Análisis de Supervivencia

Funciones fundamentales de Análisis de Supervivencia

Sergio M. Nava Muñoz

2025-06-01

1 Funciones fundamentales de Análisis de Supervivencia

1.1 Introducción

En esta sección abordaremos los conceptos fundamentales para el análisis de datos de supervivencia, comenzando con funciones de probabilidad clásicas y avanzando hacia funciones específicas como la función de supervivencia y la función de riesgo.

1.1.1 Objetivos

- Recordar las funciones de densidad y distribución acumulada.
- Introducir la función de supervivencia $S(t)$ y la función de riesgo $h(t)$.
- Interpretar estas funciones desde una perspectiva probabilística.
- Visualizar ejemplos aplicados y comparativos con distintas distribuciones.

1.2 Funciones fundamentales

Antes de introducir las funciones de supervivencia y riesgo, recordemos dos funciones clave en probabilidad y estadística:

- **Función de densidad**: $f(t)$
- **Función de distribución acumulada**: $F(t) = P(T \leq t)$

1.2.1 Función de densidad $f(t)$

- Describe la distribución de probabilidad de una variable continua T
- No es una probabilidad en sí, pero su integral sí lo es:

$$P(a < T \leq b) = \int_a^b f(t)\,dt$$

- Debe cumplir:

$$f(t) \geq 0 \quad \text{y} \quad \int_{-\infty}^{\infty} f(t)\,dt = 1$$

1.2.2 Función de distribución acumulada $F(t)$

- Es la probabilidad de que la variable aleatoria tome un valor menor o igual que t:

$$F(t) = \int_{-\infty}^{t} f(u)\,du = P(T \leq t)$$

- Propiedades:
  - $F(t)$ es monótona creciente
  - $\lim_{t\to-\infty} F(t) = 0$
  - $\lim_{t\to\infty} F(t) = 1$

1.2.3 Relación entre $f(t)$ y $F(t)$

- Si f es continua:

$$f(t) = \frac{d}{dt}F(t)$$

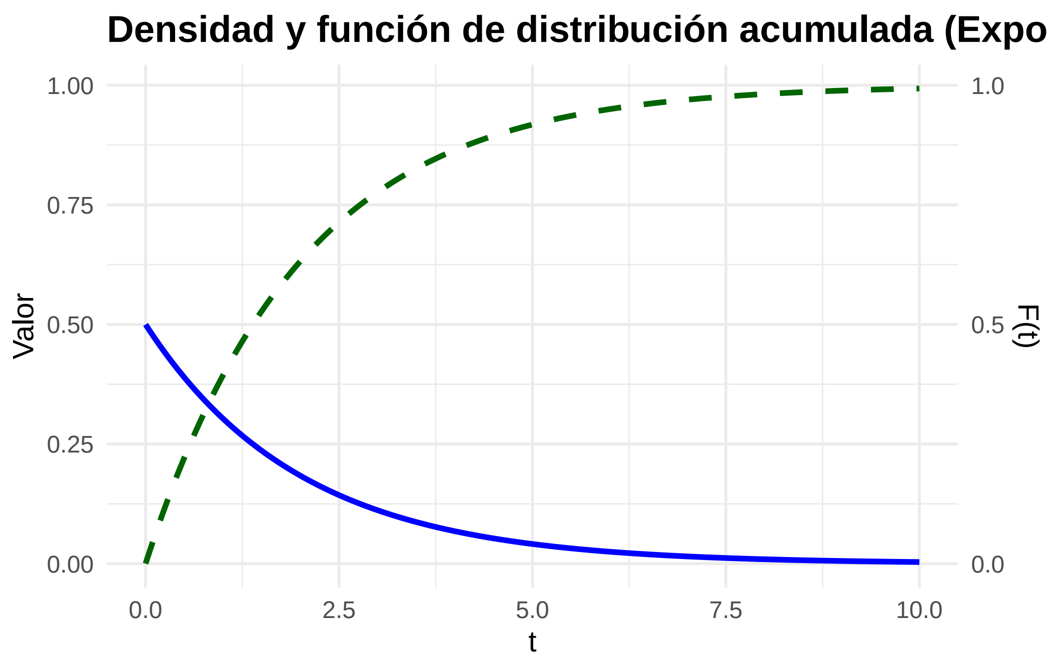
- Y también:

$$F(t) = \int_{-\infty}^{t} f(u)\,du$$

Estas relaciones son clave para definir funciones como la de supervivencia y la de riesgo, que veremos a continuación.

