

**MÉTODOS DE INFERENCIA
PARA LA DISTRIBUCIÓN
WEIBULL: APLICACIÓN EN
FIABILIDAD INDUSTRIAL**

Laura Martínez Fernández

Trabajo fin de máster

Máster en Técnicas Estadísticas

Universidade de Vigo

Julio de 2011

Estela Sánchez Rodríguez y Juan Carlos Pardo Fernández, profesores del Departamento de Estadística e Investigación Operativa de la Universidade de Vigo,

HACEN CONSTAR

que el presente trabajo titulado *Métodos de inferencia para la distribución Weibull: Aplicación en Fiabilidad Industrial* ha sido realizado por Laura Martínez Fernández bajo su dirección para su presentación como Trabajo Fin de Máster del Máster en Técnicas Estadísticas.

Fdo.: Estela Sánchez Rodríguez

Fdo.: Juan Carlos Pardo Fernández

Agradecimientos

Deseo expresar mi más sincero agradecimiento a los profesores y directores de este proyecto Estela Sánchez Rodríguez y Juan Carlos Pardo Fernández, por su dedicación y su constante apoyo a lo largo de este proyecto, que han sido fundamentales.

Gracias a mi familia, sobre todo a mis padres, que a lo largo de todos estos años me han apoyado y dado ánimos para seguir adelante.

Quiero dar las gracias también a mis amigos por la infinidad de momentos buenos que me hacen pasar y por apoyarme y darme ánimos en los momentos malos.

También quiero dedicar este proyecto a la gente que ya no está, aunque siguen estando presentes cada día que pasa.

Por último, quiero dar las gracias a Iker, que siempre saca lo mejor de mí y que me da ánimos todos los días para seguir adelante, incluso en los momentos más complicados.

Gracias a todos.

Índice general

Introducción	1
1. Conceptos básicos	5
1.1. Definiciones Preliminares	5
1.2. Modelos de distribución importantes	8
1.2.1. La distribución Exponencial	8
1.2.2. La distribución Weibull	10
1.2.3. La distribución Gumbel	12
1.2.4. La distribución Normal	14
1.2.5. La distribución Log-Normal	16
1.3. Muestra aleatoria simple y verosimilitud	18
2. Tipos de censura	21
2.1. Tipos de censura	22
2.1.1. Censura tipo I	22
2.1.2. Censura tipo II	23
2.1.3. Censura Aleatoria Independiente	23
3. Inferencia para la distribución Weibull	25
3.1. Gráficos probabilísticos	25
3.2. Estimación por máxima verosimilitud	31
3.2.1. Máxima verosimilitud para la distribución Exponencial	32

3.2.2. Máxima verosimilitud para la distribución Weibull . . .	33
3.3. Intervalos de confianza	34
3.4. Contrastes de hipótesis basados en el test de razón de verosimilitudes	35
3.4.1. Test de razón de verosimilitudes para el parámetro de forma β	37
3.4.2. Test de razón de verosimilitudes para el parámetro escala λ	38
3.5. Pruebas de bondad de ajuste	39
3.5.1. Contraste de Kolmogorov-Smirnov	39
3.5.2. Contraste de Cramér-von Misses	40
3.5.3. Contraste de Anderson-Darling	40
4. Software R	41
4.1. Paquete Renext	42
4.2. Paquete STAR	44
4.3. Paquete Sim.DiffProc	47
5. Simulaciones y aplicación a datos reales	51
5.1. Estudio de simulación	51
5.2. Aplicación a datos reales	56
Bibliografía	59

Introducción

En ingeniería, la fiabilidad es comúnmente definida como la probabilidad de que un dispositivo realice su función cuando es requerida, durante un tiempo especificado y bajo unas condiciones de entorno establecidas. Para ser más precisos necesitamos introducir los conceptos tiempo de vida útil de un producto y fallo.

Llamaremos fallo a la pérdida de alguna de las propiedades del dispositivo, la cual impide el funcionamiento de éste. El tiempo de vida útil de un producto es el período de tiempo en el que puede ser utilizado, es decir, el tiempo que transcurre entre su puesta en funcionamiento y su fallo.

La fiabilidad en la ingeniería está orientada al estudio de los tiempos de fallo. El problema reside en predecir si se puede producir un fallo y cuándo ocurrirá. Esta información es muy útil para las políticas de mantenimiento e inspección de una empresa así como los plazos de garantía de los productos. También es muy útil para predecir costes debidos al mantenimiento y a los fallos ocasionales que pueden ocurrir durante el tiempo de vida del dispositivo.

El origen de la fiabilidad se puede atribuir a los estudios para evaluar la mortalidad derivada de las epidemias y a los métodos actuariales desarrollados por las compañías de seguros, para determinar el riesgo de sus pólizas. La herramienta utilizada para el cálculo de esta fiabilidad eran las tablas de vida.

A principios de 1900 se utilizaban los métodos actuariales para estudiar la

supervivencia de pacientes con determinados tratamientos y para estudiar la fiabilidad de los ferrocarriles. La teoría matemática de la fiabilidad se desarrolla por las demandas de la tecnología. El área de mantenimiento de máquinas es donde la fiabilidad se aplica con sofisticadas técnicas matemáticas.

En 1939 Walodie Weibull, propuso una distribución para describir la duración de los materiales, que más tarde se denominaría distribución de Weibull. Esta distribución, la cual trataremos en este proyecto, es utilizada en infinidad de aplicaciones debido a su gran versatilidad.

Epstein y Sobel (1953) empezaron a trabajar con la distribución Exponencial como modelo para estudiar el tiempo de vida útil de un dispositivo. La distribución exponencial no tiene memoria, es decir, el tiempo de vida útil de un determinado dispositivo no influye en la probabilidad de que éste falle. La popularidad de esta distribución es el gran uso que se ha hecho de ella en trabajos de fiabilidad debido a su simplicidad en los cálculos.

En los años noventa, la investigación de la fiabilidad toma nuevas direcciones gracias a M. B. Mendel. Sus métodos se pueden encontrar en publicaciones sobre problemas de fiabilidad en la ingeniería, entre los que destacan los de Shortle y Mendel (1996).

Nuestro proyecto se estructura en cinco capítulos diferenciados. En el Capítulo 1 se definen conceptos básicos en estadística como la función de distribución y densidad y se introducen los conceptos de función de fiabilidad y riesgo. Asimismo, se caracterizan modelos de distribuciones utilizados en el estudio de tiempos de vida útil, como la distribución Exponencial, Weibull, Gumbel, Normal y Log-Normal y las relaciones existentes entre algunas de éstas. Además, se introducen conceptos como muestra aleatoria simple y función de verosimilitud, que se utilizarán posteriormente en el desarrollo de este proyecto. En el Capítulo 2 se trata el concepto de censura, que aparece frecuentemente en los estudios de fiabilidad, centrándose en la censura por la derecha y estudiando los tipos de censura que se pueden presentar. En to-

do caso, las aplicaciones de este proyecto se harán para muestras completas y muestras con censura tipo II. El Capítulo 3 se centra en la distribución Weibull, donde se proporcionan dos métodos, el gráfico y el de máxima verosimilitud para proporcionar estimaciones de los parámetros de la Weibull tanto para datos completos como para censurados. En el Capítulo 4 se revisan los paquetes del software R relacionados con este tema. En el Capítulo 5 se realiza un estudio de simulación para estudiar los intervalos de confianza basados en la normalidad asintótica para los estimadores de máxima verosimilitud de los parámetros de la distribución *Weibull* y una aplicación de las técnicas de inferencia a datos reales procedentes de una empresa del sector de la automoción.

