Estimación: Estimación por máxima verosimilitud del parámetro pi de la distribución Binomial

$X_1, X_2, \ldots, X_n$ donde $X_i \sim Bin(n=1, \pi)$ $\Rightarrow$ $\pi = ?$

$$f(x_i|\pi) = \cancel{\binom{1}{x_i}}^{1} \pi^{x_i}(1-\pi)^{1-x_i} = \pi^{x_i}(1-\pi)^{1-x_i}$$

f. Verosimilitud = $L(\pi / \mathbf{x}) = \prod_{i=1}^{n} f(x_i|\pi)$

likelihood

creíble

Verosimilitud

$$L(\pi|\mathbf{x}) = \prod_{i=1}^{n} \pi^{x_i}(1-\pi)^{1-x_i} = \underbrace{\pi^{\sum x_i}(1-\pi)^{n-\sum x_i}}_{\text{f. Verosimilitud}}$$

¿cuál es el valor de π más creíble? $\Rightarrow$ el que maximiza L

log-Verosimilitud

$$\log L(\pi|\mathbf{x}) = \sum x_i \log \pi + (n-\sum x_i)\log(1-\pi)$$

$$\frac{\partial \log L(\pi|\mathbf{x})}{\partial \pi} = \frac{\sum x_i}{\pi} + \frac{(n-\sum x_i)}{1-\pi}(-1) = 0$$

$$\frac{\sum x_i}{\pi} = \frac{n-\sum x_i}{1-\pi}$$

$$\sum x_i - \cancel{\pi \sum x_i} = n\pi - \cancel{\pi\sum x_i}$$

$$\bar{x} = \frac{\sum x_i}{n} = \hat{\pi}$$

↳ estimador

```r
(datos |> pull(Seguro) |> mean() -> pX_est)
```

```
[1] 0.575
```

* Prueba de Bondad de Ajuste

Estimación: Estimación por máxima verosimilitud del parámetro lambda de la distribución Exponencial

$f(x_i|\lambda) = \lambda e^{-\lambda x_i}$ para una muestra $x_1, x_2, \ldots, x_n$

$\lambda = ?$

$L(\lambda|\mathbf{x}) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i}$

likelihood

$$\text{f. } \underbrace{\text{Verosimilitud}}_{\text{creíble}} = L(\lambda/\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\lambda)$$

$\log L(\lambda|\mathbf{x}) = n \log \lambda - \lambda \sum x_i$

$\frac{\partial \log L(\lambda|\mathbf{x})}{\partial \lambda} = \frac{n}{\lambda} - \sum x_i = 0$

$\frac{n}{\lambda} = \sum x_i$

$\frac{1}{\bar{x}} = \boxed{\frac{n}{\sum x_i} = \hat{\lambda}}$

↳ Estimador

```r
((datos |> pull(Espera) |> mean())**-1 -> lambdaE_est)
```

```
[1] 2.083333
```