1.2.4 Ejemplo en R: distribución exponencial con parámetro $\lambda = 0.5$

```
library(ggplot2)
library(dplyr)
t <- seq(0, 10, length.out = 400)
lambda <- 0.5
datos <- data.frame(
  t = t,
  densidad = dexp(t, rate = lambda),
  acumulada = pexp(t, rate = lambda)
)
ggplot(datos, aes(x = t)) +
  geom_line(aes(y = densidad), color = "blue", size = 1) +
  geom_line(aes(y = acumulada), color = "darkgreen", size = 1, linetype = "dashed") +
  labs(
    title = "Densidad y función de distribución acumulada (Exponencial)",
    y = "Valor",
    x = "t"
  ) +
  theme_minimal() +
  theme(plot.title = element_text(size = 14, face = "bold")) +
  scale_y_continuous(sec.axis = sec_axis(~., name = "F(t)", breaks = c(0, 0.5, 1)))
```



1.3 Funciones fundamentales en análisis de supervivencia

En análisis de supervivencia, las variables aleatorias de interés T son no negativas, y se caracterizan no solo por $f(t)$ o $F(t)$, sino también por funciones **más interpretables**:

- $S(t)$: función de supervivencia
- $h(t)$: función de riesgo o tasa de falla
- $H(t)$: riesgo acumulado

2 Función de supervivencia $S(t)$

2.1 Función de Supervivencia

La función de supervivencia $S(t)$ y la función de riesgo instantáneo $h(t)$ son fundamentales para modelar procesos de falla en este tipo de análisis, ver Klein & Moeschberger (2003).

$$S(t) = P(T > t) = 1 - F(t)$$

Representa la probabilidad de **sobrevivir más allá del tiempo** t.

Propiedades clave:

- Monótona no creciente
- $S(0) = 1$, $\lim_{t\to\infty} S(t) = 0$

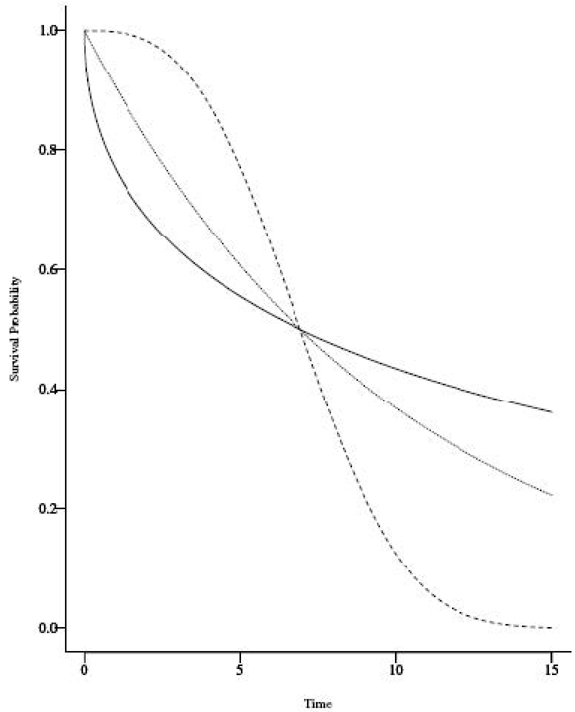


Figure 2.1 *Weibull Survival functions for* $\alpha = 0.5$, $\lambda = 0.26328$ (———); $\alpha = 1.0$, $\lambda = 0.1$ (······); $\alpha = 3.0$, $\lambda = 0.00208$ (------).