Este proyecto surge como continuación de un proyecto fin de carrera presentado en la Escuela Técnica Superior de Ingeniería Industrial de la Universidad de Vigo por Hermida (2007).

Los libros de referencia de este proyecto son Lawless (2003) y Griful Ponsati (2004). Otras referencias bibliográficas asociadas a este campo son Nelson (1982), Kalbfleish y Prentice (2002), Prabhakar y otros (2004), Lee y Wenyu Wang (2003), Klein y Moeschberg (1982), Nachlas (1995), Epstein y Sobel (1953) y Jiang y otros (1999).

Capítulo 1

Conceptos básicos

En este capítulo introduciremos algunos conceptos generales de teoría de fiabilidad, así como los modelos de distribución más utilizados para este tipo de estudios.

1.1. Definiciones Preliminares

En general, asumiremos que el tiempo de vida es una variable continua, la cual denotaremos por T , y que toma valores en el intervalo $[0, \infty)$. Esta distribución viene caracterizada por las funciones definidas a continuación.

La función de distribución de la variable aleatoria continua T , $F(t)$, representa la probabilidad acumulada de fallo hasta el tiempo t , es decir

$$F(t) = P(T \leq t).$$

Esta función es continua, monótona no decreciente y además verifica que

$$\lim_{t \rightarrow -\infty} F(t) = 0$$

y

$$\lim_{t \rightarrow \infty} F(t) = 1.$$

La función de fiabilidad en el tiempo t , $R(t)$, representa la proporción de unidades que no han fallado antes del instante t y viene dada por

$$R(t) = P(T > t) = \int_t^{\infty} f(x) dx = 1 - F(t).$$

Esta función es continua, monótona decreciente y además verifica que

$$R(0) = 1$$

y

$$R(\infty) = \lim_{t \rightarrow \infty} (1 - F(t)) = 0.$$

El p -cuantil de la distribución de T es el valor t_p que verifica

$$F(t_p) = P(T \leq t_p) = p,$$

lo que equivale a la expresión

$$t_p = F^{-1}(p).$$

donde F^{-1} es la función inversa de F .

A partir de la función de distribución definimos la función de densidad $f(t)$ de la variable aleatoria continua T como la derivada de la función de distribución, es decir, $f(t) = F'(t)$.

Un concepto muy importante a tener en cuenta cuando trabajamos con distribuciones asociadas a tiempos de vida es la función de riesgo o tasa de fallo $h(t)$, que se define como la probabilidad instantánea de que una componente falle en el instante t . En términos de probabilidad se interpreta como el límite de la probabilidad condicionada de que T falle antes del tiempo $t + \Delta t$ sabiendo que no había fallado en el instante t . Esto no es más que el cociente entre la función de densidad $f(t)$ y la función de fiabilidad $R(t)$. Entonces, la función de riesgo es

$$h(t) = \frac{f(t)}{R(t)}.$$

Definimos la función de riesgo acumulada, $H(t)$, como

$$H(t) = \int_0^t h(x) dx .$$

La función de riesgo acumulada y la función de distribución verifican la siguiente relación

$$H(t) = -\log(1 - F(t)).$$

La función de riesgo es una característica muy importante de las distribuciones asociadas a variables correspondientes a tiempos de vida.

- Funciones de riesgo decrecientes surgen cuando al principio del funcionamiento la probabilidad de fallo es alta debido a posibles defectos de fabricación.
- Funciones de riesgo constantes tienden a aparecer en conjuntos donde los fallos son debidos a un fenómeno aleatorio como accidentes o shocks.
- Funciones de riesgo crecientes aparecen cuando se produce un envejecimiento del componente.

Estas tres formas de riesgo o fallo se combinan para generar la curva de bañera o *bathtub curve*, que se corresponde con la figura (1.1), la cual es una curva típica en fiabilidad. La primera zona se denomina mortalidad infantil, la siguiente zona de vida útil y la última es la zona de desgaste.

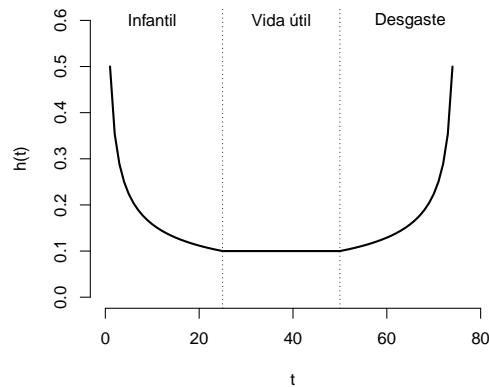


Figura 1.1: Curva de la bañera

1.2. Modelos de distribución importantes

En esta sección vamos a tratar algunos modelos de distribución importantes en trabajos de fiabilidad.

1.2.1. La distribución Exponencial

Históricamente, este modelo de distribución fue muy utilizado en el trabajo con tiempos de vida útil, debido a la simplicidad de los métodos estadísticos que proporciona. La función de densidad de una variable aleatoria con distribución Exponencial es

$$f(t) = \lambda e^{-\lambda t}, \quad (1.1)$$

con $t \geq 0$, siendo $\lambda \geq 0$ el parámetro del modelo. Su función de distribución es

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

y su función de fiabilidad es

$$R(t) = e^{-\lambda t}.$$

Su función de riesgo es constante y su valor es el parámetro del modelo, λ ,

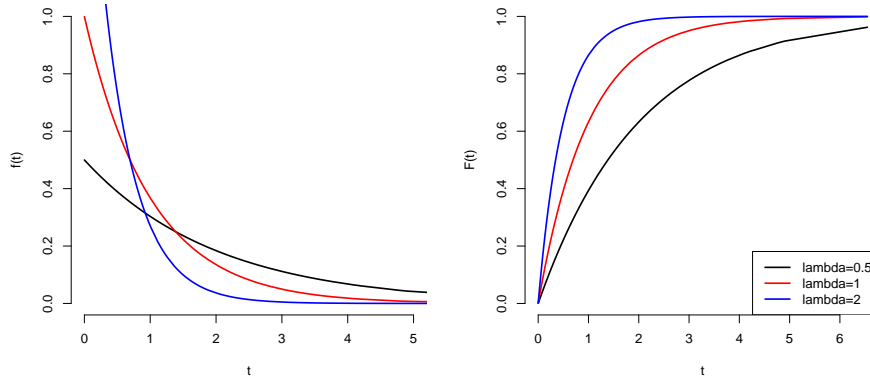


Figura 1.2: Funciones de densidad (izquierda) y distribución (derecha) Exponencial con $\lambda = 0.5, 1$ y 2 .

es decir

$$h(t) = \lambda.$$

Emplearemos la notación $T \sim Exp(\lambda)$ para indicar que T es una variable aleatoria con función de densidad (1.1). La media y la varianza de esta distribución vienen dadas por los valores $1/\lambda$ y $1/\lambda^2$, respectivamente. El p -cuantil es

$$t_p = -\frac{1}{\lambda} \log(1 - p).$$

Cuando el parámetro λ es igual a 1 la llamaremos distribución Exponencial estándar, cuyas funciones de densidad y distribución se corresponden con la línea de color rojo de la figura (1.2).