✦ Prueba de Bondad de Ajuste

$$\hat{\alpha} \quad \hat{\beta} \quad \hat{\gamma}$$

$$X \sim D\left(\underbrace{\alpha \mid \beta \mid \gamma}_{\theta}\right)$$

Distribución Normal

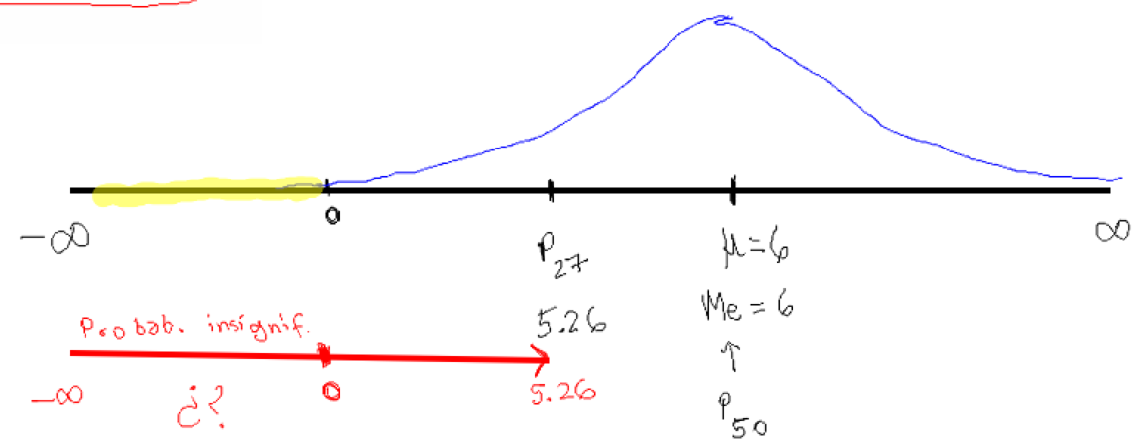
Percentil 27 de la duración de batería

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

```
[1] 5.264624
```

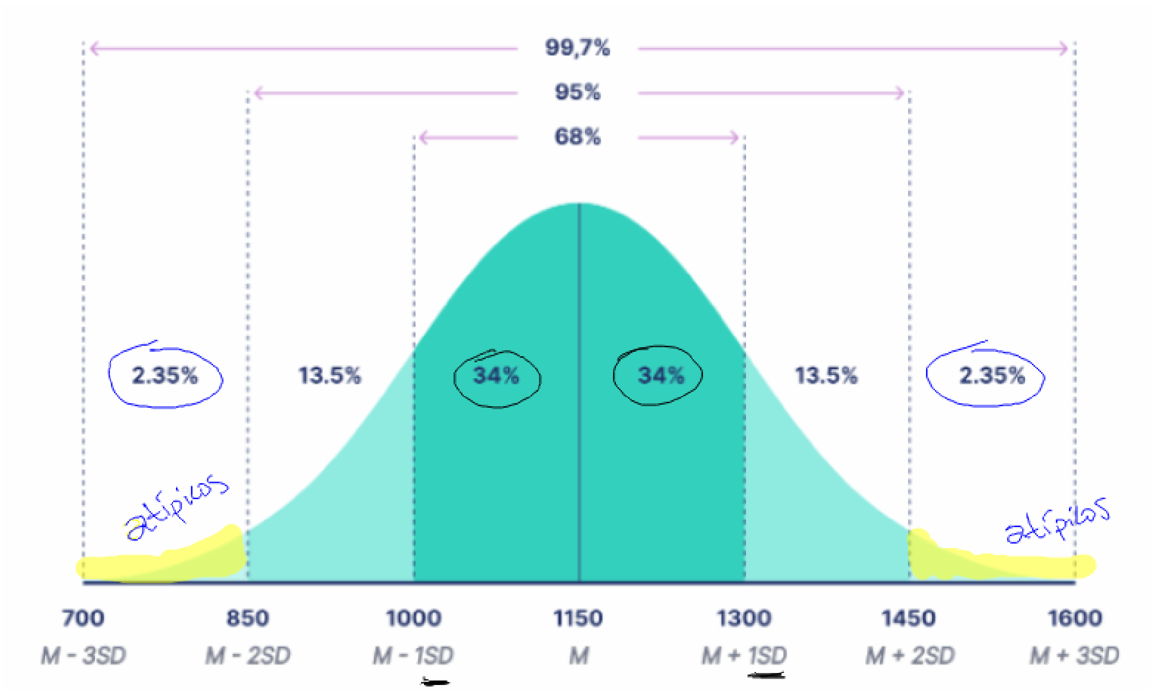
Esto significa que el 27% de las baterías dura como máximo 5.26 horas.

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



$P(X<0) = ¿?$

```
> pnorm(q = 0, mean = 6, sd = 1.2)
[1] 2.866516e-07
```



6	6+1.2	6+2.4	6+3.6
	7.2	8.4	9.6

$X \sim N(\mu = 6, \sigma = 1.2)$

Entre $\underbrace{4.8}_{6-1.2}$ y $\underbrace{7.2}_{6+1.2}$ horas = 68%

Entre 6 y 8.4 horas → 47.5%

Escalar

- $\boxed{\text{Estandarizar}}$ $\frac{X - \bar{X}}{S}$ → aprox entre -4 y 4
- Normalizar: $\frac{X - \min}{\max - \min}$ → entre 0 y 1, inclusive

Aproximaciones a la distribución 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.

