sum(x**2*fx) - (sum(x*fx))**2
```

```
[1] 0.06225022
```

Probabilidad de seleccionar el segundo archivo más grande

$$P(X=2)=\frac{1}{2^{5.1}}*\frac{1}{\zeta(5.1)}=0.054$$

```r
dzeta(x = 2, shape = 4.1)
```

```
[1] 0.02819347
```

```r
1/2**5.1 * 1/zeta(5.1)
```

```
[1] 0.02819347
```

Probabilidad de seleccionar el primero, segundo, tercero, cuarto y quinto archivos más grandes

```r
dzeta(x = 1:5, shape = 4.1) |> round(3)
```

```
[1] 0.967 0.028 0.004 0.001 0.000
```

Probabilidad de seleccionar hasta el tercer archivo más grande:

$$P(X \leq 3) = 0.994$$

```r
pzeta(q = 3, shape = 4.1)
```

```
[1] 0.9987032
```

```r
dzeta(x = 1:3, shape = 4.1) |> sum()
```

```
[1] 0.9987032
```

```r
1/(1**5.1*zeta(5.1)) + 1/(2**5.1*zeta(5.1)) + 1/(3**5.1*zeta(5.1))
```

```
[1] 0.9987032
```

Percentil 98 del rango de elección de archivos

```r
qzeta(p = 0.98, shape = 4.1)
```

```
[1] 2
```

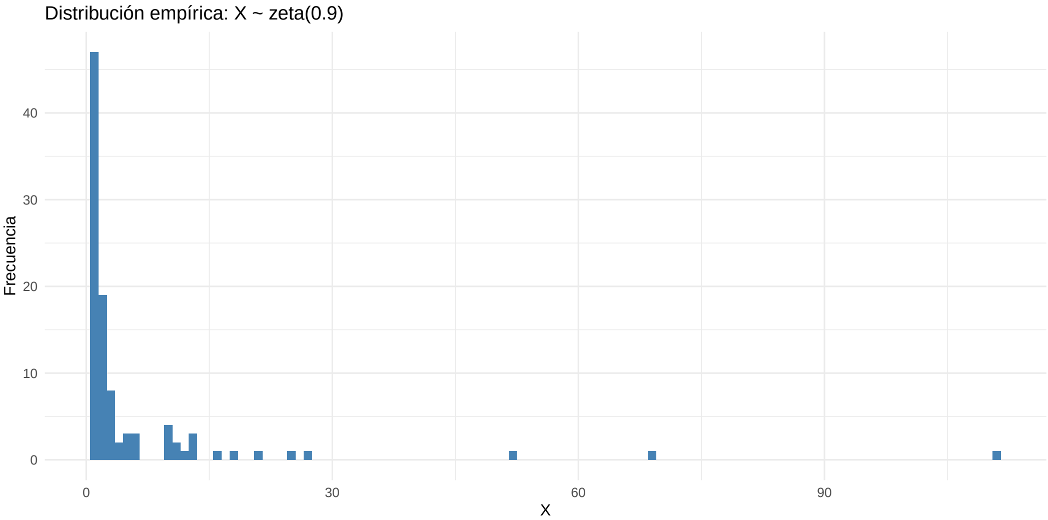
Ejemplo de simulación aleatoria.

Supongamos que deseamos simular 12 datos de $Zeta(s = 0.9)$

```r
rzeta(n = 12, shape = 0.9)
```

```
 [1] 1 1 2 1 1 1 1 7 2 1 1 1
```

```r
library(ggplot2)
set.seed(21)
x <- rzeta(n = 100, shape = 0.9)
df <- as.data.frame(table(x))
df$x <- as.numeric(as.character(df$x))

ggplot(df, aes(x = x, y = Freq)) +
  geom_bar(stat = "identity", fill = "steelblue", width = 1) +
  labs(title = "Distribución empírica: X ~ zeta(0.9)",
       x = "X",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ zeta(0.9)
Frecuencia
40
30
20
10
0
0
30
60
90
X

# Distribución Uniforme Continua

- Distribución continua que modela situaciones donde todos los valores dentro de un intervalo son igualmente probables
- Representa la máxima ignorancia, incertidumbre o entropía: no se conoce ninguna concentración de probabilidad dentro del intervalo
- Comúnmente usada como modelo inicial en simulaciones, como componente de algoritmos estocásticos y como distribución a priori no informativa en inferencia bayesiana.
- La mayoría de los algoritmos de simulación comienzan generando variables $U \sim Unif(0, 1)$, que luego se transforman para obtener otras distribuciones.

## Función de densidad

Una variable aleatoria continua $X$ tiene distribución uniforme en el intervalo $[a, b]$ si su función de densidad es:

$$f_X(x) = \frac{1}{b-a} I_{[a,b]}(x)$$

Notación: $X \sim U(a, b)$

## Ejemplos

- T = Tiempo de espera para el bus si llega en cualquier momento dentro de un intervalo de 10 minutos $\rightarrow T \sim Unif(0, 10)$
- U = Temperatura ambiente durante una hora establecida si se mantiene entre 18 y 22 grados Celsius $\rightarrow U \sim Unif(18, 22)$
- C = Tiempo de carga de un archivo entre 30 y 50 segundos cuando no hay congestión $\rightarrow C \sim Unif(30, 50)$

## Función de distribución acumulada

$$F_X(x) = \frac{x-a}{b-a} I_{[a,b]}(x) + I_{(b,\infty)}(x)$$

## Media

$$\mu_X = E(X) = \frac{a+b}{2}$$

## Varianza

$$\sigma_x^2 = V(X) = \frac{(b-a)^2}{12}$$

## Función generatriz de momentos

$$M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

## Función característica

$$\phi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

## Implementación en R

- Paquete: stats
- Funciones:
  - dunif: densidad en un punto
  - punif: probabilidad acumulada
  - qunif: cuantil
  - runif: muestra aleatoria

## Ejemplo 9

En una planta embotelladora, una máquina realiza inspecciones automáticas de botellas a intervalos aleatorios entre 12 y 20 segundos. No hay preferencia por ningún valor dentro de ese rango.

**Variable aleatoria**

V.A. X = Tiempo (en segundos) hasta la próxima inspección, $X \sim Unif(12, 20)$

**Función de densidad**

$$f(x) = \frac{1}{8} I_{[12,20]}(x)$$

**Función de distribución acumulada**

$$F(x) = \frac{x - 12}{8} I_{[12,20]}(x) + I_{(20,\infty)}(x)$$

El tiempo medio hasta la próxima inspección:

$$\mu_X = E(X) = \frac{12 + 20}{2} = 16$$

Su varianza:

$$\sigma_X^2 = V(X) = \frac{(20 - 12)^2}{12} = 5.333$$

Probabilidad de que una inspección ocurra entre 13 y 16 segundos:

$$P(13 \leq X \leq 16) = \int_{13}^{16} \frac{1}{8} dx = \frac{16 - 13}{8} = \frac{3}{8} = 0.375$$

```
[1] 0.375
```

$$P(13 \leq X \leq 16) = P(X \leq 16) - P(X \leq 13) = F(16) - F(13) = \frac{16 - 12}{8} - \frac{13 - 12}{8} = \frac{3}{8} = 0.375$$

```r
punif(16, min = 12, max = 20) - punif(13, min = 12, max = 20)
```

```
[1] 0.375
```

Probabilidad de que una inspección ocurra luego de los 17 segundos

$$P(X > 17) = 1 - P(X \leq 17) = 1 - F(17) = 1 - \frac{17 - 12}{8} = 0.375$$

```r
punif(q = 17, min = 12, max = 20, lower.tail = FALSE)
```

```
[1] 0.375
```

Percentil 90 del tiempo de inspección:

$$F(x) = 0.90 \rightarrow \frac{x-12}{8} = 0.9 \rightarrow x = 8 \times 0.9 + 12 = 19.2$$

```r
qunif(p = 0.9, min = 12, max = 20)
```

```
[1] 19.2
```

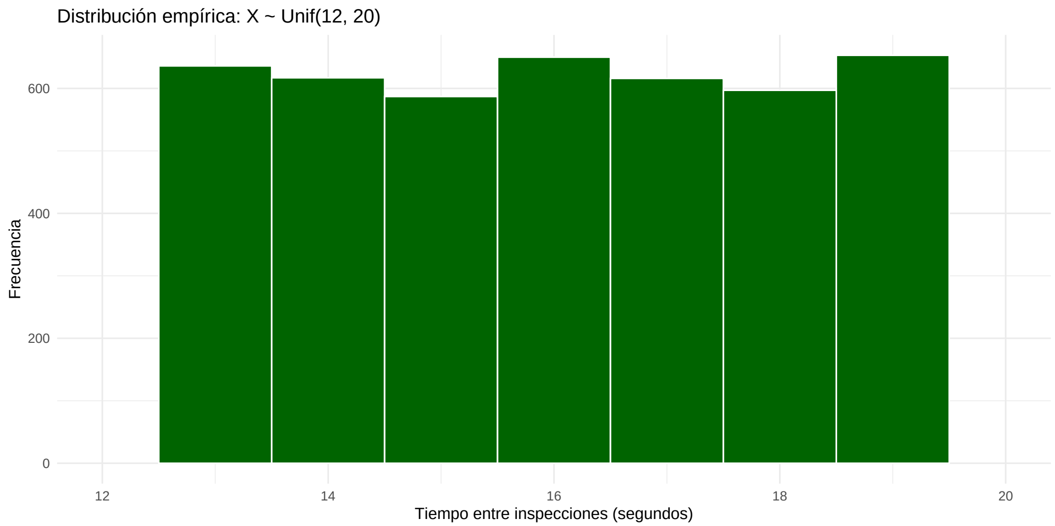
Esto significa que el 90% de las inspecciones ocurre hasta los 19.2 segundos.