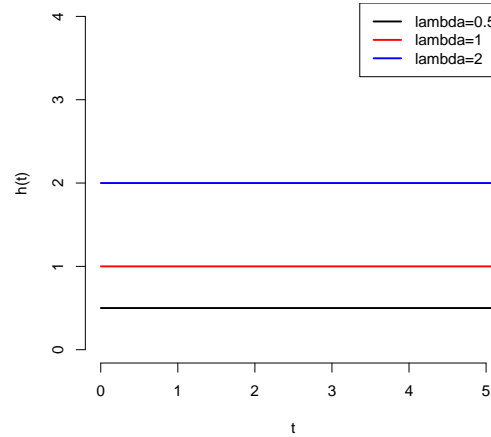


Figura 1.3: Función de riesgo de la distribución Exponencial con $\lambda = 0.5, 1$ y 2 .

1.2.2. La distribución Weibull

La distribución Weibull es quizá el modelo más utilizado para tratar problemas con tiempos de vida en fiabilidad industrial y al que nos ceñiremos en este proyecto. Una variable aleatoria continua T tendrá una distribución Weibull, y lo denotaremos $T \sim Weibull(\lambda, \beta)$, si su función de densidad es

$$f(t) = \lambda\beta(\lambda t)^{\beta-1} \exp\{-(\lambda t)^\beta\}, \quad (1.2)$$

con $t \geq 0$, siendo $\lambda > 0$ y $\beta > 0$ los parámetros del modelo, que se denominan de escala y forma, respectivamente. Su función de distribución, que es

$$F(t) = 1 - \exp\{-(\lambda t)^\beta\}. \quad (1.3)$$

Las funciones de densidad y distribución del modelo Weibull están representadas en la figura (1.4). Su función de fiabilidad es

$$R(t) = \exp\{-(\lambda t)^\beta\}, \quad (1.4)$$

y su función de riesgo, que se corresponde con la figura (1.5), es

$$h(t) = \lambda\beta(\lambda t)^{\beta-1}. \quad (1.5)$$

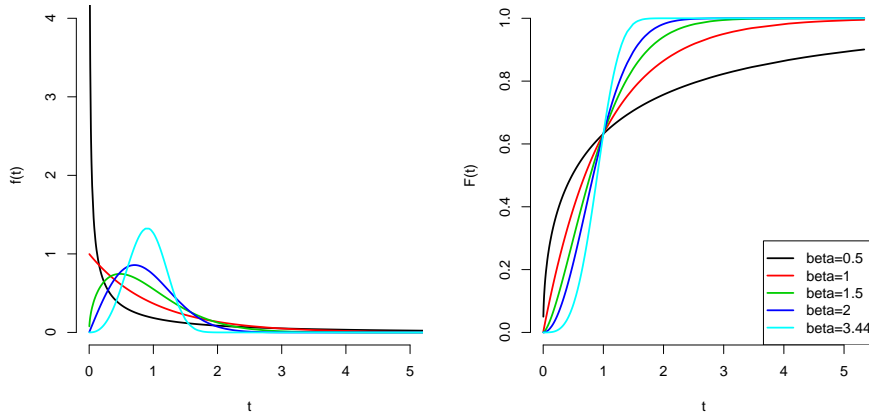


Figura 1.4: Funciones de densidad (izquierda) distribución (derecha) Weibull con $\beta = 0.5, 1, 1.5, 2$ y 3.44 y $\lambda = 1$.

La función de riesgo adopta las siguientes formas en función del parámetro de forma, β :

- Si $\beta < 1$ la función de riesgo es decreciente, es decir, la tasa de fallo disminuye al aumentar el tiempo.
- Si $\beta = 1$ la función de riesgo es constante, por lo que no depende del tiempo. En este caso, la distribución Weibull coincide con la Exponencial.
- Si $\beta > 1$ la función de riesgo es creciente. En particular, si $1 < \beta < 2$ la función de riesgo crece rápido en el origen y muy poco a medida que t crece; para $\beta = 2$ el riesgo crece linealmente con el tiempo y para $\beta > 2$ crece un poco con t próximo a cero y después rápido. Es oportuno considerar la posibilidad $\beta = 3.44$, ya que en este caso, la distribución Weibull se parece a la Normal.

El p-cuantil de esta distribución es

$$t_p = \frac{1}{\lambda} [-\log(1-p)]^{\frac{1}{\beta}}.$$

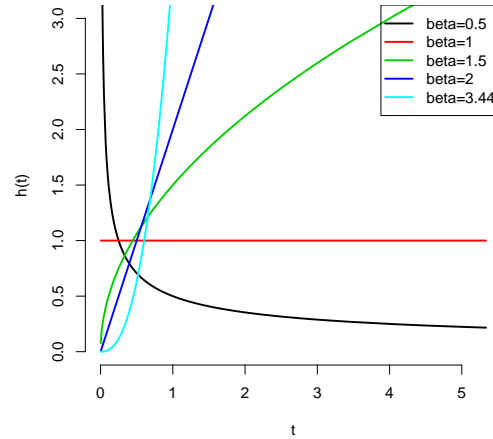


Figura 1.5: Función de riesgo de la distribución Weibull con $\beta = 0.5, 1, 1.5, 2$ y 3.44 y $\lambda = 1$.

El valor $1/\lambda$ es aproximadamente el percentil 63.2% y se interpreta como el valor de la variable del tiempo de vida en el que han fallado el 63.2% de las unidades. En el capítulo referente a la inferencia volveremos a trabajar con esta distribución.

1.2.3. La distribución Gumbel

Nuestro interés en esta distribución no se centra en su uso directo como modelo para variables aleatorias asociadas a tiempos de vida, sino por su relación con la distribución Weibull. La variable aleatoria T tendrá distribución Gumbel, y lo denotaremos por $T \sim Gumbel(\tau, \omega)$, si su función de densidad es

$$f(t) = \omega^{-1} \exp \left[- \left(\frac{t - \tau}{\omega} \right) - \exp \left[- \left(\frac{t - \tau}{\omega} \right) \right] \right],$$

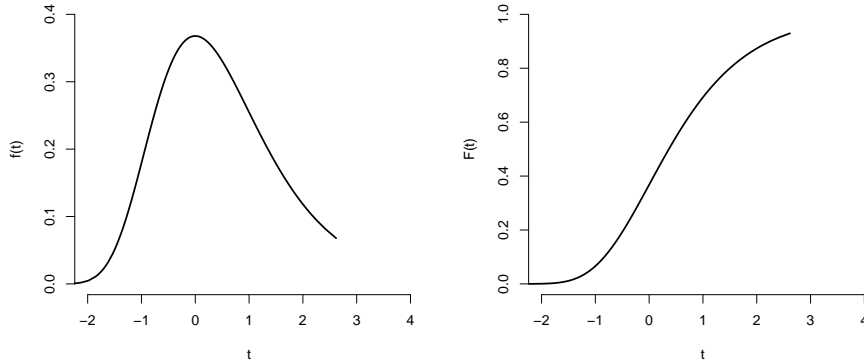


Figura 1.6: Funciones de densidad (izquierda) y distribución (derecha) de la distribución Gumbel con $\tau = 0$ y $\omega = 1$.

con $t \in \mathbb{R}$, siendo $\omega > 0$ y $\tau \in \mathbb{R}$ los parámetros del modelo, que se denominan de escala y localización, respectivamente. Su función de distribución es

$$F(t) = \exp \left[- \exp \left[- \left(\frac{t - \tau}{\omega} \right) \right] \right],$$

y su función de fiabilidad es

$$R(t) = 1 - \exp \left[- \exp \left[- \left(\frac{t - \tau}{\omega} \right) \right] \right].$$

El p-cuantil de esta distribución es

$$y_p = \tau + \omega \log[-\log(1 - p)].$$

Si T es una variable aleatoria con distribución $Weibull(\lambda, \beta)$, entonces, la variable aleatoria transformada, $Y = \log T$ tendrá distribución Gumbel con parámetros $\tau = \log \lambda$ y $\omega = \beta^{-1}$. Cuando trabajamos con variables asociadas a tiempos de vida, es conveniente tratar con los logaritmos de estas variables. Cuando $\tau = 0$ y $\omega = 1$, esta distribución se llama distribución Gumbel estándar, cuyas funciones de densidad y distribución están representadas en la figura (1.6) y su función de riesgo está representada en la figura (1.7).