$\{0, 1, \ldots, 10\} \leftarrow X \sim Bin(n=10, \pi = 0.3) \rightarrow n\pi = 3, \quad n(1-\pi) = 7$

Si aproximamos a la Normal (en este caso no se debería): $X \sim N(\mu = 3, \sigma^2 = 2.1)$

$P(X<0) = 0.02$

```
> pnorm(q = 0, mean = 3, sd = sqrt(2.1))
[1] 0.01921697
```

$X \sim Binom(n = 40, \pi = 0.4) \rightarrow n\pi = 16, \quad n(1-\pi) = 24$

Aproximamos a la Normal: $X \sim N(\mu = 16, \sigma^2 = 9.6)$

$P(X<0) = 0.000000121 \approx 0$

```
> pnorm(q = 0, mean = 16, sd = sqrt(9.6))
[1] 1.208782e-07
```

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

$$P(X = 8) = 0.0626 \approx 0.0604$$

$P(X=8) \approx P(7.5 < X < 8.5)$

Normal

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

Probabilidad exacta

```
[1] 0.06262327
```

```
pnorm(q=20.5, mean=24, sd=sqrt(16.8)) - pnorm(q=19.5, mean=24, sd=sqrt(16
```

Prob. aprox.

```
[1] 0.06044993
```

$n \uparrow$ Aprox Binom – Normal mejora

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 $n\frac{A}{N} \geq 5$ y $n\left(1-\frac{A}{N}\right) \geq 5$. Sin embargo se debe usar la corrección por continuidad.

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

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

Hiperg $\longrightarrow$ discreta

$\Downarrow$ corrección

Normal $\longrightarrow$ continua

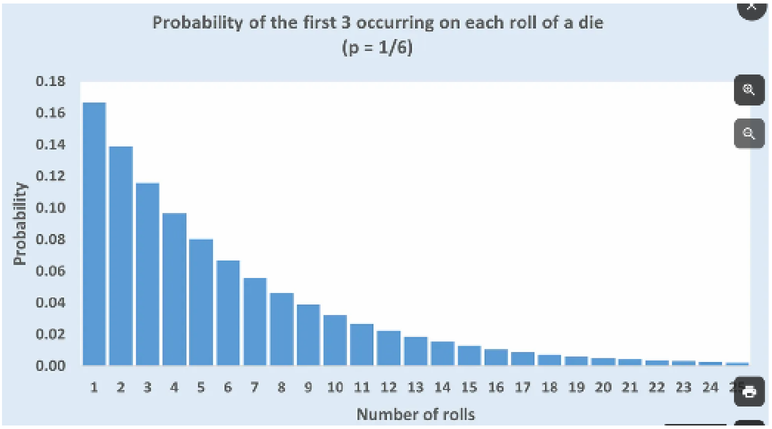
geométrica $\quad f(x) = \pi(1-\pi)^x$

$x=0 : \quad \pi$

$x=1 : \quad \pi(1-\pi)$

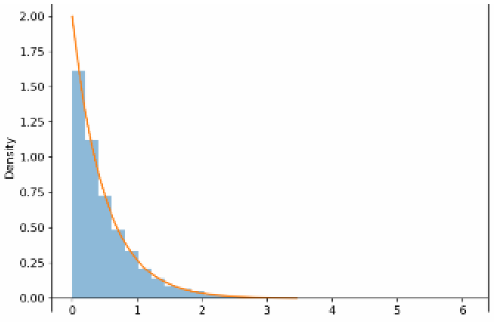
$x=2 : \quad \pi(1-\pi)^2 = \pi(1-\pi)(1-\pi)$