Simulación de 7 tiempos de inspección aleatorios:

```r
set.seed(111)
runif(n = 7, min = 12, max = 20)
```

```
[1] 16.74385 17.81185 14.96338 16.11939 15.02131 15.34670 12.08526
```

```r
library(ggplot2)
set.seed(2)
x <- runif(5000, min = 12, max = 20)
ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 1, fill = "darkgreen", color = "white") +
  xlim(12,20)+
  labs(title = "Distribución empírica: X ~ Unif(12, 20)",
       x = "Tiempo entre inspecciones (segundos)",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Unif(12, 20)
Frecuencia
600
400
200
0
12
14
16
18
20
Tiempo entre inspecciones (segundos)

# Distribución triangular

- Distribución continua que modela situaciones donde los valores cercanos a un punto central son más probables, y la probabilidad decrece linealmente hacia los extremos.
- Representa conocimiento parcial: se conoce un mínimo, un máximo y un valor más probable.
- Es una alternativa más informativa que la uniforme, pero más simple que la normal.

## Función de densidad

Una variable aleatoria continua $X$ tiene distribución Triangular en el intervalo $[a, b]$ con centro en $c$ si su función de densidad es:

$$f_X(x) = \frac{2(x-a)}{(b-a)(c-a)} I_{[a,c]}(x) + \frac{2(b-x)}{(b-a)(b-c)} I_{(c,b]}(x)$$

Notación: $X \sim Tri(a, c, b)$

## Ejemplos

- T $=$ Tiempo de espera de un bus, sabiendo que nunca llega antes de 5 minutos ni después de 15, y que usualmente llega alrededor de 10 minutos: $T \sim Tri(5, 10, 15)$
- C $=$ Costo de reparación de una máquina, con mínimo 200 soles, máximo 800 y valor más probable 500: $C \sim Tri(200, 500, 800)$
- D $=$ Duración de una llamada en un call center, entre 2 y 10 minutos, con moda en 4: $D \sim Tri(2, 4, 10)$

## Función de distribución acumulada

$$F_X(x) = \frac{(x-a)^2}{(b-a)(c-a)} I_{[a,c]}(x) + \left(1 - \frac{(b-x)^2}{(b-a)(b-c)}\right) I_{(c,b])}(x) + I_{(b,\infty)}(x)$$

## Media

$$\mu_X = E(X) = \frac{a+b+c}{3}$$

## Varianza

$$\sigma_X^2 = V(X) = \frac{a^2+b^2+c^2-ab-ac-bc}{18}$$

## Función generatriz de momentos

$$M_X(t) = \frac{2(e^{tc}(b-a) - e^{ta}(b-c) - e^{tb}(c-a)}{t^2(b-a)(c-a)(b-c)}$$

## Función característica

$$\phi_X(t) = \frac{2(e^{itc}(b-a) - e^{ita}(b-c) - e^{itb}(c-a)}{(it)^2(b-a)(c-a)(b-c)}$$

## Implementación en R

- Paquete: triangle
- Funciones:
  - dtriangle: densidad en un punto
  - ptriangle: probabilidad acumulada
  - qtriangle: cuantil
  - rtriangle: muestra aleatoria

## Ejemplo 10

En una planta embotelladora, una inspección nunca ocurre antes de 12 s ni después de 20 s, pero el tiempo más habitual es 15 s.

**Variable aleatoria**

V.A. X = Tiempo (en segundos) hasta la próxima inspección, $X \sim Tri(12, 15, 20)$

**Función de densidad**

$$f_X(x) = \frac{2(x-12)}{24} I_{[12,15]}(x) + \frac{2(20-x)}{40} I_{(15,20]}(x)$$

**Función de distribución acumulada**

$$F_X(x) = \frac{(x-12)^2}{24} I_{[12,15]}(x) + \left(1 - \frac{(20-x)^2}{40}\right) I_{(15,20])}(x) + I_{(20,\infty)}(x)$$

**Media**

$$\mu_X = E(X) = \frac{12 + 15 + 20}{3} = 15.67$$

**Varianza**

$$\sigma_X^2 = V(X) = \frac{12^2 + 20^2 + 15^2 - 12 \times 20 - 12 \times 15 - 15 \times 20}{18} = \frac{769 - 720}{18} = 2.722$$

## Función generatriz de momentos

$$M_X(t) = \frac{2(8e^{15t} - 5e^{12t} - 3e^{20t})}{120t^2}$$

## Función característica

$$\phi_X(t) = \frac{2(8e^{15it} - 5e^{12it} - 3e^{20it})}{120(it)^2}$$

Probabilidad de inspección entre 14 y 17 segundos

$$P(14 \leq X \leq 17) = P(X \leq 17) - P(X \leq 14) = 0.775 - 0.167 = 0.608$$

```
library(triangle)
ptriangle(q = 17, a=12, b=20, c=15) -
ptriangle(q = 14, a=12, b=20, c=15)
```

```
[1] 0.6083333
```

Percentil 90

```r
qtriangle(0.9, a=12, b=20, c=15)
```

```
[1] 18
```

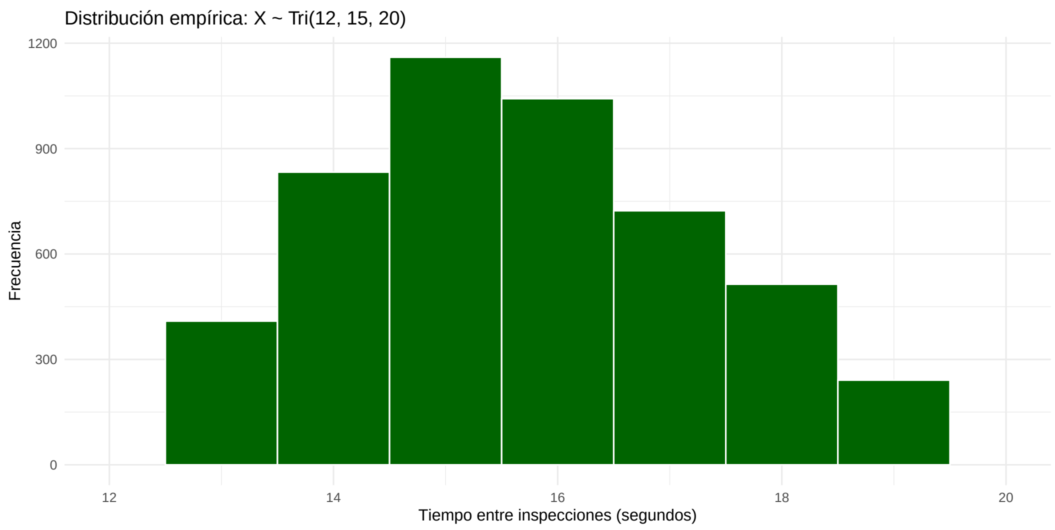
El 90% de las inspecciones se da como máximo en 18 segundos

Simulación de 7 tiempos de inspección

```r
set.seed(111)
rtriangle(7, a=12, b=20, c=15)
```

```
[1] 15.96506 16.69232 14.98163 15.59511 15.01066 15.17646 12.50576
```

```r
library(ggplot2)
set.seed(2026)
x <- rtriangle(5000, a = 12, c = 15, b = 20)
ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 1, fill = "darkgreen", color = "white") +
  xlim(12,20)+
  labs(title = "Distribución empírica: X ~ Tri(12, 15, 20)",
       x = "Tiempo entre inspecciones (segundos)",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Tri(12, 15, 20)
1200
900
600
300
0
Frecuencia
12
14
16
18
20
Tiempo entre inspecciones (segundos)

# Distribución Normal

- Distribución continua ampliamente utilizada en estadística y ciencia de datos, conocida como la campana de Gauss.
- Su importancia se debe al Teorema Central del Límite, que establece que la suma (o media) de muchas variables independientes tiende a seguir una distribución normal, aun si las variables originales no lo sean.

## Función de densidad

Se define la V.A. X con distribución Normal con parámetros $\mu$ y $\sigma^2$, cuya función de densidad viene dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \qquad x \in (-\infty, \infty)$$

Notación: $X \sim N(\mu, \sigma^2)$

$f(x)$ es simétrica respecto a $\mu$, unimodal, y sus colas son infinitas.

## Función de distribución acumulada

No posee una expresión cerrada, se denota como:

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

donde $\Phi$ es la FDA de la distribución Normal estándar.

## Ejemplos

- P $=$ Presión arterial sistólica en adultos sanos, en mm Hg
  $P \sim N(\mu = 120, \sigma^2 = 225)$
- T $=$ Tiempo de resolución de un examen, en minutos $T \sim N(\mu = 90, \sigma^2 = 100)$
- E $=$ Estatura de un estudiante de pregrado, en centímetros
  $E \sim N(\mu = 168, \sigma^2 = 90.25)$

## Media

$$\mu_x = E(X) = \mu = Me$$

## Varianza

$$\sigma_X^2 = V(X) = \sigma^2 > 0$$

## Asimetría

$$As = 0$$

## Curtosis

$$\beta_2 = 3 \Rightarrow \gamma_2 = \beta_2 - 3 = 0$$

## Función generatriz de momentos

$$M_X(t) = exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

## Función generatriz

$$\phi_X(t) = exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right)$$

## Estandarización

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

## Implementación en R

- Paquete: stats
- Funciones:
  - dnorm: densidad en un punto
  - pnorm: probabilidad acumulada
  - qnorm: cuantil
  - rnorm: muestra aleatoria

## Ejemplo 10

La duración de la batería de cierto modelo de laptop sigue una distribución normal con media de 6 horas y desviación estándar de 1.2 horas.

**Variable aleatoria**

X = Duración de la batería, en horas.

$$X \sim N(6, 1.44)$$

**Función de densidad**

$$f(x) = \frac{1}{1.2\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-6}{1.2}\right)^2\right) \qquad x \in (-\infty, \infty)$$

Probabilidad de que una batería dure menos de 7 horas

$$P(X < 7) = 0.7977$$

```
pnorm(q = 7, mean = 6, sd = 1.2)
```

```
[1] 0.7976716
```

Probabilidad de que una batería dure más de 3 horas y 15 minutos

$$P(X > 3.25) = 1 - P(X < 3.25) = 0.9813$$

```
pnorm(q = 3.5, mean = 6, sd = 1.2, lower.tail = F)
```

```
[1] 0.9813896
```

```
1-pnorm(q = 3.5, mean = 6, sd = 1.2)
```

```
[1] 0.9813896
```

Probabilidad de que la duración esté entre 7.5 y 9 horas:

$$P(7.5 < X < 9) = 0.099$$

```r
pnorm(9, mean = 6, sd = 1.2) - pnorm(7.5, mean = 6, sd = 1.2)
```

```
[1] 0.09944011
```

Percentil 27 de la duración de batería

```r
qnorm(p = 0.27, mean = 6, sd = 1.2)
```

```
[1] 5.264624
```

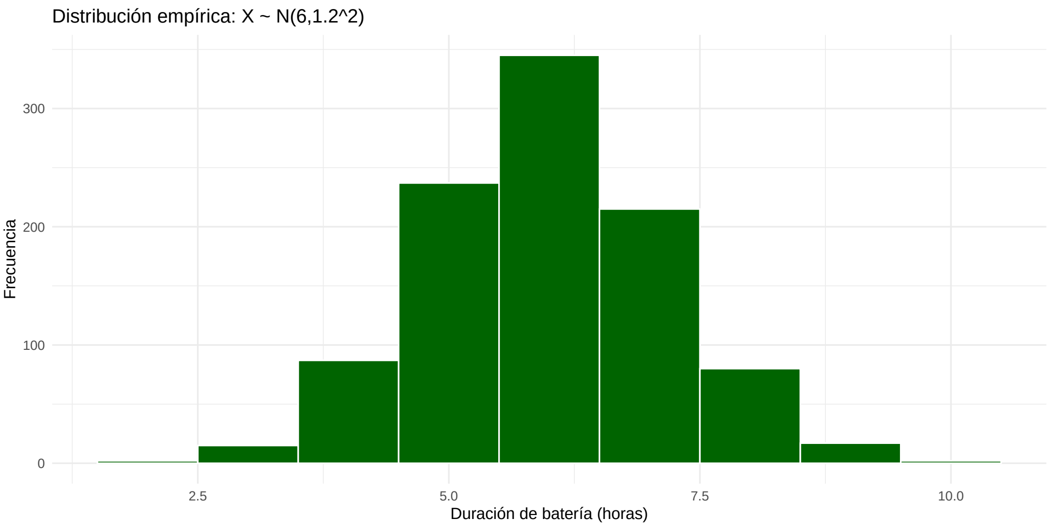
Esto significa que el 27% de las baterías dura como máximo 5.26 horas.