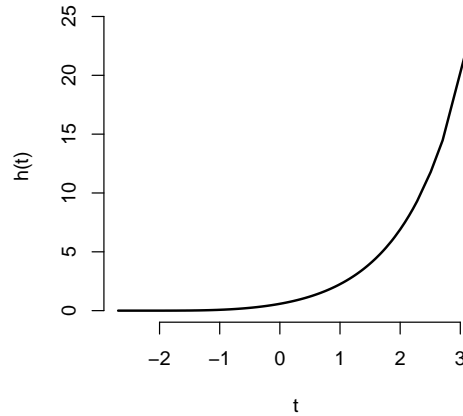


Figura 1.7: Función de riesgo de la distribución Gumbel con $\tau = 0$ y $\omega = 1$.

1.2.4. La distribución Normal

Es la distribución utilizada con más frecuencia en estadística, aunque no ocurre así en estudios de fiabilidad debido a su carácter simétrico, ya que habitualmente los tiempos de vida presentan un comportamiento asimétrico.

Una variable aleatoria continua T tiene distribución Normal, y lo denotaremos $T \sim N(\mu, \sigma)$, si su función de densidad es

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2} \left(\frac{t - \mu}{\sigma}\right)^2\right],$$

con $t \in \mathbb{R}$, siendo μ y σ los parámetros del modelo, que se denominan media y desviación típica, respectivamente. Sus funciones de distribución y de fiabilidad son

$$F(t) = \int_{-\infty}^t \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right] dx,$$

y

$$R(t) = \int_t^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right] dx,$$

respectivamente, para las cuales no existen expresiones analíticas explícitas, aunque sí se pueden aproximar numéricamente. Véanse las representaciones gráficas de la figura (1.8). La función de riesgo, dado que la función de distribución Normal no tiene forma explícita, no puede ser calculada de manera directa mediante el cociente entre la función de densidad y la función de fiabilidad, aunque sí se puede aproximar numéricamente. Esta función está representada en la figura (1.9). Cuando $\mu = 0$ y $\sigma = 1$, esta distribución se denomina distribución Normal estándar, cuya función de densidad, denotada por $\varphi(t)$ es

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \exp \left[\frac{-t^2}{2} \right].$$

Su función de distribución es

$$\Phi(t) = \int_{-\infty}^t \varphi(u) du. \quad (1.6)$$

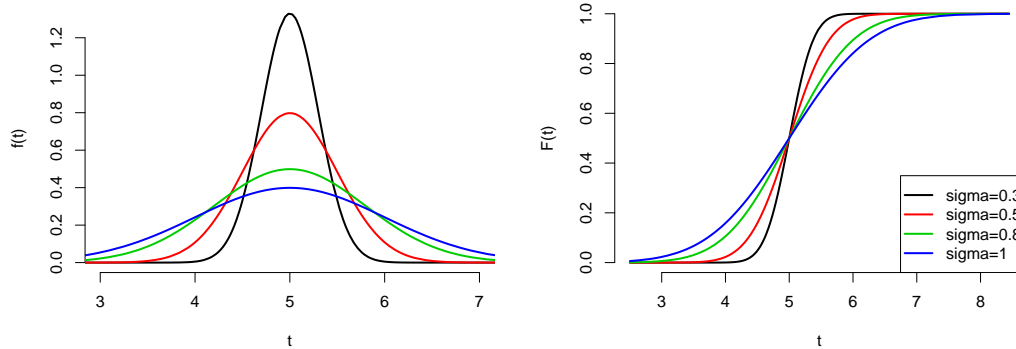


Figura 1.8: Funciones de densidad (izquierda) y distribución (derecha) de la distribución Normal con $\mu = 5$ y $\sigma = 0.3, 0.5, 0.8$ y 1 .

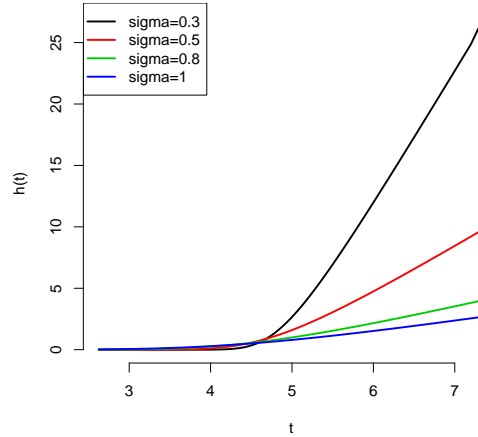


Figura 1.9: Función de riesgo de la distribución Normal con $\mu = 5$ y $\sigma = 0.3, 0.5, 0.8$ y 1 .

1.2.5. La distribución Log-Normal

Este modelo de distribución se ha utilizado en multitud de aplicaciones asociadas a ingeniería, medicina y otras áreas. Diremos que la variable aleatoria T sigue una distribución Log-Normal, y lo denotaremos como $LogN(\mu, \sigma)$, si la variable $Y = \log T$ sigue una distribución Normal con media μ y desviación típica σ . Por lo tanto, si la función de densidad de la variable aleatoria Y es

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \left(\frac{y - \mu}{\sigma}\right)^2\right],$$

con $y \in \mathbb{R}$, entonces la función de densidad de $T = \exp(Y)$ es

$$f(t) = \frac{1}{\sigma t \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{\log t - \mu}{\sigma}\right)^2\right],$$

con $t > 0$, que es la densidad de la distribución Log-Normal. Su función de distribución es

$$F(t) = \Phi\left(\frac{\log t - \mu}{\sigma}\right),$$

donde Φ es la función de distribución Normal estándar dada en la expresión (1.6). Las funciones de densidad y distribución están representadas en la figura (1.10). La función de fiabilidad es

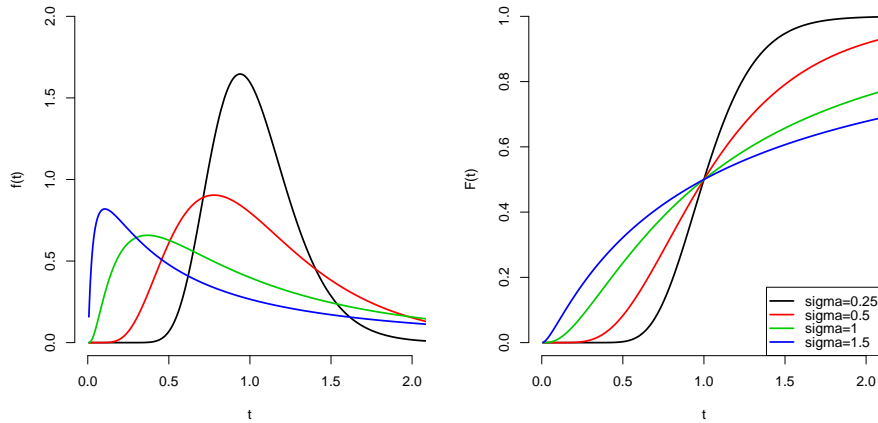


Figura 1.10: Funciones de densidad (izquierda) y distribución (derecha) de la distribución Log-Normal para $\mu = 0$ y $\sigma = 0.25, 0.5, 1$ y 1.5 .

$$R(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right).$$

La función de riesgo, $h(t)$, que se corresponde con la figura (1.11), se caracteriza por valer cero en $t = 0$, y como podemos comprobar, para el caso $\sigma = 1.5$, la función de riesgo crece hasta un máximo y luego decrece, acercándose a cero a medida que t tiende a ∞ .

