


# Introducción a la Estimación por Máxima Verosimilitud

Sergio M. Nava Muñoz
nava@cimat.mx
CIMAT

2025-06-01



## Table of contents





## Objetivo

- Introducir el concepto de máxima verosimilitud (MLE)
- Aplicar MLE a modelos de tiempo de supervivencia
- Usar R para estimar parámetros en presencia de censura
- Interpretar estimaciones y su relación con funciones de supervivencia



## ¿Qué es la verosimilitud?

- Es una función que mide **cuán probable** es observar los datos dados ciertos parámetros. Este enfoque es introducido en Moore (2016) como base para la estimación paramétrica en supervivencia.
- Dado un modelo con función de densidad $f(t;\theta)$, la **verosimilitud** para un conjunto de datos $t_1, \dots, t_n$ es:

$$L(\theta) = \prod_{i=1}^{n} f(t_i;\theta)$$

- Se busca el valor $\hat{\theta}$ que **maximiza** $L(\theta)$ o, más comúnmente, $\log L(\theta)$



## Caso con censura

- Si hay censura, se observa:
  - Tiempo $t_i$
  - Indicador $\delta_i = 1$ si ocurrió el evento, $0$ si censurado
- La función de verosimilitud se ajusta:

$$L(\theta) = \prod_{i=1}^{n} [f(t_i;\theta)]^{\delta_i} [S(t_i;\theta)]^{1-\delta_i}$$

Este desarrollo puede encontrarse también en Klein & Moeschberger (2003) como parte de la teoría general de modelos paramétricos en supervivencia.



## Ejemplo: distribución exponencial

- Supón $T \sim \text{Exp}(\lambda)$, entonces:
  - $f(t) = \lambda e^{-\lambda t}$
  - $S(t) = e^{-\lambda t}$
- Verosimilitud con censura:

$$L(\lambda) = \prod_{i=1}^{n} [\lambda e^{-\lambda t_i}]^{\delta_i} [e^{-\lambda t_i}]^{1-\delta_i} = \lambda^d e^{-\lambda \sum t_i}$$

- $d = \sum \delta_i$, número de eventos

# Derivación del estimador MLE

Para una discusión general sobre el principio de máxima verosimilitud, véase Casella & Berger (2002)

- Log-verosimilitud:

$$\ell(\lambda) = d \log \lambda - \lambda \sum t_i$$

- Derivando e igualando a 0:

$$\frac{d}{d\lambda}\ell(\lambda) = \frac{d}{\lambda} - \sum t_i = 0 \Rightarrow \hat{\lambda} = \frac{d}{\sum t_i}$$



# Código R: estimación con censura

```r
tt <- c(7, 6, 6, 5, 2, 4)
status <- c(0, 1, 0, 0, 1, 1)

d <- sum(status)
suma_t <- sum(tt)
lambda_hat <- d / suma_t
lambda_hat
```

```
[1] 0.1
```



# Comparación con `survreg`

```r
library(survival)
fit <- survreg(Surv(tt, status) ~ 1, dist = "exponential")
summary(fit)
```

```
Call:
survreg(formula = Surv(tt, status) ~ 1, dist = "exponential")
            Value Std. Error    z       p
(Intercept) 2.303      0.577 3.99 6.7e-05

Scale fixed at 1

Exponential distribution
Loglik(model)= -9.9   Loglik(intercept only)= -9.9
Number of Newton-Raphson Iterations: 4
n= 6
```

- La estimación $\hat{\lambda}$ se relaciona con $\text{scale}^{-1}$



# ¿Dónde está en `survreg()`?

```r
fit <- survreg(Surv(tt, status) ~ 1, dist = "exponential")
summary(fit)
```

- El modelo AFT estima:

$$\log(T) = \mu + \varepsilon, \quad \text{con } \mu = \text{Intercepto}$$

- Para la distribución exponencial:

$$\mu = \log\left(\frac{1}{\lambda}\right) \Rightarrow \hat{\lambda} = e^{-\mu}$$

```r
lambda_survreg <- exp(-fit$coefficients)
lambda_survreg
```

```
(Intercept)
        0.1
```

- En tu salida: `Intercept = 2.303`
  Entonces:
  $$\hat{\lambda} = e^{-2.303} \approx 0.1$$



# Interpretación de resultados

- $\hat{\lambda}$ es la tasa de riesgo constante estimada
- Su inverso es la **media de supervivencia**:

$$\hat{E}(T) = \frac{1}{\hat{\lambda}}$$



# Actividad práctica

1. Simula un conjunto de datos de supervivencia con censura
2. Calcula el estimador de máxima verosimilitud para $\lambda$
3. Usa `survreg` para confirmar

CIMAT

## Conclusiones

- MLE permite incorporar eventos y censura de forma natural
- Las expresiones son simples en modelos paramétricos como el exponencial
- Herramientas de R hacen este proceso accesible



CIMAT

## Lecturas recomendadas

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Duxbury.
Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). Springer.
Moore, D. F. (2016). *Applied survival analysis using r* (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-31245-3