Simulación de 6 tiempos de duración de batería

```r
set.seed(555)
rnorm(n = 6, mean = 6, sd = 1.2)
```

```
[1] 5.604168 6.604376 6.449243 8.266344 3.864117 7.062757
```

```r
library(ggplot2)
set.seed(555)
x <- rnorm(n = 1000, mean = 6, sd = 1.2)
ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 1, fill = "darkgreen", color = "white") +
  labs(title = "Distribución empírica: X ~ N(6,1.2^2)",
       x = "Duración de batería (horas)",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ N(6,1.2^2)
Frecuencia
300
200
100
0
2.5
5.0
7.5
10.0
Duración de batería (horas)

## Aproximaciones a la Normal

**Aproximación de la distribución Binomial a la Normal**

Si $X \sim Bin(n, \pi)$, entonces $X \approx N(\mu = n\pi, \sigma^2 = n\pi(1 - \pi))$ siempre que $n\pi \geq 5$ y $n(1 - \pi) \geq 5$. Sin embargo se debe usar la corrección por continuidad.

$$P(X = a) = P(a - 0.5 \leq X \leq a + 0.5)$$

$$P(X \leq a) = P(X \leq a + 0.5)$$

Por ejemplo, si $X \sim Bin(80, 0.3)$ entonces $X \approx N(\mu = 24, \sigma^2 = 16.8)$

$$P(X = 20) = 0.0626 \approx 0.0604$$

```r
dbinom(x = 20, size = 80, prob = 0.3)
```

```
[1] 0.06262327
```

```r
pnorm(q=20.5, mean=24, sd=sqrt(16.8)) - pnorm(q=19.5, mean=24, sd=sqrt(16.8))
```

```
[1] 0.06044993
```

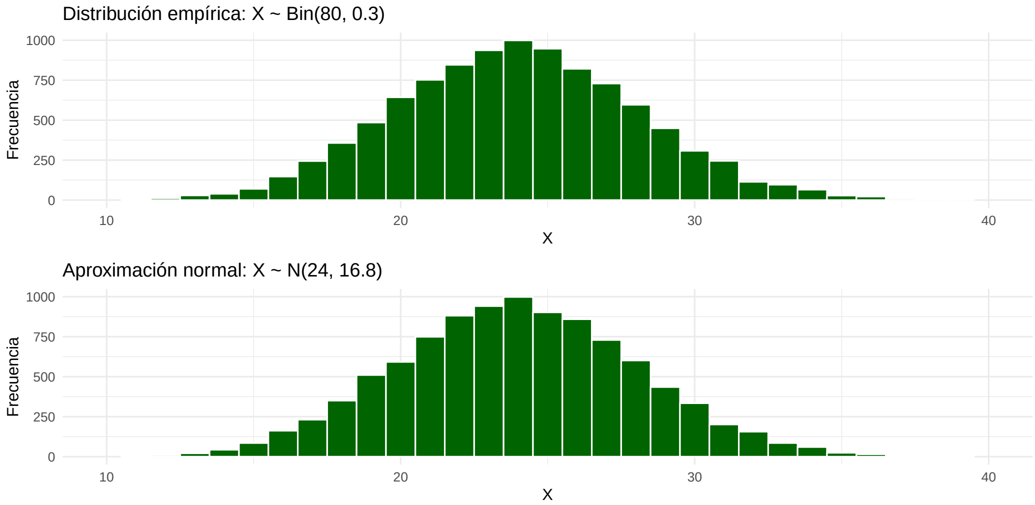
$$P(X \leq 15) = 0.016 \approx 0.019$$

```r
pbinom(q = 15, size = 80, prob = 0.3)
```

```
[1] 0.01606465
```

```r
pnorm(q = 15.5, mean = 24, sd = sqrt(16.8))
```

```
[1] 0.01904952
```



Distribución empírica: X ~ Bin(80, 0.3)
Frecuencia
1000
750
500
250
0
10
20
30
40
X
Aproximación normal: X ~ N(24, 16.8)
Frecuencia
1000
750
500
250
0
10
20
30
40
X

## Aproximación de la distribución Hipergeométrica a la Normal

Si $X \sim Hiper(N, n, A)$, entonces $X \approx N\left(\mu = n\frac{A}{N}, \sigma^2 = n\frac{A}{N}\left(1 - \frac{A}{N}\right)\left(\frac{N-n}{N-1}\right)\right)$ siempre que $n\frac{A}{N} \geq 5$. Sin embargo se debe usar la corrección por continuidad.

$$P(X = a) = P(a - 0.5 \leq X \leq a + 0.5)$$

$$P(X \leq a) = P(X \leq a + 0.5)$$

Por ejemplo, si $X \sim Hiper(N = 250, n = 60, A = 30)$ entonces $X \approx N(\mu = 7.2, \sigma^2 = 4.835)$. Luego, $P(X = 10) = 0.0775 \approx 0.0811$.

```r
dhyper(x = 10, m = 30, n = 220, k = 60)
```

```
[1] 0.07745748
```

```r
pnorm(q=10.5, mean=7.2, sd=sqrt(4.835)) - pnorm(q=9.5, mean=7.2, sd=sqrt(4.835))
```

```
[1] 0.08107484
```

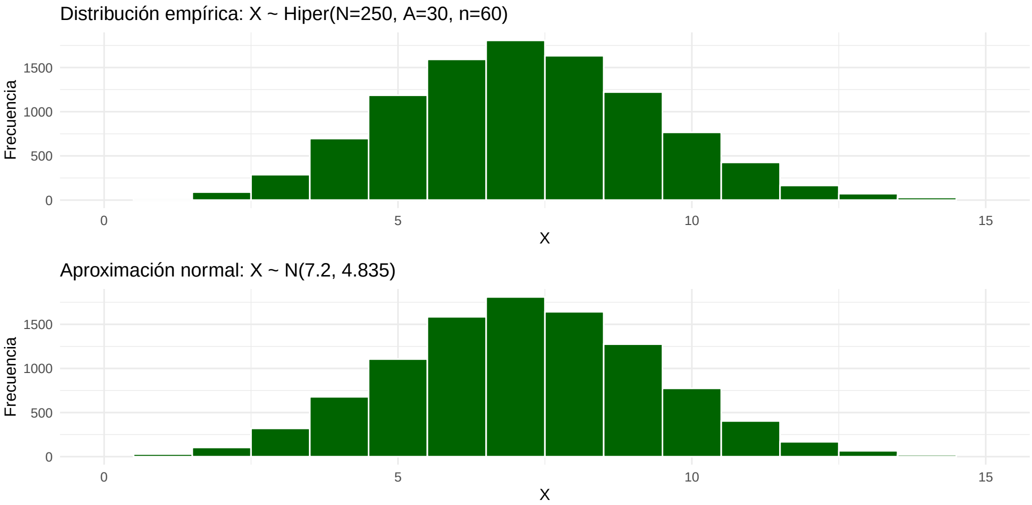
$$P(X \leq 10) = 0.9299 \approx 0.9333$$

```r
phyper(q = 10, m = 30, n = 220, k = 60)
```

```
[1] 0.9299182
```

```r
pnorm(q = 10.5, mean=7.2, sd=sqrt(4.835))
```

```
[1] 0.9332932
```


Distribución empírica: X ~ Hiper(N=250, A=30, n=60)
Frecuencia
0
500
1000
1500
0
5
10
15
X
Aproximación normal: X ~ N(7.2, 4.835)
Frecuencia
0
500
1000
1500
0
5
10
15
X

## Aproximación de la distribución Poisson a la Normal

Si $X \sim Pois(\lambda)$, entonces $X \approx N(\mu = \lambda, \sigma^2 = \lambda)$ siempre que $\lambda \geq 15$. Sin embargo se debe usar la corrección por continuidad.

$$P(X = a) = P(a - 0.5 \leq X \leq a + 0.5)$$

$$P(X \leq a) = P(X \leq a + 0.5)$$

Por ejemplo, si $X \sim Pois(50)$ entonces $X \approx N(\mu = 50, \sigma^2 = 50)$

$$P(X = 55) = 0.0422 \approx 0.0439$$

```
dpois(x = 55, lambda = 50)
```

```
[1] 0.04216435
```

```
pnorm(q=55.5, mean=50, sd=sqrt(50)) - pnorm(q=54.5, mean=50, sd=sqrt(50))
```

```
[1] 0.04392082
```

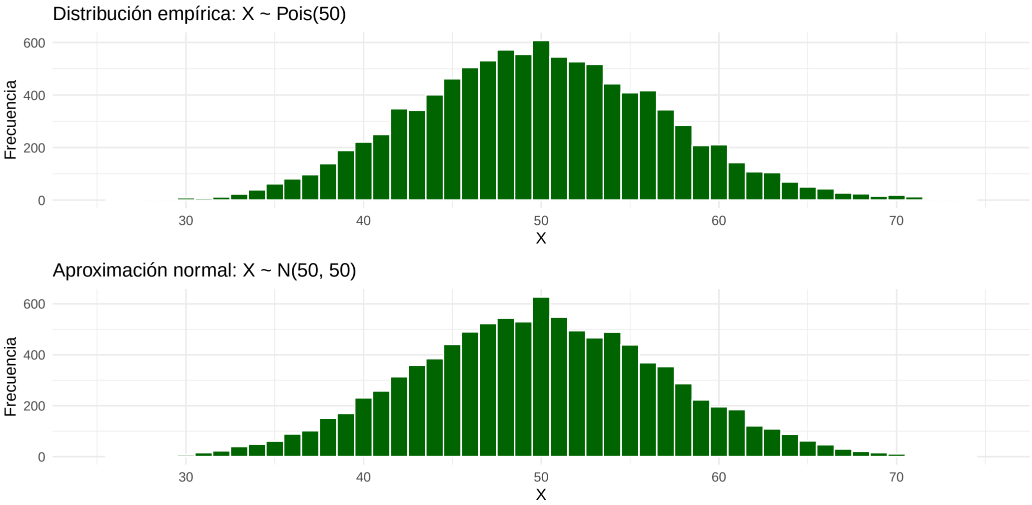
$$P(X \leq 60) = 0.9278 \approx 0.9312$$