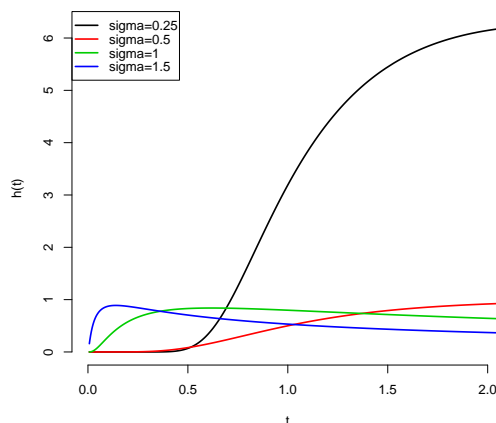


Figura 1.11: Función de riesgo de la distribución Log-Normal para $\mu = 0$ y $\sigma = 0.25, 0.5, 1$ y 1.5 .

1.3. Muestra aleatoria simple y verosimilitud

Definimos una muestra aleatoria simple de tamaño n de una población T como un conjunto de n variables aleatorias independientes T_1, \dots, T_n que tienen la misma distribución que T . Cuando reordenamos los elementos de esta muestra de menor a mayor, obtendremos el conjunto $T_{(1)}, \dots, T_{(n)}$, al que llamaremos muestra ordenada y que utilizaremos en los siguientes capítulos.

Supongamos que T_1, \dots, T_n es una muestra aleatoria simple de n elementos, los cuales proceden de una determinada población con función de densidad $f_\theta(t)$, donde $\theta = (\theta_1, \dots, \theta_k)'$ es un vector de parámetros desconocidos que toman valores en un conjunto Ω . La función de verosimilitud para θ , que denotaremos por $L(\theta)$ es

$$L(\theta) = \prod_{i=1}^n f_\theta(T_i). \quad (1.7)$$

Evidentemente, la función $L(\theta)$ depende de la muestra, pero es común suprimirla en la notación.

Si $\hat{\theta}$ es un elemento de Ω donde la función de verosimilitud (1.7) alcanza un máximo, entonces $\hat{\theta}$ se denomina estimador de máxima verosimilitud del parámetro θ , que en muchos modelos de distribución existe y es único. Muchas veces es más cómodo trabajar con la función log-verosimilitud, que también alcanza un máximo en $\hat{\theta}$. Esta función de log-verosimilitud tiene la expresión

$$l(\theta) = \log(L(\theta)) = \sum_{i=1}^n \log(f_{\theta}(T_i)). \quad (1.8)$$

En la mayoría de las ocasiones, el estimador de máxima verosimilitud $\hat{\theta}$ se obtiene resolviendo las k ecuaciones

$$\frac{\partial l(\theta)}{\partial \theta_i} = 0, \quad i = 1, \dots, k,$$

donde θ_i es la i -ésima componente del vector de parámetros $\theta = (\theta_1, \dots, \theta_k)'$. El método de máxima verosimilitud es muy utilizado en la práctica debido a que proporciona estimadores consistentes, asintóticamente eficientes, insesgados y normalmente distribuidos en muestras grandes.

Ahora definiremos la matriz de información de Fisher, cuya matriz inversa es la matriz de varianzas-covarianzas que utilizaremos para construir los intervalos de confianza basados en las estimaciones de máxima verosimilitud de nuestros parámetros. La matriz de información de Fisher para un estimador de máxima verosimilitud $\hat{\theta}$ del parámetro θ es

$$I = \left[-\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'} \right]_{\theta=\hat{\theta}}$$

que es una matriz cuadrada con dimensiones $k \times k$. Cada elemento de la matriz es el opuesto de la segunda derivada parcial de la función de log-verosimilitud evaluada posteriormente en los estimadores de máxima verosimilitud del modelo de distribución correspondiente. La matriz de varianzas-covarianzas del vector aleatorio $\hat{\theta}$ es

$$Var(\hat{\theta}) = \left[-\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'} \right]_{\theta=\hat{\theta}}^{-1}, \quad (1.9)$$

que es la inversa de la matriz de información de Fisher. Los elementos diagonales de esta última matriz se corresponden con las varianzas de los estimadores de máxima verosimilitud de cada componente del parámetro.

Capítulo 2

Tipos de censura

Normalmente, los datos asociados a tiempos de vida presentan observaciones incompletas. La estimación de las características de la fiabilidad cambian respecto a la estimación clásica de muestras completas. Un fenómeno que produce datos incompletos es la censura.

Formalmente, se dice que una observación es censurada por la derecha de un valor C si el valor exacto de tal observación no es conocida pero sí se sabe que excede del tiempo C . Análogamente, una observación es censurada por la izquierda de C cuando sólo se sabe que el valor de la observación es menor que C . También puede aparecer la censura por intervalo, donde los datos censurados presentan censura por la derecha y por la izquierda. Es más común que aparezca la censura por la derecha que por la izquierda con datos asociados a tiempos de vida. A partir de ahora, cuando nos refiramos a censura, nos estaremos refiriendo específicamente a censura por la derecha. Para poder hablar de censura tendremos que tener en cuenta la forma en cómo se obtuvieron los datos. El problema fundamental es determinar la distribución de la muestra y correspondiente función de verosimilitud de un proceso para así determinar los métodos estadísticos que derivan de ello.

2.1. Tipos de censura

Por simplicidad, supondremos que se dispone de una muestra aleatoria simple T_1, \dots, T_n de tiempos de vida. Los tipos de censura que vamos a explicar a continuación son censura tipo I, censura tipo II y censura aleatoria independiente.

2.1.1. Censura tipo I

Este mecanismo de censura aparece cuando se observa a cada uno de los n individuos hasta un tiempo predeterminado $C > 0$, observando si los tiempos de vida T_i , con $i = 1, \dots, n$, asociados a cada individuo son mayores o menores que este valor C . Si $T_i > C$ estaremos en condiciones de saber que cada T_i es mayor que ese valor prefijado C pero no su valor exacto, por lo que diremos que éste es un dato censurado.

Estos datos pueden ser convenientemente representados por n pares de variables aleatorias (Z_i, Δ_i) , donde

$$Z_i = \min(T_i, C),$$

y

$$\Delta_i = \begin{cases} 1 & \text{si } T_i \leq C \\ 0 & \text{si } T_i > C \end{cases}$$

es el indicador censura, que valdrá 1 si el dato es observado y 0 si el dato es censurado.