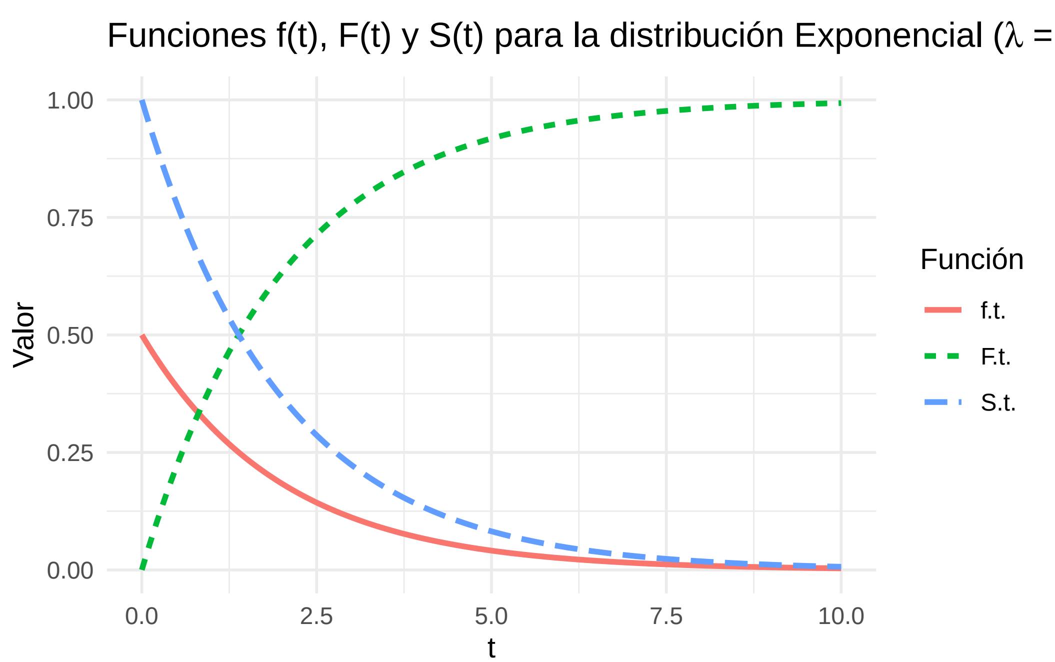
2.2 Ejemplo: función de supervivencia para distribución exponencial

Sea $T \sim \mathrm{Exp}(\lambda = 0.5)$, es decir:

$$f(t) = \lambda e^{-\lambda t}, \quad F(t) = 1 - e^{-\lambda t}, \quad S(t) = e^{-\lambda t}$$

```r
library(ggplot2)
library(tidyr)
t <- seq(0, 10, length.out = 200)
lambda <- 0.5
```

```r
datos <- data.frame(
  t = t,
  `f(t)` = dexp(t, rate = lambda),
  `F(t)` = pexp(t, rate = lambda),
  `S(t)` = 1 - pexp(t, rate = lambda)
)
datos_long <- pivot_longer(datos, cols = -t, names_to = "funcion", values_to = "valor")
ggplot(datos_long, aes(x = t, y = valor, color = funcion, linetype = funcion)) +
  geom_line(size = 1) +
  labs(
    title = expression(paste("Funciones f(t), F(t) y S(t) para la distribución Exponencial ("
    x = "t",
    y = "Valor",
    color = "Función",
    linetype = "Función"
  ) +
  theme_minimal()
```



3 Función de riesgo $h(t)$

3.1 Función de Riesgo

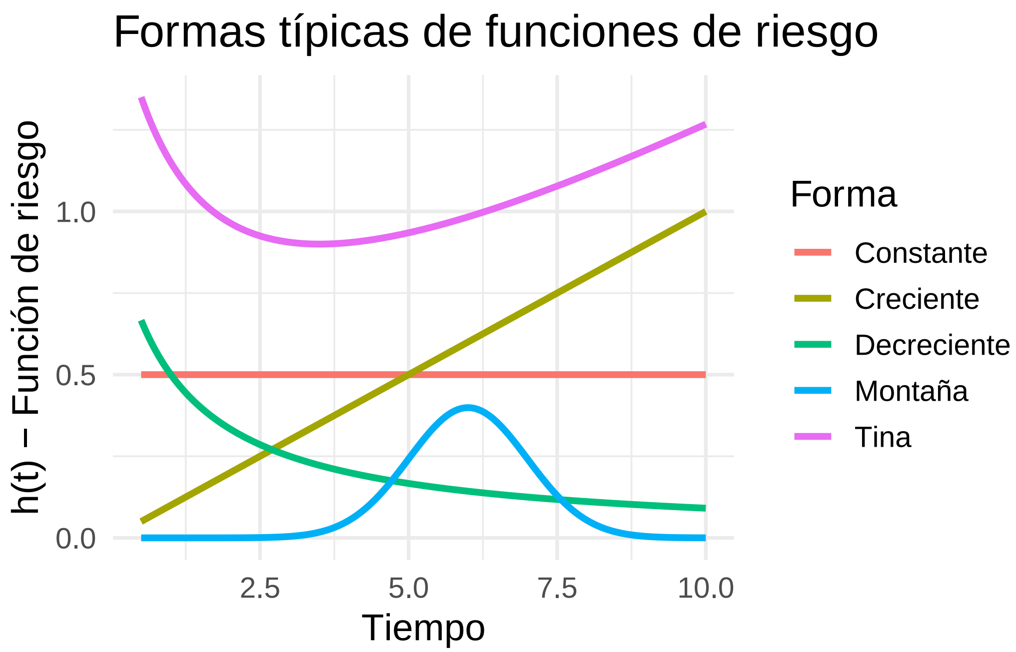
$h(t) = \frac{f(t)}{S(t)}$

- También conocida como:
  - Tasa de falla condicional (confiabilidad)
  - Tasa de mortalidad (demografía)
  - Función de intensidad (procesos estocásticos)

Interpretación:
Tasa instantánea de ocurrencia del evento, dado que se ha sobrevivido hasta t.