```
ppois(q = 60, lambda = 50)
```

```
[1] 0.9278398
```

```
pnorm(q = 60.5, mean = 50, sd = sqrt(50))
```

```
[1] 0.9312181
```


Distribución empírica: X ~ Pois(50)
Frecuencia
600
400
200
0
30
40
50
60
70
X
Aproximación normal: X ~ N(50, 50)
Frecuencia
600
400
200
0
30
40
50
60
70
X

# Distribución lognormal

- Distribución continua utilizada para modelar variables positivas cuyo logaritmo sigue una distribución normal.
- Aparece naturalmente en fenómenos de crecimiento multiplicativo, tiempos de vida, ingresos económicos y tamaños de partículas
- Si $Y = log(X)$ es normal, entonces $X$ es lognormal.

## Función de densidad

Se define la V.A. X con distribución lognormal con parámetros $\mu$ y $\sigma^2$, cuya función de densidad viene dada por:

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right] I_{(0,\infty)}(x)$$

Notación: $X \sim LogN(\mu, \sigma^2)$
La función de densidad $f_X(x)$ es **asimétrica a la derecha**, **unimodal** y presenta **cola larga hacia valores grandes**; su soporte es $(0, \infty)$.

## Función de distribución acumulada

$$F_X(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) I_{(0,\infty)}(x)$$

donde $\Phi$ es la función de distribución acumulada de la Normal estándar.

## Media

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

## Mediana

$$Me(X) = \exp\left(\mu\right)$$

## Varianza

$$Var(X) = \exp(2\mu + \sigma^2)\left[\exp(\sigma^2) - 1\right]$$

## Asimetría

$$As = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}$$

## Implementación en R

- Paquete: stats
- Funciones:
  - dlnorm: densidad en un punto
  - plnorm: probabilidad acumulada
  - qlnorm: cuantil
  - rlnorm: muestra aleatoria

## Ejemplo 11

En una empresa de mensajería, el tiempo de entrega $X$ (en horas) es siempre positivo y presenta cola larga a la derecha: la mayoría de entregas ocurren relativamente rápido, pero ocasionalmente hay demoras importantes. Se modela con una lognormal con
$\mu = 1.5$ y $\sigma = 0.4$

**Variable aleatoria** X = Tiempo de entrega (en horas)

$$X \sim LogN(\mu = 1.5, \sigma = 0.4)$$

**Función de densidad**

$$f_X(x) = \frac{1}{0.4x\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - 1.5}{0.4}\right)^2\right] I_{(0,\infty)}(x)$$

**Media**

$$\mu_x = E(X) = \exp\left(1.5 + \frac{0.4^2}{2}\right) = 4.855$$

**Mediana**

$$Me(X) = \exp(1.5) = 4.482$$

**Varianza**

$$Var(X) = \exp\left(3 + 0.4^2\right)[\exp\left(0.4^2\right) - 1] = 4.09$$

**Asimetría**

$$As = (e^{0.4^2} + 2)\sqrt{e^{0.4^2} - 1} = 1.322$$

Probabilidad de entregar en menos de 6 horas

$$P(X < 6) = F_X(6) = \Phi\left(\frac{log(6) - 1.5}{0.4}\right) = \Phi(0.729) = 0.767$$

```r
plnorm(6, meanlog = 1.5, sdlog = 0.4)
```

```
[1] 0.7671211
```

```r
pnorm((log(6)-1.5)/0.4)
```

```
[1] 0.7671211
```

Probabilidad de que la entrega tome entre 4 y 8 horas

$$P(4 \leq X \leq 8) = F_X(8) - F_X(4) = 0.538$$

```r
plnorm(8, meanlog = 1.5, sdlog = 0.4) -
plnorm(4, meanlog = 1.5, sdlog = 0.4)
```

```
[1] 0.5381719
```

Percentil 39 del tiempo de entrega

```
qlnorm(0.39, meanlog = 1.5, sdlog = 0.4)
```

```
[1] 4.00792
```

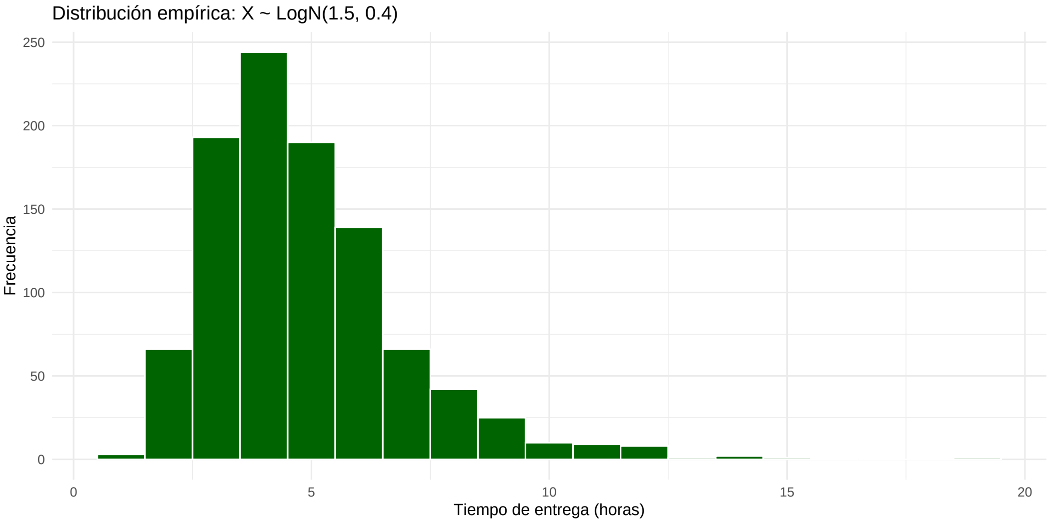
El 39% de los pedidos se entrega en un máximo de 4 horas

Simulación de 10 tiempos de entrega

```r
set.seed(2027)
rlnorm(10, meanlog = 1.5, sdlog = 0.4)
```

```
 [1] 3.083449 3.395513 3.599597 6.089149 9.401226 6.059634 2.903765 3.4840
 [9] 2.378374 3.320399
```

```r
library(ggplot2)
set.seed(2027)
x <- rlnorm(n = 1000, meanlog = 1.5, sdlog = 0.4)
ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 1, fill = "darkgreen", color = "white") +
  labs(title = "Distribución empírica: X ~ LogN(1.5, 0.4)",
       x = "Tiempo de entrega (horas)",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ LogN(1.5, 0.4)
Frecuencia
250
200
150
100
50
0
0
5
10
15
20
Tiempo de entrega (horas)

# Distribución exponencial

- Distribución continua que modela el tiempo entre dos eventos sucesivos en un proceso de Poisson.
- Se utiliza para describir tiempos de espera hasta que ocurra un evento, como la llegada de clientes, fallas de un sistema, o llamadas telefónicas.
- Es una distribución sin memoria: la probabilidad de que ocurra un evento en el futuro no depende de cuánto tiempo ha pasado.

## Función de densidad

Se define la V.A.C. X como el tiempo entre eventos, con parámetro $\lambda > 0$, cuya función de densidad viene dada por:

$$f_X(x) = \lambda \exp(-\lambda x) I_{(0,\infty)}(x)$$

## Función de distribución acumulada

$$F_X(x) = P(X \leq x) = 1 - \exp(-\lambda x) I_{(0,\infty)}(x)$$

## Ejemplos

- T = Tiempo hasta que llegue la próxima llamada a una central telefónica con $\lambda = 3 \rightarrow T \sim Exp(\lambda = 3)$
- C = Tiempo entre llegadas de clientes en una tienda si se estima 1 cliente cada 4 minutos $\rightarrow T \sim Exp\left(\lambda = \frac{1}{4}\right)$
- S = Tiempo de espera para cruzar una calle si el semáforo cambia al azar y en promedio lo hace cada 45 segundos. $\rightarrow S \sim Exp\left(\frac{1}{45}\right)$

## Media

$$\mu_X = E(X) = \frac{1}{\lambda}$$

## Varianza

$$\sigma_X^2 = V(X) = \frac{1}{\lambda^2}$$

## Propiedad de falta de memoria

$$P(X > s + t | X > s) = P(X > t)$$

Es decir, el tiempo restante no depende de cuánto ya se ha esperado.

# Implementación en R

- Paquete: stats
- Funciones:
  - dexp: densidad en un punto
  - pexp: probabilidad acumulada
  - qexp: cuantil
  - rexp: muestra aleatoria

## Ejemplo 12

Un cajero automático recibe, en promedio, un cliente cada 6 minutos. Se desea modelar el tiempo que pasa entre la salida de un cliente y la llegada del siguiente.

**Variable aleatoria**

X = Tiempo entre llegadas de clientes (en minutos), $X \sim Exp\left(\lambda = \frac{1}{6}\right)$

**Función de densidad**

$$f(x) = \frac{1}{6}\exp\left(-\frac{x}{6}\right) I_{(0,\infty)}(x)$$

**Función de distribución acumulada**

$$F(x) = \left(1 - \exp\left(-\frac{x}{6}\right)\right) I_{(0,\infty)}(x)$$

El tiempo medio entre cliente y cliente es:

$$\mu_X = E(X) = \frac{1}{1/6} = 6$$

Su varianza es:

$$\sigma_X^2 = V(X) = \frac{1}{(1/6)^2} = 36$$

La probabilidad de que el próximo cliente llegue antes de los 4 minutos:

$$P(X < 4) = F(4) = 1 - \exp\left(-\frac{4}{6}\right) = 0.48658$$

```r
pexp(q = 4, rate = 1/6)
```

```
[1] 0.4865829
```

La probabilidad de que el próximo cliente llegue luego de los 9 minutos:

$$P(X > 9) = 1 - F(9) = \exp\left(-\frac{9}{6}\right) = 0.2231302$$

```r
pexp(q = 9, rate = 1/6, lower.tail = FALSE)
```

```
[1] 0.2231302
```

```r
1 - pexp(q = 9, rate = 1/6)
```

```
[1] 0.2231302
```

La probabilidad de que el próximo cliente llegue entre los 3 y 7 minutos siguientes

$$P(3 < X < 7) = F(7) - F(3) = \left(1 - \exp\left(-\frac{7}{6}\right)\right) - \left(1 - \exp\left(-\frac{3}{6}\right)\right) = 0.2951274$$

```r
pexp(q = 7, rate = 1/6) - pexp(q = 3, rate = 1/6)
```

```
[1] 0.2951274
```

El percentil 33 del tiempo entre llegadas

```r
qexp(p = 0.33, rate = 1/6)
```

```
[1] 2.402865
```

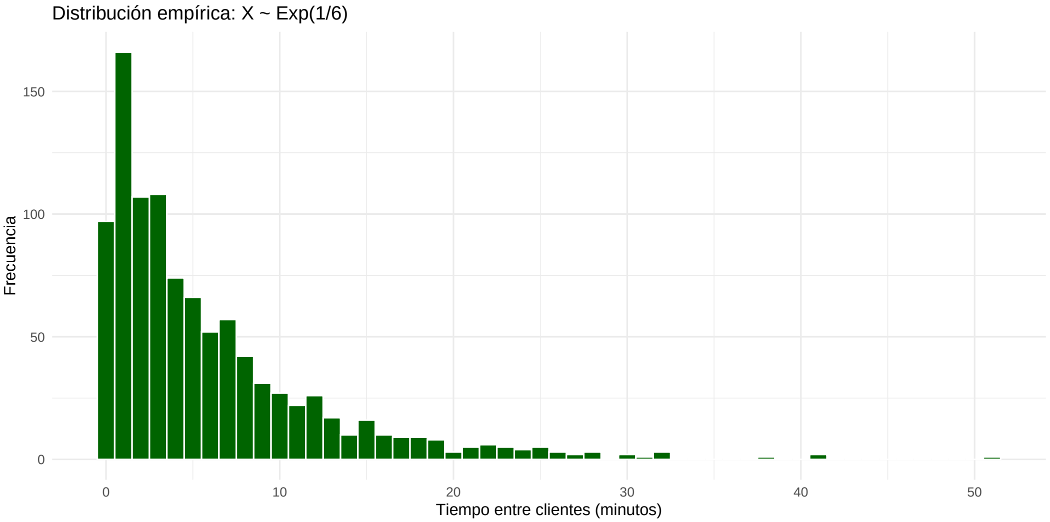
Significa que el 33% de los tiempos entre clientes que llegan al cajero es como máximo de 2.40 minutos (2 minutos y 24 segundos).