La contribución del par de variables (Z_i, Δ_i) a la función de verosimilitud es

$$f(Z_i)^{\Delta_i} R(Z_i)^{1-\Delta_i},$$

donde los datos observados contribuyen en la expresión anterior con su función de densidad y los datos censurados con su función de fiabilidad. Por lo

tanto, la función de verosimilitud en el caso de censura tipo I es

$$\prod_{i=1}^n f(T_i)^{\Delta_i} R(C)^{1-\Delta_i} = \prod_{i=1}^n f(Z_i)^{\Delta_i} R(Z_i)^{1-\Delta_i}$$

2.1.2. Censura tipo II

Este tipo de censura se presenta cuando comenzamos el estudio de los n individuos al mismo tiempo y lo damos por finalizado una vez que observamos r fallos, donde $r \leq n$ es un número prefijado. Este tipo de pruebas permiten ahorrar tiempo y dinero, ya que en algunas ocasiones puede que los individuos tarden mucho tiempo en fallar. A pesar de que algunas pruebas se formulan con censura tipo II, ésta presenta la desventaja de que el tiempo total de estudio, $T_{(r)}$, es desconocido al empezar dicha prueba. Formalmente, supongamos que tenemos los r tiempos de vida ordenados más pequeños, $T_{(1)} \leq \dots \leq T_{(r)}$, de una muestra aleatoria de orden n . Si T_1, \dots, T_n tienen distribución continua y son independientes e idénticamente distribuidos con función de densidad $f(t)$ y función de fiabilidad $R(t)$, entonces la función de verosimilitud de la muestra ordenada $T_{(1)}, \dots, T_{(r)}$ es

$$\frac{n!}{(n-r)!} f(T_{(1)}) \dots f(T_{(r)}) [R(T_{(r)})]^{n-r}. \quad (2.1)$$

2.1.3. Censura Aleatoria Independiente

Este tipo de procesos se dan mucho en la práctica y son aquellos en los que se supone que cada individuo tiene asociado un tiempo de vida T y un tiempo de censura C , donde T y C son variables aleatorias continuas con funciones de densidad $f(t)$ y $g(t)$ y funciones de fiabilidad $R(t)$ y $S(t)$, respectivamente. Todos los tiempos de vida así como los tiempos censurados son mutuamente independientes. Este tipo de censura es una generalización de la censura tipo I, ya que en vez de considerar un tiempo prefijado C constante, tendremos asociado para cada T_i un valor C_i , con $i = 1, \dots, n$. Como pasaba con la

censura tipo I, tendremos que

$$Z_i = \min(T_i, C_i)$$

y

$$\Delta_i = \begin{cases} 1 & \text{si } T_i \leq C_i \\ 0 & \text{si } T_i > C_i. \end{cases}$$

Los datos de las observaciones sobre n individuos vendrán dados por los pares (Z_i, Δ_i) , $i = 1, \dots, n$. Entonces, la contribución del par (Z_i, Δ_i) a la función de verosimilitud es

$$[f(Z_i)S(Z_i)]^{\Delta_i} [g(Z_i)R(Z_i)]^{1-\Delta_i},$$

donde un dato censurado contribuye con su función de densidad y la función de fiabilidad del dato no censurado y un dato observado contribuye con su función de densidad y la función de fiabilidad de la censura. En este caso, la función de verosimilitud es

$$\prod_{i=1}^n [f(Z_i)S(Z_i)]^{\Delta_i} [g(Z_i)R(Z_i)]^{1-\Delta_i}.$$

Capítulo 3

Inferencia para la distribución Weibull

En este capítulo nos centraremos en la distribución Weibull. Existen diferentes métodos para la obtención de estimadores puntuales adecuados para los parámetros de forma y escala tanto para datos completos como para datos censurados. En este capítulo estudiaremos un método gráfico y la estimación por máxima verosimilitud. También hablaremos de cómo construir intervalos de confianza para los parámetros y de contrastes de hipótesis aplicados a los parámetros de esta distribución, así como su bondad de ajuste.

3.1. Gráficos probabilísticos

En esta sección presentaremos un método gráfico para determinar si es o no adecuado el modelo Weibull para un conjunto de datos. Además de su uso para establecer visualmente la bondad del ajuste del modelo Weibull, se suelen utilizar para estimar gráficamente los parámetros de dicho modelo. Este tipo de herramientas gráficas son aplicables tanto para muestras completas como para muestras censuradas. Consideraremos que T es una variable aleatoria *Weibull* (λ, β) con características (1.2), (1.3), (1.4) y (1.5). Si tomamos

logaritmos a ambos lados de la expresión (1.4) obtendremos

$$\log(R(t)) = \log(1 - F(t)) = -(\lambda t)^\beta,$$

podremos reescribir la expresión anterior como sigue, teniendo en cuenta las propiedades de los logaritmos,

$$\log\left(\frac{1}{1 - F(t)}\right) = (\lambda t)^\beta.$$

Volviendo a aplicar logaritmos a ambos lados de la expresión anterior

$$\log\left(\log\left(\frac{1}{1 - F(t)}\right)\right) = \log(\lambda t)^\beta,$$

o equivalentemente

$$\log\left(\log\left(\frac{1}{1 - F(t)}\right)\right) = \beta \log(\lambda) + \beta \log(t). \quad (3.1)$$

Teniendo en cuenta que

$$H(t) = -\log(1 - F(t)) = \log\left(\frac{1}{1 - F(t)}\right),$$

la expresión (3.1) quedaría

$$\log(H(t)) = \beta \log(\lambda) + \beta \log(t), \quad (3.2)$$

por lo que el logaritmo de la tasa de fallo acumulada $\log(H(t))$ del modelo Weibull se puede expresar como una recta en función de $\log t$ con pendiente β . Si la muestra es completa, se dispondrán de n tiempos de fallo y si la muestra presenta censura (en nuestro caso censura tipo II), solo dispondremos de r tiempos observados y $n - r$ tiempos censurados.

Un estudio completo sería un caso particular del estudio censurado que estamos considerando solo con tomar $r = n$. Supongamos entonces que se dispone de r tiempos T_1, T_2, \dots, T_r ordenados de menor a mayor

$$T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(r)}.$$

El estimador empírico de la función de distribución es

$$\hat{F}(t) = \frac{\text{n}^\circ \text{observaciones } \leq t}{n}, \quad (3.3)$$

con $t \leq T_{(r)}$. De la expresión (3.1) deducimos directamente el estimador de la tasa de fallo acumulada es

$$\hat{H}(t) = -\log(1 - \hat{F}(t)).$$

Dado que estamos considerando una muestra censurada, no podremos estimar $F(t)$ y $H(t)$ de forma completa, sino que solo podremos dar estimaciones basándonos en los r datos no censurados. Definimos ahora el concepto de rango de $T_{(i)}$ como el número de elementos de la muestra menores o iguales que $T_{(i)}$ y se denota $\text{rango}(T_{(i)})$. El estimador de la función de distribución empírico tiene expresiones muy sencillas si se evalúa en una muestra ordenada y en este caso es ¹

$$\hat{F}(T_{(i)}) = \frac{\text{rango}(T_{(i)})}{n}, \quad i = 1, \dots, r.$$

y el estimador de la tasa de fallo acumulada es

$$\hat{H}(T_{(i)}) = -\log(1 - \hat{F}(T_{(i)})), \quad i = 1, \dots, r.$$

Esta función representa el porcentaje empírico de fallos ocurridos ante el tiempo de fallo correspondiente al rango i -ésimo, por lo que la tasa de fallo acumulada en la muestra ordenada será

$$\hat{H}(T_{(i)}) = -\log\left(\frac{n-i}{n}\right), \quad i = 1, \dots, r.$$

Teniendo en cuenta las expresiones anteriores, el método gráfico de ajuste de la distribución Weibull consta de los siguientes pasos:

¹En la práctica, utilizaremos los valores $\hat{F}(T_{(i)}) = i/(n+1)$ ó $\hat{F}(T_{(i)}) = (i-0.3)/(n+0.4)$, ya que son los únicos valores que podremos utilizar con la función `weibplot`, de la que hablaremos en el capítulo correspondiente a software R.