3.2 Ejemplos de formas de riesgo

Forma del riesgo	Interpretación
Riesgo creciente	Envejecimiento
Riesgo decreciente	Rejuvenecimiento
Riesgo tipo "tina de baño"	Mortalidad neonatal y senil
Riesgo tipo "montaña"	Recaída tras tratamiento



3.3 Ejemplo: función de riesgo para distribuciones comunes

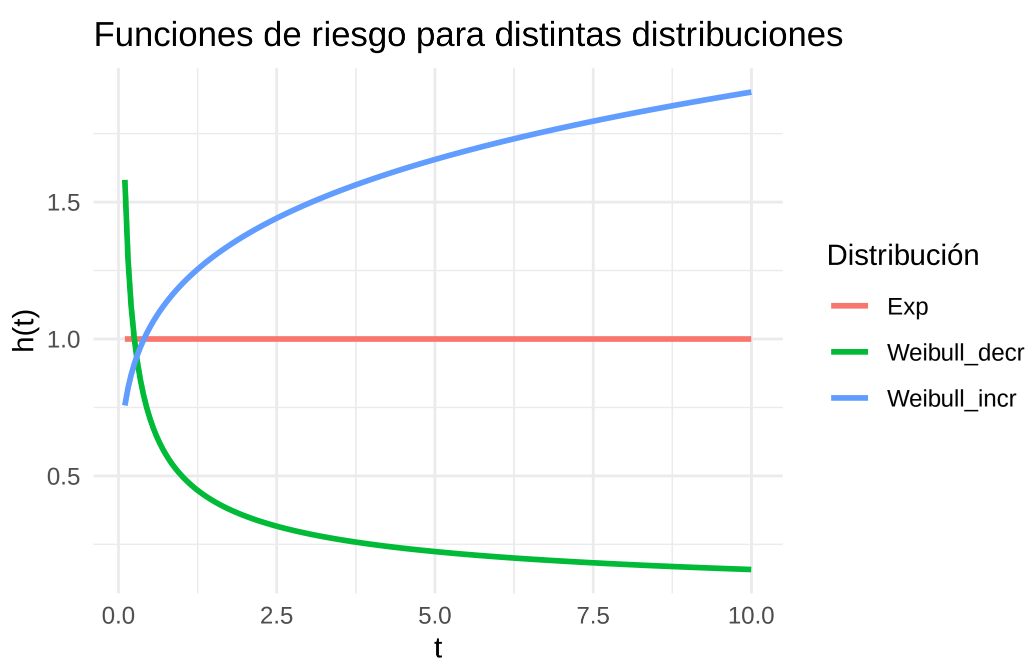
$$h(t) = \frac{f(t)}{S(t)}$$

```
library(survival)

t <- seq(0.1, 10, length.out = 200)
lambda <- 1
k1 <- 0.5  # riesgo decreciente
k2 <- 1.2    # riesgo creciente

hazard_weibull <- function(t, lambda, k) {
  (k / lambda) * (t / lambda)^(k - 1)
}

datos_hazard <- data.frame(
  t = t,
  Exp = rep(lambda, length(t)),
  Weibull_decr = hazard_weibull(t, lambda, k1),
  Weibull_incr = hazard_weibull(t, lambda, k2)
)
```

```r
library(tidyr)
datos_long <- pivot_longer(datos_hazard, -t, names_to = "Distribucion", values_to = "h")

ggplot(datos_long, aes(x = t, y = h, color = Distribucion)) +
  geom_line(size = 1) +
  labs(
    title = "Funciones de riesgo para distintas distribuciones",
    x = "t",
    y = "h(t)",
    color = "Distribución"
  ) +
  theme_minimal()
```

Para la distribución exponencial con $\lambda = 0.5$, $h(t) = \lambda$, constante.

Comparémosla con la distribución Weibull, donde el riesgo puede aumentar o disminuir con el tiempo.