Una muestra aleatoria de 5 tiempos de llegada

```r
set.seed(753)
rexp(n = 5, rate = 1/6)
```

```
[1]  0.04490393  2.09205825  5.62120365 10.65256807  7.35046952
```

```r
library(ggplot2)
set.seed(753)
x <- rexp(n = 1000, rate = 1/6)
ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 1, fill = "darkgreen", color = "white") +
  labs(title = "Distribución empírica: X ~ Exp(1/6)",
       x = "Tiempo entre clientes (minutos)",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Exp(1/6)
Frecuencia
150
100
50
0
0
10
20
30
40
50
Tiempo entre clientes (minutos)

# Distribución Gamma

- Distribución continua que generaliza la exponencial, modelando el tiempo hasta que ocurren $\alpha$ eventos en un proceso de Poisson.
- Ampliamente utilizada en teoría de colas, fiabilidad, meteorología, biomedicina, y modelamiento de tiempos de espera acumulados.
- Si $\alpha = 1$, la distribución Gamma se reduce a la Exponencial

## Función de densidad

Se define la V.A.C. X con distribución Gamma con parámetros $\alpha > 0$ y $\beta > 0$, cuya función de densidad viene dada por:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) I_{[0,\infty)}(x)$$

## Ejemplos

- T: Tiempo que transcurre hasta que llegan 3 estudiantes a una clase, si cada llegada en promedio cada 2 minutos $\rightarrow T \sim Gamma(\alpha = 3, \beta = 2)$
- F: tiempo total hasta 5 fallas en un sistema de producción, con tiempo medio de 1.5 días entre fallas $\rightarrow F \sim Gamma(\alpha = 5, \beta = 1.5)$

## Media

$$\mu_X = E(X) = \alpha\beta$$

## Varianza

$$\sigma_X^2 = V(X) = \alpha\beta^2$$

## Asimetría

Asimétrica a la derecha (menos conforme $\alpha$ crece)

# Implementación en R

- Paquete: stats
- Funciones:
  - dgamma: densidad en un punto
  - pgamma: probabilidad acumulada
  - qgamma: cuantil
  - rgamma: muestra aleatoria

## Ejemplo 13

En una central de monitoreo de emergencias, se sabe que, en promedio, ocurre un evento cada 10 minutos. Queremos modelar el tiempo hasta que ocurran 3 eventos consecutivos.

**Variable aleatoria** X =Tiempo que transcurre hasta 3 eventos

$$X \sim Gamma(\alpha = 3, \beta = 10)$$

**Función de densidad**

$$f(x) = \frac{1}{\Gamma(3)10^3} x^2 \exp\left(-\frac{x}{10}\right) I_{[0,\infty)}(x)$$

El tiempo medio que transcurre hasta 3 eventos es:

$$\mu_X = E(X) = 3 \times 10 = 30$$

Su varianza:

$$\sigma_X^2 = V(X) = 3 \times 10^2 = 300$$

Probabilidad de que ocurran los 3 eventos antes de 25 minutos:

$$P(X < 25) = 0.4562$$

```r
pgamma(q = 25, shape = 3, scale = 10)
```

```
[1] 0.4561869
```

Percentil 11 del tiempo que transcurre hasta que sucedan 3 eventos

```r
qgamma(p = 0.11, shape = 3, scale = 10)
```

```
[1] 11.50693
```

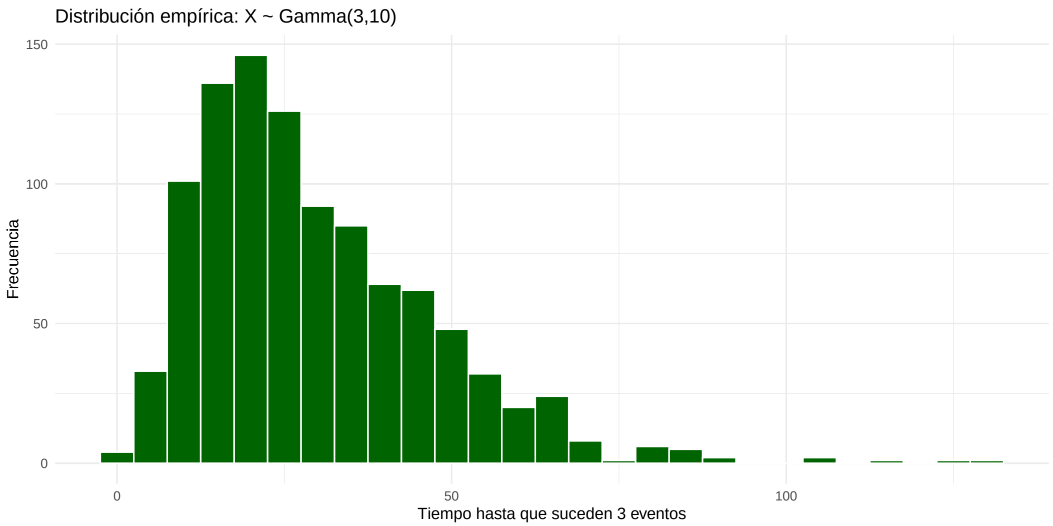
El 11% de los tiempos que transcurren hasta que sucedan 3 eventos es menor o igual a 11.51 minutos.

Muestra aleatoria de 7 tiempos que transcurren hasta que sucedan 3 eventos:

```r
set.seed(852)
rgamma(n = 7, shape = 3, scale = 10)
```

```
[1] 25.997993 57.500235 22.171333 25.254647 16.790934  9.310327 47.792846
```

```r
library(ggplot2)
set.seed(852)
x <- rgamma(n = 1000, shape = 3, scale = 10)
ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 5, fill = "darkgreen", color = "white") +
  labs(title = "Distribución empírica: X ~ Gamma(3,10)",
       x = "Tiempo hasta que suceden 3 eventos",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Gamma(3,10)
Frecuencia
150
100
50
0
0
50
100
Tiempo hasta que suceden 3 eventos

# Distribución Weibull

- Distribución continua muy utilizada para modelar tiempos hasta la ocurrencia de un evento, especialmente tiempos de falla.
- Es fundamental en análisis de confiabilidad, ingeniería, biomedicina, meteorología y supervivencia.
- Generaliza a la exponencial ya que si $\alpha = 1$, se reduce a la distribución Exponencial.

## Función de densidad

Se define la V.A.C. $X$ con distribución Weibull con parámetros $\alpha > 0$ (forma) y $\beta > 0$ (escala), cuya función de densidad es:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right] I_{[0,\infty)}(x)$$

Notación: $X \sim Weibull(\alpha, \beta)$

Si $\alpha < 1$, la tasa de falla es decreciente, mientras que si $\alpha > 1$, es creciente. Por otro lado, $\beta$ controla la magnitud del tiempo (estira o comprime la distribución).

Tasa de fallo creciente: A mayor tiempo de vida, mayor probabilidad de falla. “El sistema envejece, se desgasta, se deteriora”.

Tasa de fallo constante: La probabilidad de falla no depende del tiempo transcurrido. Las fallas son aleatorias

Tasa de fallo decreciente: A mayor tiempo de vida, menor probabilidad de falla. “Los débiles fallan primero, surgen fallas tempranas”

## Ejemplos

- T $=$ tiempo hasta la falla de un componente electrónico, con tasa de falla creciente, $T \sim Weibull(\alpha = 2, \beta = 500)$
- L $=$ tiempo de vida útil de una batería industrial $L \sim Weibull(\alpha = 1.5, \beta = 1000)$

## Función de distribución acumulada

$$F(x) = P(X \leq x) = 1 - \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right] I_{(0,\infty)}(x)$$

## Función de supervivencia

$$S(x) = P(X > x) = \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right] I_{(0,\infty)}(x)$$

## Función de riesgo

$$h(x) = \frac{f(x)}{S(x)} = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} I_{(0,\infty)}(x)$$

## Media

$$\mu_X = E(X) = \beta\Gamma\left(1 + \frac{1}{\alpha}\right)$$

## Mediana

$$Me(X) = \beta\,(\log 2)^{1/\alpha}$$

## Varianza

$$\sigma_X^2 = V(X) = \beta^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \left(\Gamma\left(1 + \frac{1}{\alpha}\right)\right)^2\right]$$

## Asimetría

Generalmente asimétrica a la derecha, pero disminuye a medida que $\alpha$ aumenta

# Implementación en R

- Paquete: stats
- Funciones:
  - dweibull: densidad en un punto
  - pweibull: probabilidad acumulada
  - qweibull: cuantil
  - rweibull: muestra aleatoria

## Ejemplo 14

En una planta industrial, el tiempo hasta que ocurre una falla en una máquina sigue una Weibull con parámetros $\alpha = 2$ y $\beta = 1000$ horas.

**Variable aleatoria** X = Tiempo hasta la falla

$$X \sim Weibull(\alpha = 2, \beta = 1000)$$

**Función de densidad**

$$f(x) = \frac{2}{1000}\left(\frac{x}{1000}\right)^{2-1} \exp\left[-\left(\frac{x}{1000}\right)^{2}\right] I_{[0,\infty)}(x)$$

El tiempo medio hasta la falla es:

$$E(X) = 1000\Gamma\left(1 + \frac{1}{2}\right) = 886.23$$

```
1000*gamma(1.5)
```

```
[1] 886.2269
```

El tiempo mediano hasta la falla es:

$$Me(X) = 1000(\log 2)^{1/2} = 832.5546$$

Su varianza:

$$\sigma_X^2 = V(X) = 1000^2\left[\Gamma\left(1+\frac{2}{2}\right) - \left(\Gamma\left(1+\frac{1}{2}\right)\right)^2\right] = 214601.8$$

```
1000^2*(gamma(1+2/2) - (gamma(1+1/2))^2)
```

```
[1] 214601.8
```

$$\sigma_X = \sqrt{214601.8} = 463.2513$$

Función de distribución acumulada:

$$F(x) = P(X \leq x) = 1 - \exp\left[-\left(\frac{x}{1000}\right)^2\right] I_{(0,\infty)}(x)$$

Función de supervivencia

$$S(x) = P(X > x) = \exp\left[-\left(\frac{x}{1000}\right)^2\right] I_{(0,\infty)}(x)$$

Función de riesgo

$$h(x) = \frac{f(x)}{S(x)} = \frac{2}{1000}\left(\frac{x}{1000}\right)^{2-1} I_{(0,\infty)}(x) = \frac{2x}{10^8}$$

Probabilidad de que la máquina falle antes de 800 horas:

$$P(X < 800) = F(800) = 1 - \exp\left[-\left(\frac{800}{1000}\right)^2\right] = 0.4727$$

```r
pweibull(q = 800, shape = 2, scale = 1000)
```

```
[1] 0.4727076
```

Probabilidad de que la máquina falle luego de las 1000 horas:

$$P(X > 1000) = S(1000) = \exp\left[-\left(\frac{1000}{1000}\right)^2\right] = 0.3679$$

```r
pweibull(q = 1000, shape = 2, scale = 1000, lower.tail = F)
```

```
[1] 0.3678794
```

```r
1- pweibull(q = 1000, shape = 2, scale = 1000)
```

```
[1] 0.3678794
```