1. Determinar si el estudio es completo (trabajaremos con n datos) o censurado (trabajaremos con r datos, los no censurados).
2. Considerar la muestra ordenada de los tiempos de fallo:
 - $T_{(1)}, T_{(2)}, \dots, T_{(n)}$ en el caso de que estemos ante un estudio completo
 - $T_{(1)}, T_{(2)}, \dots, T_{(r)}$ en el caso de que estemos ante un estudio censurado.

3. Calcular la función de distribución empírica evaluada en cada tiempo observado

$$\hat{F}(T_{(i)}) = \frac{\text{rango}(T_{(i)})}{n}, i = 1, \dots, n \ (i = 1, \dots, r).$$

En el caso de que no haya tiempos repetidos tendremos que $\text{rango}(T_{(i)})=i$.

4. Calcular la tasa de fallo acumulada empírica en cada uno de los tiempos observados

$$\hat{H}(T_{(i)}) = -\log\left(\frac{n-i}{n}\right), i = 1, \dots, n \ (i = 1, \dots, r)$$

5. Representar los pares de puntos $(\log(T_{(i)}), \log(\hat{H}(T_{(i)})))$, $i = 1, \dots, n$ ($i = 1, \dots, r$) en los ejes coordenados.
6. Calcular la recta de ajuste por mínimos cuadrados para comprobar la existencia de relación lineal. El ajuste de esta recta será adecuada para nuestro conjunto de tiempos de fallo si el coeficiente de determinación R^2 es próximo a 1. Si esto ocurre, el modelo Weibull será un modelo plausible para los tiempos de fallo.
7. Una vez que sabemos que nuestros datos proceden de una distribución Weibull, podremos estimar λ y β , parámetros de escala y forma, respectivamente. Para ello consideremos que la recta ajustada en el gráfico

es

$$y = \hat{a}t + \hat{b},$$

donde \hat{a} es la pendiente de la recta y \hat{b} es una constante. Volviendo a la relación (3.2), entonces

$$y = \hat{a}t + \hat{b} = \log[H(t)] = \beta \log(\lambda) + \beta \log(t),$$

y las estimaciones de los parámetros de la distribución Weibull son

$$\hat{\beta} = \hat{a},$$

con lo que la estimación del parámetro de forma, β , coincide con la pendiente de la recta ajustada y

$$\hat{\lambda} = \exp\left(\frac{\hat{b}}{\hat{a}}\right),$$

es la estimación del parámetro de escala.

Veamos un ejemplo como ilustración del método gráfico que acabamos de describir, recogido de Griful Ponsati (2004).

Ejemplo 1. *En una prueba de vida interesa el tiempo de perforación, en minutos, de un fluido aislante interpuesto entre dos electrodos. Los datos de tiempo son 0.74, 1.13, 0.09, 0.47, 0.73, 2.38, 1.4, 0.39. El experimento termina cuando fallan todas las unidades.*

Estamos ante un estudio completo, ya que sabemos que el estudio termina cuando han fallado todas las unidades. Lo primero que haremos será ordenar los tiempos de menor a mayor para a continuación seguir los pasos citados anteriormente para la construcción del gráfico Weibull. Para hacer este ejercicio utilizaremos la función `weibplot` de R, de la que hablaremos con detalle en el siguiente capítulo.

Tabla 3.1: Cálculos para la obtención del Gráfico Weibull

T_i	$Rango(T_{(i)})$	$\hat{F}(T_{(i)})$	$\log(T_{(i)})$	$\log\left(-\log\left(1 - \hat{F}(T_{(i)})\right)\right)$
0.09	1	0.11	-2.4079	-2.14
0.39	2	0.22	-0.9416	-1.38
0.47	3	0.33	-0.755	-0.92
0.73	4	0.44	-0.3147	-0.55
0.74	5	0.56	-0.3011	-0.2
1.13	6	0.67	0.1222	0.103
1.4	7	0.78	0.3365	0.41
2.38	8	0.89	0.8671	0.79

Los cálculos realizados para la obtención del gráfico están reflejados en la tabla(3.1), donde la estimación de la función de distribución empírica $\hat{F}(T_{(i)})$ se calcula utilizando el argumento 'exp' de la función `weibplot`, cuyo valor es $i/(n + 1)$. Valiéndonos de la función anterior, obtenemos la figura (3.1). La disposición de los puntos en el gráfico se ajusta a la línea recta, por lo que podremos asegurar que nuestro conjunto de datos sigue una distribución Weibull. Una vez que sabemos que nuestro conjunto de datos tiene dicha distribución, nos faltará estimar los parámetros de ésta. Para ello, tendremos que saber la expresión de la línea recta sobre la que caen los puntos, que en este caso es

$$y = 0.95 \log(t) - 0.081,$$

por lo que las estimaciones de los parámetros en este caso serán $\hat{\beta} = 0.95$ y $\hat{\lambda} = e^{-0.081/0.95} = 0.92$. Además, el coeficiente de determinación es $R^2 = 0.95$, lo que indica un buen ajuste, ya que es un valor próximo a uno.

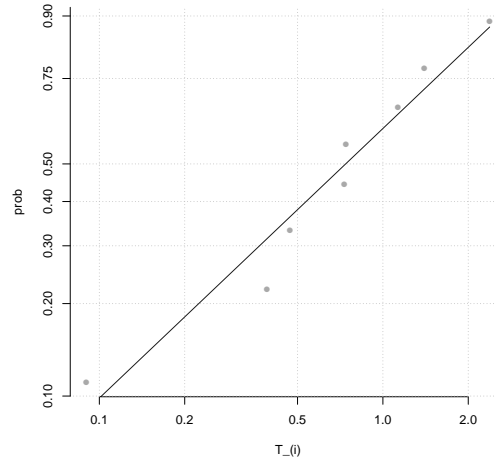


Figura 3.1: Gráfico Weibull con $\lambda = 0.92$ y $\beta = 0.95$.

3.2. Estimación por máxima verosimilitud

Otro método muy utilizado en la estimación de parámetros es el método de máxima verosimilitud. Este método es muy utilizado en la práctica debido a que proporciona estimadores consistentes, asintóticamente eficientes, insesgados y normalmente distribuidos en muestras grandes. Veremos la formulación del método de máxima verosimilitud tanto para muestras completas como para muestras con censura tipo II para la distribución $Exp(\lambda)$ y para la distribución $Weibull(\lambda, \beta)$.

Supongamos que T_1, T_2, \dots, T_n es una muestra aleatoria simple de n elementos y la muestra ordenada $T_{(1)}, T_{(2)}, \dots, T_{(n)}$, los cuales proceden de un determinado modelo de distribución con función de densidad $f(t; \theta)$, función de verosimilitud (1.7) y función de log-verosimilitud (1.8).