$x=3 : \quad \pi(1-\pi)^3 = \pi(1-\pi)(1-\pi)(1-\pi)$



binomial negativa: sí se aproxima a $N(\mu, \sigma^2)$ pero con $r\pi \geq 5$ y $r(1-\pi) \geq 5$

exponencial: No $\longrightarrow$



gamma: sí para α y β grandes

Distribución muestral de la media

- Caso 1: Si es una población con distribución normal: Entonces para cualquier n, $\overline{X}$ se distribuye como una Normal con media μ y varianza $\frac{\sigma^2}{n}$
- Caso 2. Si la población no proviene de una distribución normal. Entonces, para $n \geq 30$ suficientemente grande y por el El teorema del límite central, $\overline{X}$ se distribuye aproximadamnte como una Normal con media μ y varianza $\frac{\sigma^2}{n}$

1) $X_i \sim N(\mu, \sigma^2) \Rightarrow \overline{X} \sim N(\mu, \sigma^2/n)$

2) $X_i \sim D(\mu, \sigma^2) \Rightarrow \overline{X} \sim N(\mu, \sigma^2/n)$, $n \geq 30$ (TLC)

cualq. otra dist.

3) $X_i \sim D(\mu, \sigma^2)$ y $n < 30 \Rightarrow$ Alternativas no paramétricas, Remuestreo (computacional)

cualq. otra dist

Una clínica registra el tiempo de atención médica de sus pacientes como una variable que sigue una distribución normal con media 18 minutos y desviación estándar 6 minutos. Se toma una muestra aleatoria de n=9 pacientes. ¿Cuál es la probabilidad de que el tiempo promedio de atención sea mayor a 20 minutos?

$X \sim N(\mu = 18, \sigma^2 = 6^2)$, $n = 9$ $\Rightarrow$ $\bar{X} \sim N(\mu = 18, \sigma^2 = \underbrace{36/9}_{4})$ $\sigma = 2$

$P(\bar{X} > 20) = 0.1587$

```
> pnorm(q = 20, mean = 18, sd = 2, lower.tail = F)
[1] 0.1586553
> 1-pnorm(q = 20, mean = 18, sd = 2)
[1] 0.1586553
```

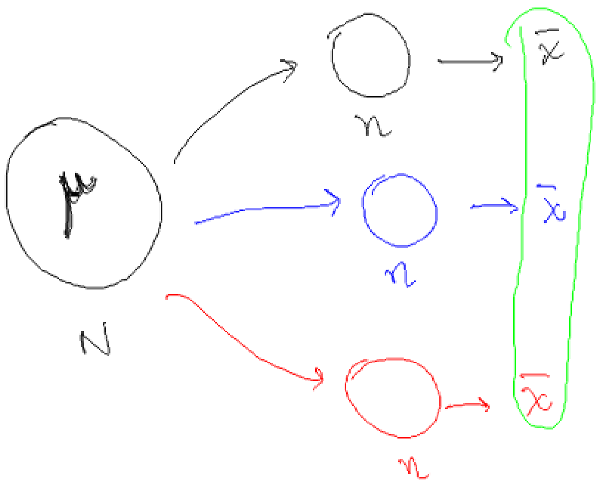
En una ciudad, se estudian los ingresos diarios de trabajadores independientes. Se sabe que los ingresos son altamente asimétricos a la derecha, con algunos trabajadores que ganan mucho más que el promedio. Estudios previos han estimado que la media poblacional es de 120 soles y la desviación estándar es de 90 soles. Un investigador toma una muestra aleatoria de $\boxed{n = 40}$ trabajadores, ¿cuál es la probabilidad de que el ingreso promedio de la muestra esté entre 100 y 140 soles?