Percentil 20 del tiempo hasta la falla:

```r
qweibull(p = 0.20, shape = 2, scale = 1000)
```

```
[1] 472.3807
```

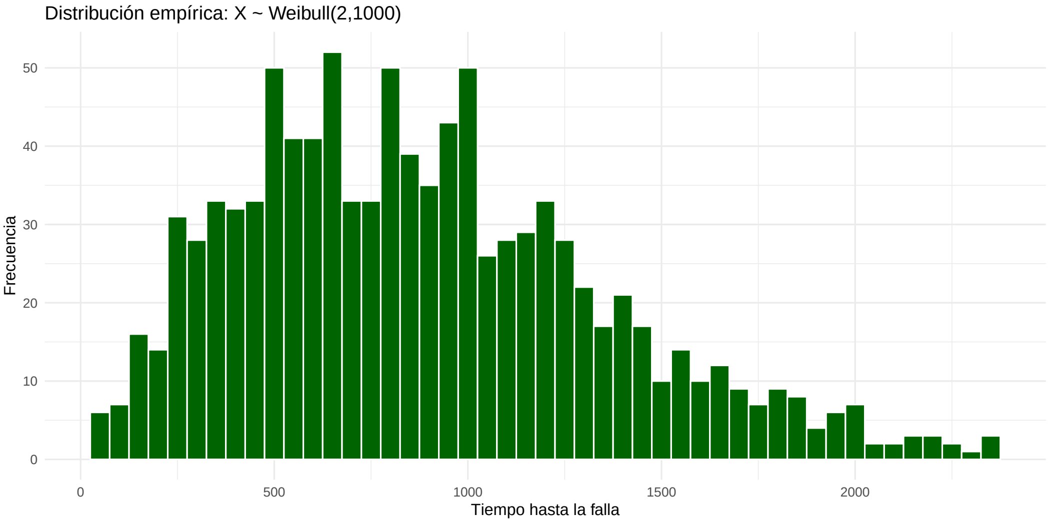
El 20% de los tiempos de vida es menor o igual a las 472.38 horas.

Muestra de 7 tiempos hasta la falla:

```r
set.seed(852)
rweibull(n = 7, shape = 2, scale = 1000)
```

```
[1]  802.8158  920.3126  229.1968 2194.4111  922.6066  852.0811 1544.6043
```

```r
library(ggplot2)
set.seed(852)
x <- rweibull(n = 1000, shape = 2, scale = 1000)

ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 50, fill = "darkgreen", color = "white") +
  labs(title = "Distribución empírica: X ~ Weibull(2,1000)",
       x = "Tiempo hasta la falla",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Weibull(2,1000)
Frecuencia
50
40
30
20
10
0
0
500
1000
1500
2000
Tiempo hasta la falla

# Distribución beta

- Distribución continua definida en el intervalo $(0, 1)$, utilizada para modelar proporciones, probabilidades o porcentajes.
- Es fundamental en estadística bayesiana, modelamiento de tasas, aprendizaje automático, biomedicina y ciencias sociales.
- Es fundamental en estadística bayesiana, modelamiento de tasas, aprendizaje automático, biomedicina y ciencias sociales.

## Función de densidad

Se define la V.A.C. $X$ con distribución Beta con parámetros $\alpha > 0$ y $\beta > 0$ , cuya función de densidad es:

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{(0,1)}(x)$$

donde:

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

- $\alpha$ controla el comportamiento cerca de 0.
- $\beta$ controla el comportamiento cerca de 1.
- Si $\alpha < 1$ y $\beta < 1$, la distribución tiene forma de U.
- Si $\alpha = \beta = 1$, se reduce a la distribución $U(0, 1)$.

## Ejemplos

- P $=$ proporción de estudiantes que aprueban un examen. $P \sim Beta(\alpha = 5, \beta = 2)$
- Q $=$ proporción de tiempo que un sistema permanece operativo.
  $Q \sim Beta(\alpha = 2, \beta = 3)$

## Media

$$\mu_X = E(X) = \frac{\alpha}{\alpha + \beta}$$

## Mediana

$$Me(X) \approx \frac{\alpha - \frac{1}{3}}{\alpha + \beta - \frac{2}{3}}$$

## Moda

Para $\alpha > 1$ y $\beta > 1$

$$Mo(X) = \frac{\alpha - 1}{\alpha + \beta - 2}$$

## Varianza

$$\sigma_X^2 = V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

## Asimetría

- Si $\alpha = \beta$, la distribución es simétrica.
- Si $\alpha > \beta$, la distribución es asimétrica a la izquierda.
- Si $\alpha < \beta$, la distribución es asimétrica a la derecha.

# Implementación en R

- Paquete: stats
- Funciones:
  - dbeta: densidad en un punto
  - pbeta: probabilidad acumulada
  - qbeta: cuantil
  - rbeta: muestra aleatoria

## Ejemplo 15

En un curso universitario, se modela la proporción de preguntas respondidas correctamente por un estudiante en un examen.

**Variable aleatoria** X = proporción de respuestas correctas

$$X \sim Beta(\alpha = 4, \beta = 6)$$

**Función de densidad**

$$f(x) = \frac{1}{B(4,6)} x^{4-1}(1-x)^{6-1} I_{(0,1)}(x)$$

La proporción media de preguntas respondidas correctamente es:

$$E(X) = \frac{4}{4+6} = 0.4$$

La proporción mediana de preguntas respondidas correctamente es aproximadamente:

$$Me(X) \approx \frac{4-\frac{1}{3}}{4+6-\frac{2}{3}} = 0.3929$$

La moda de la proporción de preguntas respondidas correctamente es:

$$Mo(X) = \frac{4-1}{4+6-2} = 0.375$$

Probabilidad de que el estudiante acierte menos del 30% de las preguntas:

$$P(X < 0.3) = 0.2703$$

```r
pbeta(q = 0.3, shape1 = 4, shape2 = 6)
```

```
[1] 0.2703409
```

Percentil 25

```r
qbeta(p = 0.25, shape1 = 4, shape2 = 6)
```

```
[1] 0.2909859
```

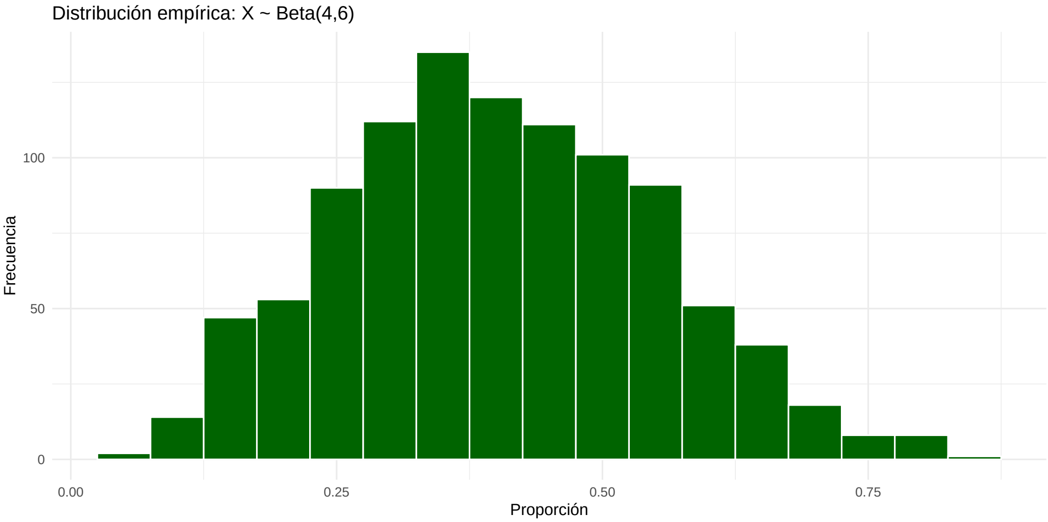
El 25% de las proporciones observadas es menor o igual a 0.291.

Muestra aleatoria de tamaño 7

```r
set.seed(1)
rbeta(n = 7, shape1 = 4, shape2 = 6)
```

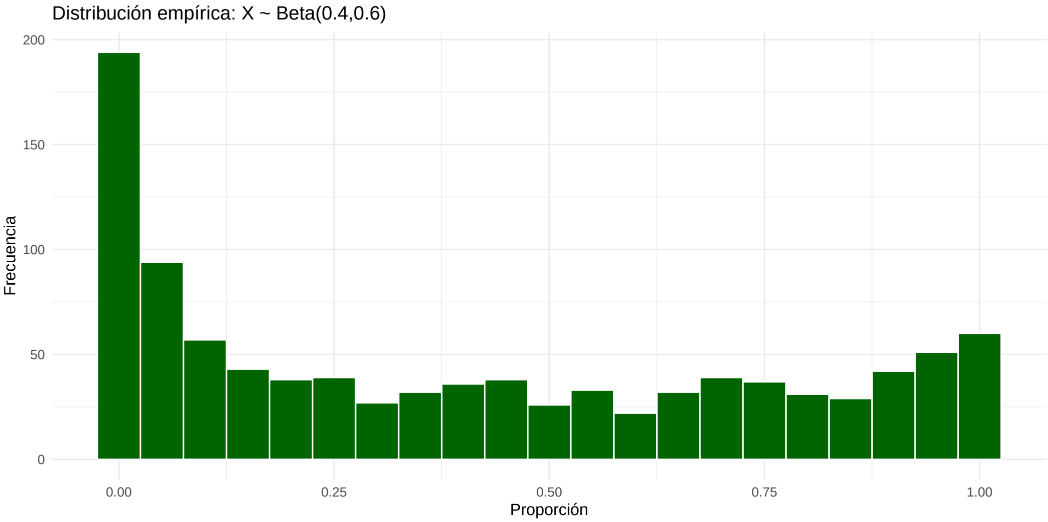
```
[1] 0.2947676 0.4327115 0.2617810 0.4593374 0.2641333 0.4888232 0.5370657
```

```r
library(ggplot2)
set.seed(852)
x <- rbeta(n = 1000, shape1 = 4, shape2 = 6)

ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 0.05, fill = "darkgreen", color = "white") +
  labs(title = "Distribución empírica: X ~ Beta(4,6)",
       x = "Proporción",
       y = "Frecuencia") +
  theme_minimal()
```



Distribución empírica: X ~ Beta(4,6)
Frecuencia
100
50
0
0.00
0.25
0.50
0.75
Proporción

```r
library(ggplot2)
set.seed(1)
x <- rbeta(n = 1000, shape1 = 0.4, shape2 = 0.6)

ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 0.05, fill = "darkgreen", color = "white") +
  labs(title = "Distribución empírica: X ~ Beta(0.4,0.6)",
       x = "Proporción",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Beta(0.4,0.6)
Frecuencia
200
150
100
50
0
0.00
0.25
0.50
0.75
1.00
Proporción

# Distribución Cauchy

- Distribución continua definida en todo $\mathbb{R}$, utilizada para modelar fenómenos con colas pesadas y valores extremos frecuentes.
- Aparece en física, procesamiento de señales, finanzas, y como ejemplo clásico de distribución sin media ni varianza.
- Es un caso particular de la t de Student con 1 grado de libertad.
- A diferencia de la Normal, no cumple el Teorema Central del Límite.