3.4 Otra forma de visualización

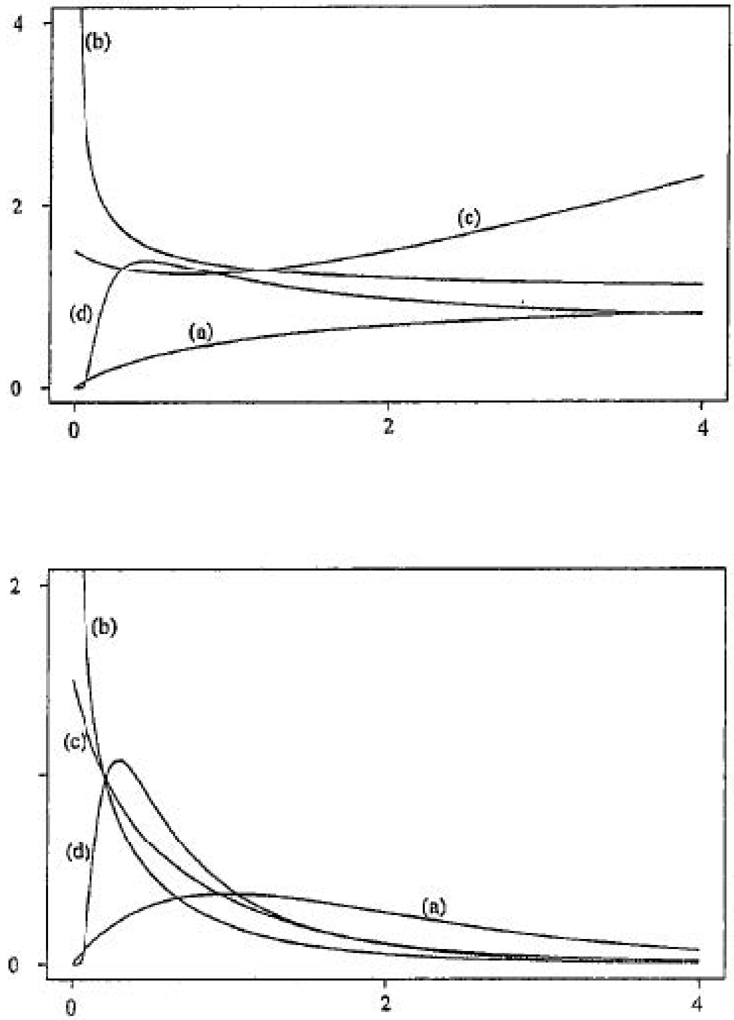


Figure 1.2. Some hazard and probability density functions.

4 Tiempo discreto

4.1 Riesgo en tiempo discreto

Para T discreta con soporte $\{u_1, u_2, \dots\}$:

$$h(t) = P(T = t \mid T \geq t)$$

$$h_k = \frac{P(T = u_k)}{P(T \geq u_k)} = \frac{f(u_k)}{S(u_{k-1})}$$

Usando $f(u_k) = S(u_{k-1}) - S(u_k)$, se obtiene:

$$h_k = 1 - \frac{S(u_k)}{S(u_{k-1})}$$

4.2 Relaciones discretas clave

Función de supervivencia:

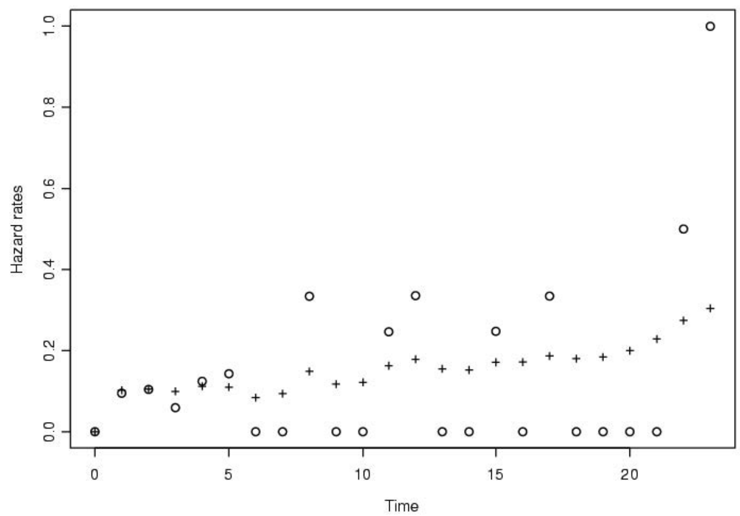
$$S(t) = \prod_{u_k \leq t} (1 - h_k)$$

Función de densidad:

$$f(u_j) = h_j \prod_{k<j} (1 - h_k)$$

> En demografía, $h(t)$ representa la probabilidad de morir en el momento t dado que se ha sobrevivido hasta t.