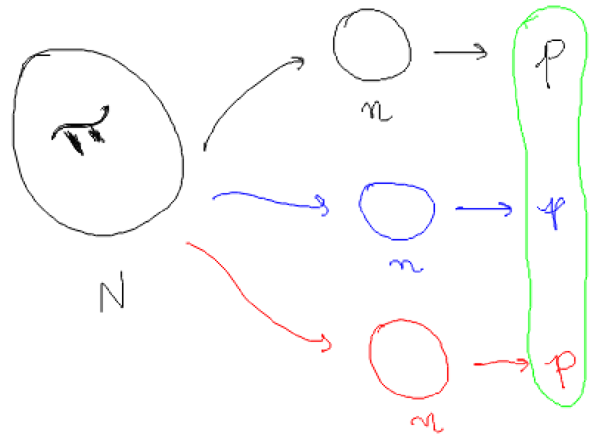
$$X \sim D(\mu = 120,\ \sigma = 90),\quad n = 40 \Rightarrow \bar{X} \sim N(\mu = 120,\ \sigma^2 = \underbrace{90^2/40}_{202.5})$$

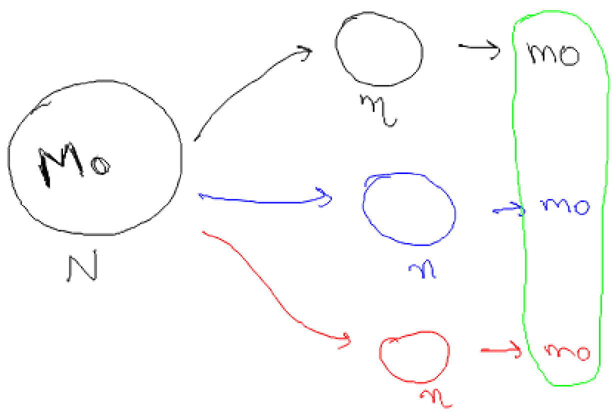
$$P(100 < \bar{X} < 140) = P(\bar{X} < 140) - P(\bar{X} < 100) = 0.84$$

```
> pnorm(q = 140, mean = 120, sd = sqrt(202.5)) -
+   pnorm(q = 100, mean = 120, sd = sqrt(202.5))
[1] 0.8401145
```

$P(100 < X < 140)$ → sí se necesita X normal o aprox. normal.







no existe forma analítica para calcular, p. ej $P(mo < 2)$

$\Downarrow$

métodos computacionales

Una encuesta nacional indica que aproximadamente el 62% de los adultos leen al menos un libro al mes. Un investigador realiza una encuesta a n = 150 adultos seleccionados al azar en una región del país. ¿Cuál es la probabilidad de que menos del 58% de los encuestados en esta muestra lean al menos un libro al mes?

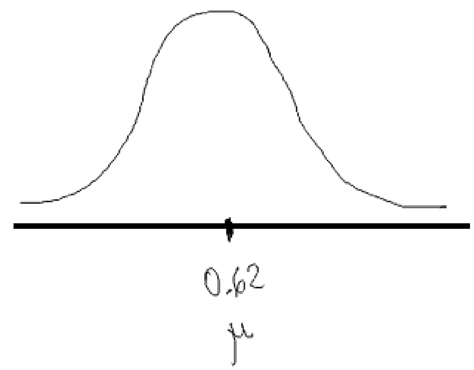
$$\left.\begin{array}{l}\pi = 0.62 \\ n = 150\end{array}\right\} \quad \hat{p} \sim N\left(\mu = 0.62,\ \sigma^2 = \frac{0.62 \times 0.38}{150}\right)$$

$$P(\hat{p} < 0.58) = 0.1564$$

$\downarrow$

pnorm

```
> pnorm(q = 0.58, mean = 0.62, sd = sqrt(0.62*0.38/150))
[1] 0.1564167
```



Probabilidad de que $t_{(10)}$ sea menor que -2

grados de libertad

degrees of freedom

$$P(t_{(10)} < -2) = 0.0367$$

```r
pt(q = -2, df = 10)
```

```
[1] 0.03669402
```

Probabilidad de que $t_{(100)}$ sea menor que -1

$$P(t_{(100)} < -1) = 0.16 \approx P(Z < -1)$$

```r
pt(q = -1, df = 100)
```

```
[1] 0.1598621
```

```r
pnorm(q = -1)
```

$\approx$

```
[1] 0.1586553
```

$gl = 100 \Rightarrow t_{(100)} \sim N(0,1)$