## Función de densidad

Se define la V.A.C. $X$ con distribución Cauchy con parámetros $x_0 \in \mathbb{R}$ y $\gamma > 0$ (escala), cuya función de densidad es:

$$f(x) = \frac{1}{\pi\gamma} \frac{1}{1 + \left(\frac{x - x_0}{\gamma}\right)^2} I_{(-\infty,\infty)}(x)$$

Notación: $X \sim \mathrm{Cauchy}(x_0, \gamma)$

## Función de distribución acumulada

$$F(x) = P(X \leq x) = \frac{1}{\pi} arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}$$

## Ejemplos

- T: error de medición con valores extremos ocasionales, $T \sim Cauchy(x_0 = 0, \gamma = 1)$
- R: desviación respecto a una frecuencia central en una señal, $R \sim Cauchy(x_0 = 10, \gamma = 2)$

## Media

No existe

## Mediana

$$Me(X) = x_0$$

## Moda

$$Mo(X) = x_0$$

## Varianza

No existe

# Implementación en R

- Paquete: stats
- Funciones:
  - dcauchy: densidad en un punto
  - pcauchy: probabilidad acumulada
  - qcauchy: cuantil
  - rcauchy: muestra aleatoria

## Ejemplo 16

En un experimento físico, se modela el error de medición de un sensor altamente sensible, donde pueden aparecer errores extremadamente grandes.

**Variable aleatoria** X = error de medición

$$X \sim Cauchy(x_0 = 0, \gamma = 1)$$

**Función de densidad**

$$f(x) = \frac{1}{\pi} \times \frac{1}{1 + x^2} I_{(-\infty,\infty)}(x)$$

**Función de distribución acumulada**

$$F(x) = P(X \leq x) = \frac{1}{\pi} arctan\,(x) + \frac{1}{2}$$

**Mediana**:

$$Me(X) = 0$$

**Moda**:

$$Mo(X) = 0$$

Probabilidad de que el error esté entre -1 y 1.5:

$$P(-1 < X < 1.5) = F(1.5) - F(-1) = 0.5628$$

```
(1/pi*atan(1.5)+1/2) - (1/pi*atan(-1)+1/2)
```

```
[1] 0.562833
```

```
pcauchy(1.5, location = 0, scale = 1) - pcauchy(-1, location = 0, scale =
```

```
[1] 0.562833
```

Percentil 14

```r
qcauchy(0.14, location = 0, scale = 1)
```

```
[1] -2.125108
```

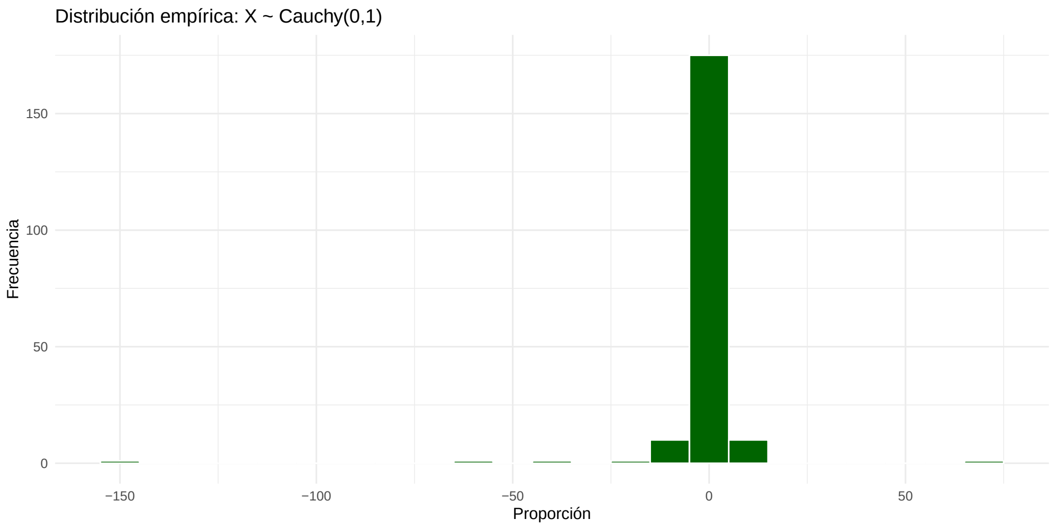
El 14% de los errores de medición son como máximo iguales a -2.125

Muestra aleatoria de tamaño 5

```r
set.seed(1111)
rcauchy(n = 5, location = 0, scale = 1)
```

```
[1]  9.1909260  3.5639343 -0.3007645  0.4595041 -1.0728555
```

```r
library(ggplot2)
set.seed(2026)
x <- rcauchy(n = 200, location = 0, scale = 1)

ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 10, fill = "darkgreen", color = "white") +
  labs(title = "Distribución empírica: X ~ Cauchy(0,1)",
       x = "Proporción",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Cauchy(0,1)
Frecuencia
150
100
50
0
−150
−100
−50
0
50
Proporción

# Distribución Laplace

- Distribución continua definida en todo $\mathbb{R}$, utilizada para modelar datos simétricos con colas más pesadas que la Normal, pero más ligeras que Cauchy.
- Aparece en econometría, procesamiento de señales, robustez estadística y machine learning.
- También se conoce como distribución doble exponencial.
- A diferencia de la distribución Cauchy, sí posee media y varianza finitas.

## Función de densidad

Se define la V.A.C. $X$ con distribución Laplace con parámetros $\mu \in \mathbb{R}$ (localización) y $b > 0$ (escala), cuya función de densidad es:

$$f(x) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right) I_{(-\infty,\infty)}(x)$$

Notación: $X \sim Laplace(\mu, b)$

## Función de distribución acumulada

$$F(x) = P(X \leq x) = \frac{1}{2} \exp\left(\frac{x-\mu}{b}\right) I_{(-\infty,\mu)}(x) + \left(1 - \frac{1}{2} \exp\left(-\frac{x-\mu}{b}\right)\right) I_{[\mu,\infty)}(x)$$

## Ejemplos

- T: error de medición con variaciones bruscas, $T \sim Laplace(\mu = 0, b = 1)$
- R: desviación respecto a un valor central en una señal, $R \sim Laplace(\mu = 10, b = 2)$

## Media

$$\mu_X = E(X) = \mu$$

## Mediana

$$Me_X = \mu$$

## Moda

$$Mo_X = \mu$$

## Varianza

$$\sigma_X^2 = Var(X) = 2b^2$$

## Implementación en R

- Paquete: greybox
- Funciones:
  - dlaplace: densidad en un punto
  - plaplace: probabilidad acumulada
  - qlaplace: cuantil
  - rlaplace: muestra aleatoria

## Ejemplo 17

En un experimento de control de calidad, se modela el error de medición de un sensor, con mayor concentración cerca del valor real pero con posibles desviaciones grandes.
**Variable aleatoria** X = Variable aleatoria

$$X \sim Laplace(\mu = 0, b = 1)$$

**Función de densidad**

$$f(x) = \frac{1}{2}\exp\left(-|x|\right)I_{(-\infty,\infty)}(x)$$

**Función de distribución acumulada**

$$F(x) = P(X \leq x) = \frac{1}{2}\exp\left(x\right)I_{(-\infty,0)}(x) + \left(1 - \frac{1}{2}\exp\left(-x\right)\right)I_{[0,\infty)}(x)$$

**Media**

$$\mu_X = E(X) = 0$$

**Mediana**

$$Me_X = 0$$

**Moda**

$$Mo_X = 0$$

**Varianza**

$$\sigma_X^2 = V(X) = 2(1)^2 = 2$$

Probabilidad de que el error esté entre -2 y 0.3

$$P(-2 < X < 0.3) = F(0.3) - F(2) = \left(1 - \frac{1}{2}\exp\left(-0.3\right)\right) - \left(\frac{1}{2}\exp\left(-2\right)\right) = 0.5619$$

```
library(greybox)
(1-0.5*exp(-0.3) - 0.5*exp(-2))
```

```
[1] 0.5619232
```

```
plaplace(0.3, mu = 0, scale = 1) - plaplace(-2, mu = 0, scale = 1)
```

```
[1] 0.5619232
```

Percentil 52

```r
qlaplace(0.52, mu = 0, scale = 1)
```

```
[1] 0.04082199
```

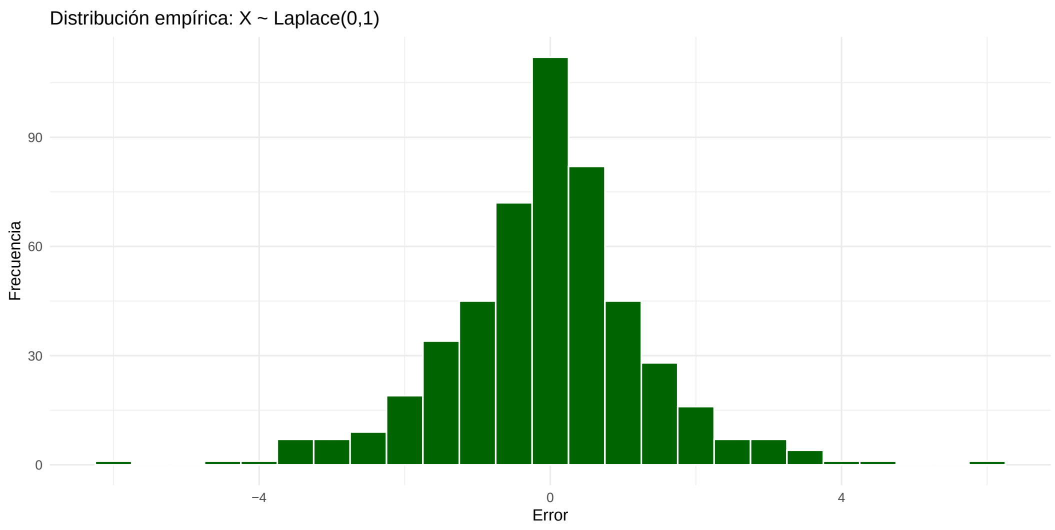
El 52% de los errores de medición del sensor toman el valor máximo de 0.0408

Muestra aleatoria de tamaño 4

```r
set.seed(1111)
rlaplace(n = 4, mu = 0, scale = 1)
```

```
[1] -0.07149033 -0.19134244  1.68204316 -1.29385804
```

```r
set.seed(159)
x <- rlaplace(n = 500, mu = 0, scale = 1)

ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 0.5, fill = "darkgreen", color = "white") +
  labs(title = "Distribución empírica: X ~ Laplace(0,1)",
       x = "Error",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Laplace(0,1)
Frecuencia
90
60
30
0
−4
0
4
Error

# Distribución logística

- Distribución continua definida en todo $\mathbb{R}$, utilizada para modelar datos simétricos con colas más pesadas que la Normal, pero más ligeras que Cauchy.
- Es fundamental en econometría, machine learning y estadística aplicada, especialmente como base del modelo de regresión logística.
- Tiene forma similar a la Normal, pero con colas ligeramente más pesadas.
- A diferencia de Cauchy, sí posee media y varianza finitas.