4.3 Ejemplos de riesgo discreto



4.4 Riesgo acumulado discreto

Dos definiciones equivalentes:

1. Suma directa:
$$H(t) = \sum_{u_k \le t} h_k$$
2. Log-transformación:
$$H(t) = -\sum_{u_k \le t} \log(1 - h_k)$$

Ambas son **monótonas no decrecientes**.

5 Tiempo contínuo

5.1 Riesgo en tiempo continuo

$$h(t) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} P(t < T \le t + \varepsilon \mid T \ge t) = \frac{f(t)}{S(t)}$$

Como $F(t) = 1 - S(t)$, entonces:

$$h(t) = -\frac{d}{dt} \log S(t)$$

Al integrar:

$$\log S(t) = -\int_0^t h(u)\, du$$

$$S(t) = \exp\left(-\int_0^t h(u)\, du\right)$$

> $h(t)\varepsilon$ es la probabilidad **aproximada** de que un evento ocurra en el siguiente instante dado que el individuo ha sobrevivido hasta t.

5.2 Riesgo acumulado continuo

$$H(t) = \int_0^t h(u)\, du \qquad \Rightarrow \qquad S(t) = \exp\{-H(t)\}$$

Si $S(\infty) = 0$, entonces $H(\infty) = \infty$.

5.3 Visualización de funciones

Hazard Rates, Survival Functions, Probability Density Functions, and Expected Lifetimes for Some Common Parametric Distributions

Distribution	*Hazard Rate* $h(x)$	*Survival Function* $S(x)$	*Probability Density Function* $f(x)$	*Mean* $E(X)$
Exponential $\lambda > 0,\ x \geq 0$	λ	$\exp[-\lambda x]$	$\lambda \exp(-\lambda x)$	$\frac{1}{\lambda}$
Weibull $\alpha, \lambda > 0,$ $x \geq 0$	$\alpha\lambda x^{\alpha-1}$	$\exp[-\lambda x^{\alpha}]$	$\alpha\lambda x^{\alpha-1}\exp(-\lambda x^{\alpha})$	$\frac{\Gamma(1+1/\alpha)}{\lambda^{1/\alpha}}$
Gamma $\beta, \lambda > 0,$ $x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - I(\lambda x, \beta)^{*}$	$\frac{\lambda^{\beta} x^{\beta-1}\exp(-\lambda x)}{\Gamma(\beta)}$	$\frac{\beta}{\lambda}$
Log normal $\sigma > 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - \Phi\left[\frac{1\mathrm{n}x - \mu}{\sigma}\right]$	$\frac{\exp\left[-\frac{1}{2}\left(\frac{1\mathrm{n}x-\mu}{\sigma}\right)^{2}\right]}{x(2\pi)^{1/2}\sigma}$	$\exp(\mu + 0.5\sigma^{2})$
Log logistic $\alpha, \lambda > 0, x \geq 0$	$\frac{\alpha x^{\alpha-1}\lambda}{1+\lambda x^{\alpha}}$	$\frac{1}{1+\lambda x^{\alpha}}$	$\frac{\alpha x^{\alpha-1}\lambda}{[1+\lambda x^{\alpha}]^{2}}$	$\frac{\pi \mathrm{Csc}(\pi/\alpha)}{\alpha\lambda^{1/\alpha}}$ if $\alpha > 1$