## Función de densidad

Se define la V.A.C. $X$ con distribución Logística con parámetros $\mu \in \mathbb{R}$ (localización) y $s > 0$ (escala), cuya función de densidad es:

$$f(x) = \frac{\exp(-\frac{x-\mu}{s})}{s\left(1 + \exp(-\frac{x-\mu}{s})\right)^2} I_{(-\infty,\infty)}(x)$$

Notación: $X \sim Logistic(\mu, s)$

## Función de distribución acumulada

$$F(x) = P(X \leq x) = \frac{1}{1 + \exp(-\frac{x-\mu}{s})} I_{(-\infty,\infty)}(x)$$

## Ejemplos

- T: error de medición con colas moderadamente pesadas, $T \sim Logistic(\mu = 0, s = 1)$
- R: desviación respecto a un valor central en una señal, $R \sim Logistic(\mu = 8, s = 1.5)$

## Media

$$\mu_X = E(X) = \mu$$

## Mediana

$$Me_X = \mu$$

## Moda

$$Mo_X = \mu$$

## Varianza

$$\sigma_X^2 = Var(X) = \frac{\pi^2 s^2}{3}$$

## Implementación en R

- Paquete: stats
- Funciones:
  - dlogis: densidad en un punto
  - plogis: probabilidad acumulada
  - qlogis: cuantil
  - rlogis: muestra aleatoria

# Ejemplo 18

En un experimento de análisis de errores, se modela el error de medición de un sensor con comportamiento simétrico y colas moderadas.
**Variable aleatoria** X = error de medición

$$X \sim Logistic(\mu = 0, s = 2)$$

**Función de densidad**

$$f(x) = \frac{\exp(-\frac{x-}{2})}{2\left(1 + \exp(-\frac{x}{2})\right)^2} I_{(-\infty,\infty)}(x)$$

**Función de distribución acumulada**

$$F(x) = P(X \leq x) = \frac{1}{1 + \exp(-\frac{x}{2})} I_{(-\infty,\infty)}(x)$$

**Media**

$$\mu_X = E(X) = 0$$

**Mediana**

$$Me_X = 0$$

**Moda**

$$Mo_X = 0$$

**Varianza**

$$\sigma_X^2 = V(X) = \frac{\pi^2 \times 2^2}{3} = 13.16$$

Probabilidad de que el error esté entre -0.5 y 2.1

$$P(-0.5 < X < 2.1) = \frac{1}{1+\exp\left(-1.05\right)} - \frac{1}{1+\exp\left(0.25\right)} = 0.303$$

```
1/(1+exp(-1.05)) - 1/(1+exp(0.25))
```

```
[1] 0.3029514
```

```
plogis(2.1,location=0,scale=2) - plogis(-0.5,location=0,scale=2)
```

```
[1] 0.3029514
```

Percentil 88

```r
qlogis(p = 0.88, location = 0, scale = 2)
```

```
[1] 3.98486
```

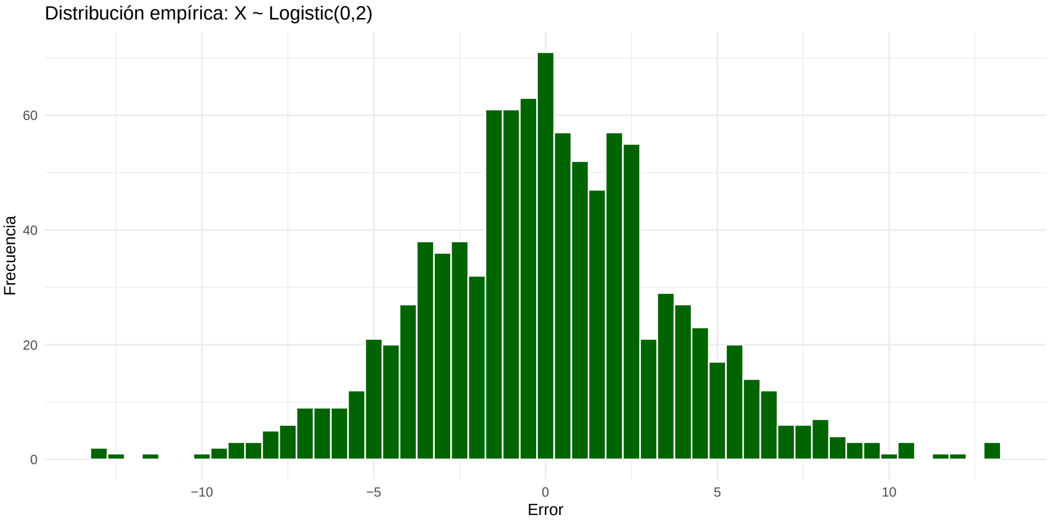
El 88% de los errores toman un valor máximo de 3.985

Muestra aleatoria de tamaño 4

```r
set.seed(2026)
rlogis(n = 4, location = 0, scale = 2)
```

```
[1]  1.6819780  0.4541859 -3.6282560 -1.8324931
```

```r
library(ggplot2)
set.seed(159)
x <- rlogis(n = 1000, location = 0, scale = 2)

ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 0.5, fill = "darkgreen", color = "white") +
  labs(title = "Distribución empírica: X ~ Logistic(0,2)",
       x = "Error",
       y = "Frecuencia") +
  theme_minimal()
```


Distribución empírica: X ~ Logistic(0,2)
Frecuencia
60
40
20
0
−10
−5
0
5
10
Error

# Distribución Pareto

- Distribución continua definida en el intervalo $[x_m, \infty)$, utilizada para modelar fenómenos con colas pesadas y concentración extrema.
- Es fundamental en economía, finanzas, ciencias sociales, ciencia de datos, teoría del riesgo y modelamiento de desigualdad.
- A diferencia de Normal, Logística o Laplace, puede no tener media ni varianza, dependiendo de sus parámetros.

## Función de densidad

Se define la V.A.C. $X$ con distribución Pareto con parámetros $\theta > 0$ y $\eta > 0$, cuya función de densidad es:

$$f(x) = \frac{\theta\eta^{\theta}}{x^{\theta+1}} I_{[\eta,\infty)}(x)$$

Notación: $X \sim Pareto(\theta, \eta)$

- $\eta$ es el valor mínimo de la variable
- $\theta$ controla el peso de la cola (concentración).

## Función de distribución acumulada

$$F(x) = P(X \leq x) = \left[1 - \left(\frac{\eta}{x}\right)^{\theta}\right] I_{[\eta,\infty)}(x)$$

## Ejemplos

- W: ingreso anual de una persona, $W \sim Pareto(\eta = 10000, \theta = 2.5)$
- S: tamaño de una empresa (empleados), $S \sim Pareto(\eta = 5, \theta = 1.8)$

## Media

$$\mu_X = E(X) = \frac{\theta\eta}{\theta - 1}$$

Solo para $\theta > 1$

## Mediana

$$Me_X = 2^{\frac{1}{\theta}} \times \eta$$

## Moda

$$Mo_X = \eta$$

## Varianza

$$\sigma_X^2 = Var(X) = \frac{\theta\eta^2}{(\theta - 1)^2(\theta - 2)}$$

Solo para $\theta > 2$

## Implementación en R

- Paquete: EnvStats
- Funciones:
  - dpareto: densidad en un punto
  - ppareto: probabilidad acumulada
  - qpareto: cuantil
  - rpareto: muestra aleatoria

## Ejemplo 19

En un estudio económico, se modela el ingreso anual de una población, sabiendo que existe un ingreso mínimo de 10 000 unidades monetarias y una fuerte desigualdad.

**Variable aleatoria** X = Ingreso anual

$$X \sim Pareto(\eta = 10000, \theta = 2.5)$$

**Función de densidad**

$$f(x) = \frac{2.5 \times 10000^{2.5}}{x^{3.5}} I_{[10000,\infty)}(x)$$

**Función de distribución acumulada**

$$F(x) = \left[1 - \left(\frac{10000}{x}\right)^{2.5}\right] I_{[10000,\infty)}(x)$$

## Media

$$\mu_X = E(X) = \frac{2.5 \times 10000}{2.5 - 1} = 16666.67$$

## Mediana

$$Me(X) = 2^{\frac{1}{2.5}} \times 10000 = 13195.08$$

## Moda

$$Mo(X) = 10000$$

## Varianza

$$\sigma_X^2 = V(X) = \frac{2.5 \times 10000^2}{1.5^2 \times 0.5} = 222222222.22$$

## Desviación estándar

$$\sigma_X = 14907.12$$

Probabilidad de que el ingreso sea mayor a 30 000:

$$P(X > 30000) = 1 - \left[1 - \left(\frac{10000}{30000}\right)^{2.5}\right] = 0.06415$$

```r
library(EnvStats)
1-ppareto(30000,location=10000,shape=2.5)
```

```
[1] 0.06415003
```

Percentil 40 de los ingresos

```r
qpareto(0.40, location = 10000, shape = 2.5)
```

```
[1] 12267.03
```

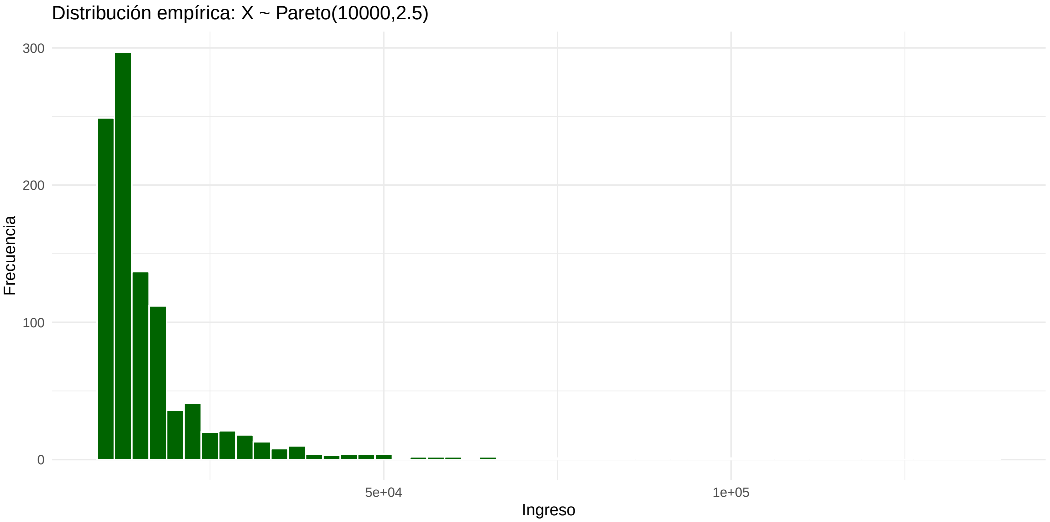
El 40% de los trabajadores gana 12267.03 soles como máximo.

Muestra aleatoria de tamaño 6

```r
set.seed(159)
rpareto(n = 6, location = 10000, shape = 2.5)
```

```
[1] 10516.34 14644.02 14838.06 10481.12 10073.65 10201.88
```

```r
library(ggplot2)
set.seed(159)
x <- rpareto(n = 1000, location = 10000, shape = 2.5)

ggplot(data.frame(x), aes(x)) +
  geom_histogram(binwidth = 2500, fill = "darkgreen", color = "white") +
  labs(title = "Distribución empírica: X ~ Pareto(10000,2.5)",
       x = "Ingreso",
       y = "Frecuencia") +
  theme_minimal()
```



Distribución empírica: X ~ Pareto(10000,2.5)
300
200
100
0
Frecuencia
5e+04
1e+05
Ingreso