5.4 Visualización de funciones (cont.)

Normal $\sigma > 0,$ $-\infty < x < \infty$	$\frac{f(x)}{S(x)}$	$1 - \Phi\left[\frac{x-\mu}{\sigma}\right]$	$\frac{\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}\right]}{(2\pi)^{1/2}\sigma}$	μ
Exponential power $\alpha, \lambda > 0, x \geq 0$	$\alpha\lambda^{\alpha} x^{\alpha-1}\exp\{[\lambda x]^{\alpha}\}$	$\exp\{1 - \exp[(\lambda x)^{\alpha}]\}$	$\alpha e \lambda^{\alpha} x^{\alpha-1}\exp[(\lambda x)^{\alpha}] - \exp\{\exp[(\lambda x)^{\alpha}]\}$	$\int_0^{\infty} S(x)dx$
Gompertz $\theta, \alpha > 0, x \geq 0$	$\theta e^{\alpha x}$	$\exp\left[\frac{\theta}{\alpha}(1 - e^{\alpha x})\right]$	$\theta e^{\alpha x}\exp\left[\frac{\theta}{\alpha}(1 - e^{\alpha x})\right]$	$\int_0^{\infty} S(x)dx$
Inverse Gaussian $\lambda \geq 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$\Phi\left[\left(\frac{\lambda}{x}\right)^{1/2}\left(1 - \frac{x}{\mu}\right)\right] - e^{2\lambda/\mu}\Phi\left\{-\left[\frac{\lambda}{x}\right]^{1/2}\left(1 + \frac{x}{\mu}\right)\right\}$	$\left(\frac{\lambda}{2\pi x^{3}}\right)^{1/2}\exp\left[\frac{\lambda(x-\mu^{2})}{2\mu^{2}x}\right]$	μ
Pareto $\theta > 0, \lambda > 0$ $x \geq \lambda$	$\frac{\theta}{x}$	$\frac{\lambda^{\theta}}{x^{\theta}}$	$\frac{\theta\lambda^{\theta}}{x^{\theta+1}}$	$\frac{\theta\lambda}{\theta - 1}$ if $\theta > 1$
Generalized gamma $\lambda > 0, \alpha > 0,$ $\beta > 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - I[\lambda x^{\alpha}, \beta]$	$\frac{\alpha\lambda^{\beta} x^{\alpha\beta-1}\exp(-\lambda x^{\alpha})}{\Gamma(\beta)}$	$\int_0^{\infty} S(x)dx$

* $I(t, \beta) = \int_0^t u^{\beta-1}\exp(-u)du/\Gamma(\beta)$.

5.5 Cuidado con la parametrización

- Las distribuciones **Weibull**, **Gamma**, **Log-normal**, etc., se usan comúnmente para modelar tiempos de vida.
- Cada una tiene formas **teóricas** bien definidas.
- Sin embargo, en R y otros lenguajes de programación:
  - La **parametrización puede cambiar**.
  - Es crucial revisar la **documentación oficial** (`?dweibull`, `?dgamma`, etc.)

Hazard Rates, Survival Functions, Probability Density Functions, and Expected Lifetimes for Some Common Parametric Distributions

Distribution	*Hazard Rate* $b(x)$	*Survival Function* $S(x)$	*Probability Density Function* $f(x)$	*Mean* $E(X)$
Exponential $\lambda > 0,\ x \geq 0$	λ	$\exp[-\lambda x]$	$\lambda \exp(-\lambda x)$	$\frac{1}{\lambda}$
Weibull $\alpha, \lambda > 0,$ $x \geq 0$	$\alpha\lambda x^{\alpha-1}$	$\exp[-\lambda x^{\alpha}]$	$\alpha\lambda x^{\alpha-1}\exp(-\lambda x^{\alpha})$	$\frac{\Gamma(1+1/\alpha)}{\lambda^{1/\alpha}}$
Gamma $\beta, \lambda > 0,$ $x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - I(\lambda x, \beta)^{*}$	$\frac{\lambda^{\beta} x^{\beta-1}\exp(-\lambda x)}{\Gamma(\beta)}$	$\frac{\beta}{\lambda}$
Log normal $\sigma > 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - \Phi\left[\frac{1n x - \mu}{\sigma}\right]$	$\frac{\exp\left[-\frac{1}{2}\left(\frac{1n x-\mu}{\sigma}\right)^{2}\right]}{x(2\pi)^{1/2}\sigma}$	$\exp(\mu + 0.5\sigma^{2})$
Log logistic $\alpha, \lambda > 0, x \geq 0$	$\frac{\alpha x^{\alpha-1}\lambda}{1+\lambda x^{\alpha}}$	$\frac{1}{1+\lambda x^{\alpha}}$	$\frac{\alpha x^{\alpha-1}\lambda}{[1+\lambda x^{\alpha}]^{2}}$	$\frac{\pi \mathrm{Csc}(\pi/\alpha)}{\alpha\lambda^{1/\alpha}}$ if $\alpha > 1$

Figure 1: Funciones

5.6 Weibull: teoría vs. R

5.6.1 Teoría:

$$f(x) = \alpha\lambda x^{\alpha-1}\exp(-\lambda x^{\alpha})$$

- λ: parámetro de escala
- α: parámetro de forma

5.6.2 R (base):

```
dweibull(x, shape, scale = 1)
```

usa

$$f(x) = (\alpha/\sigma)(x/\sigma)^{\alpha-1}\exp(-(x/\sigma)^{\alpha})$$

- El parámetro de forma α y el parámetro de escala

$$\sigma = \lambda^{-1/\alpha}$$

- La forma teórica λ no es directa en R

5.7 Gamma: teoría vs. R

5.7.1 Teoría:

$$f(x) = \frac{\lambda^{\beta}x^{\beta-1}\exp(-\lambda x)}{\Gamma(\beta)}$$

- β: forma (shape)
- λ: tasa (rate)

5.7.2 R (base):

```
dgamma(x, shape, rate = 1, scale = 1/rate)
```

Usa

$$f(x) = \frac{1}{\sigma^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\sigma}.$$

Con parámetro de forma $shape = \alpha$ y $scale = \sigma$

Entonces

- El parámetro de forma coincide, $\beta = \alpha$
- Pero $\sigma = scale = 1/rate = 1/\lambda$

5.8 Log-Normal: teoría vs. R

5.8.1 Teoría:

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

5.8.2 R (base):

```
dlnorm(x, meanlog  sdlog)
```

- `meanlog` y `sdlog` son la media y desviación estandar es escala logarítmica.

5.9 Recomendación

- Revisa siempre la documentación: `?dweibull`, `?dgamma`, `?survreg`, etc.
- Haz pruebas con valores conocidos para validar interpretación.
- Utiliza visualización para entender las funciones.

5.10 Visualización de Funciones en R

Las funciones `Surv()` y `survfit()` del paquete `survival` permiten ajustar y visualizar curvas de Kaplan-Meier de manera eficiente en R, ver Moore (2016) y Therneau & Grambsch (2000).

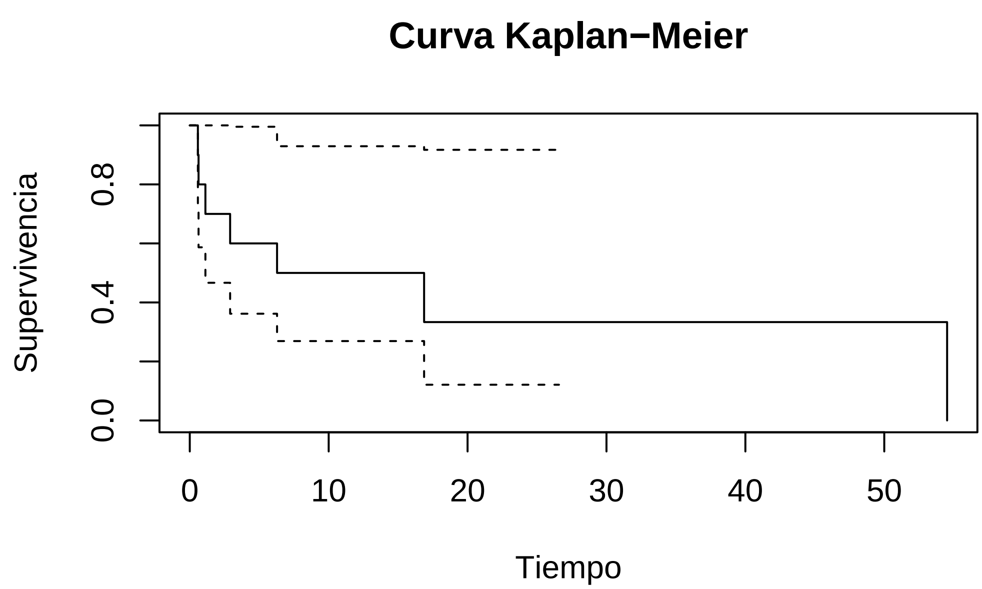
```
# Ejemplo simulado de tiempos de supervivencia
set.seed(123)
tiempos <- rexp(10, rate = 0.05)
status <- rbinom(10, 1, prob = 0.8)
data_sim <- data.frame(time = tiempos, event = status)
# Estimación Kaplan-Meier
km_fit <- survfit(Surv(time, event) ~ 1, data = data_sim)
```

Table 2: data_sim

time	event
16.8691452	1
11.5322054	0
26.5810974	0

time	event
0.6315472	1
1.1242195	1
6.3300243	0
6.2845458	1
2.9053361	1
54.5247293	1
0.5830689	1

```r
plot(km_fit,
     xlab = "Tiempo",
     ylab = "Supervivencia",
     main = "Curva Kaplan-Meier")
```



6 Referencias

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

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the cox model.* Springer.