3.2.1. Máxima verosimilitud para la distribución Exponencial

Primero hablaremos de la función de máxima verosimilitud para el modelo exponencial $Exp(\lambda)$, que es un caso particular de la distribución Weibull y veremos que la estimación de su parámetro se puede obtener analíticamente. Sea $T \sim Exp(\lambda)$, con función de densidad (1.2). Por tanto, su función de verosimilitud para el caso completo es

$$L(\lambda) = \lambda^n \prod_{i=1}^n \exp\{-\lambda T_i\}.$$

El logaritmo de la expresión anterior es

$$l(\lambda) = n \log(\lambda) - \sum_{i=1}^n (\lambda T_i),$$

y ahora haciendo la derivada respecto a λ de esta expresión e igualando el resultado a cero obtenemos que el estimador de máxima verosimilitud de λ es

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n T_i}.$$

En el caso de que tengamos una muestra con censura tipo II y $T_i \sim Exp(\lambda)$, la función de verosimilitud será la dada en la expresión (2.1), donde r es el número de datos no censurados de la muestra. En el caso de la distribución $Exp(\lambda)$, la función de verosimilitud con censura tipo II es

$$L(\lambda) = \frac{n!}{(n-r)!} \lambda^r \prod_{i=1}^r \exp[-\lambda T_{(i)}] (\exp[-\lambda T_{(r)}])^{n-r},$$

con lo que la función de log-verosimilitud es

$$l(\lambda) = \log\left(\frac{n!}{(n-r)!}\right) + r \log(\lambda) - \lambda^r \sum_{i=1}^r r \lambda T_{(i)} - (n-r)T_{(r)}$$

y el estimador de máxima verosimilitud del parámetro λ se obtiene calculando la derivada de la anterior expresión respecto a λ e igualando a cero, por lo

que este estimador es

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^r T_{(i)} + (n-r)T_{(r)}}.$$

3.2.2. Máxima verosimilitud para la distribución Weibull

Una variable aleatoria T tiene distribución $Weibull(\lambda, \beta)$ si tiene densidad (1.2). En el caso del modelo $Weibull(\lambda, \beta)$, no podremos calcular analíticamente las estimaciones mediante máxima verosimilitud de los parámetros del modelo, sino que deberemos recurrir a métodos numéricos para calcular dichas estimaciones. La función de verosimilitud para muestras completas con distribución $Weibull(\lambda, \beta)$ es

$$L(\lambda, \beta) = (\lambda\beta)^n \prod_{i=1}^n (\lambda T_i)^{\beta-1} \exp[-(\lambda T_i)]$$

y la función de log-verosimilitud es

$$l(\lambda, \beta) = n \log(\beta) + \beta n \log \lambda + (\beta - 1) \sum_{i=1}^n \log T_i - \sum_{i=1}^n (\lambda T_i).$$

Los estimadores de máxima verosimilitud se obtienen resolviendo las ecuaciones resultantes de igualar las dos derivadas parciales de $l(\lambda, \beta)$ a cero. Como resultado, el estimador de máxima verosimilitud de β , se obtiene resolviendo la ecuación

$$\frac{\sum_{i=1}^n T_i^{\hat{\beta}} \log T_i}{\sum_{i=1}^n T_i^{\hat{\beta}}} - \frac{1}{\hat{\beta}} - \frac{1}{n} \sum_{i=1}^n \log T_i = 0, \quad (3.4)$$

y aunque no es posible obtener una solución analíticamente, podremos calcular $\hat{\beta}$ utilizando métodos numéricos como el método de Newton-Raphson. Una vez estimado el parámetro de forma, obtendremos una estimación para el parámetro de escala λ mediante la expresión

$$\hat{\lambda} = \left(\frac{n}{\sum_{i=1}^n T_i^{\hat{\beta}}} \right)^{\frac{1}{\hat{\beta}}}. \quad (3.5)$$

La función de verosimilitud para muestras con censura tipo II de una distribución *Weibull*(λ, β) es la dada por la expresión (2.1). Entonces, su función de log-verosimilitud es

$$\begin{aligned} l(\lambda, \beta) &= \log n! - \log(n-r)! + r \log(\lambda) \\ &\quad + (\beta - 1) \sum_{i=1}^r \log(\lambda T_{(i)}) \\ &\quad - \sum_{i=1}^r (\lambda T_{(i)})^\beta - (n-r)(\lambda T_{(r)})^\beta. \end{aligned}$$

El estimador de máxima verosimilitud del parámetro de forma β se obtiene resolviendo la ecuación

$$\frac{\sum_{i=1}^r T_{(i)}^{\hat{\beta}} \log T_{(i)} + (n-r)T_{(r)}^{\hat{\beta}} \log(T_{(r)})}{\sum_{i=1}^r T_{(i)}^{\hat{\beta}} + (n-r)T_{(r)}^{\hat{\beta}}} - \frac{1}{\hat{\beta}} - \frac{1}{r} \sum_{i=1}^r \log T_{(i)} = 0 \quad (3.6)$$

y utilizando esta última, se obtiene que el estimador de máxima verosimilitud del parámetro escala λ es

$$\hat{\lambda} = \left(\frac{r}{\sum_{i=1}^r T_{(i)}^{\hat{\beta}} + (n-r)T_{(r)}^{\hat{\beta}}} \right)^{\frac{1}{\hat{\beta}}}.$$

Es necesario utilizar métodos iterativos como el Newton-Raphson para obtener las anteriores estimaciones.

3.3. Intervalos de confianza

Supongamos que T es una variable aleatoria con distribución *Weibull*(λ, β), entonces los estimadores de máxima verosimilitud de los parámetros de la distribución, $\hat{\lambda}$ y $\hat{\beta}$, son los obtenidos en las expresiones (3.4) y (3.5), respectivamente. Una característica salientable de los estimadores de máxima verosimilitud es que son asintóticamente normales, es decir, que si $\hat{\lambda}$ y $\hat{\beta}$ son los estimadores de máxima verosimilitud de los parámetros escala y forma

de la distribución $Weibull(\lambda, \beta)$, respectivamente, entonces

$$\hat{\lambda} \sim_{aprox} N\left(\lambda, \sqrt{Var(\hat{\lambda})}\right)$$

y

$$\hat{\beta} \sim_{aprox} N\left(\beta, \sqrt{Var(\hat{\beta})}\right),$$

donde $\sqrt{Var(\hat{\lambda})}$ y $\sqrt{Var(\hat{\beta})}$ son los elementos diagonales de la matriz inversa de la matriz de información de Fisher evaluada en $\hat{\lambda}$ y $\hat{\beta}$. Por tanto, las cantidades anteriores estandarizadas son

$$\frac{\hat{\lambda} - \lambda}{\sqrt{Var(\hat{\lambda})}} \sim_{aprox} N(0, 1) \quad (3.7)$$

y

$$\frac{\hat{\beta} - \beta}{\sqrt{Var(\hat{\beta})}} \sim_{aprox} N(0, 1). \quad (3.8)$$

Los intervalos de confianza $(1 - \alpha)\%$ bilaterales basados en los pivotes (3.7) y (3.8) son de la forma

$$\left(\hat{\lambda} - z_{\frac{\alpha}{2}} \sqrt{Var(\hat{\lambda})}, \hat{\lambda} + z_{\frac{\alpha}{2}} \sqrt{Var(\hat{\lambda})}\right)$$

y

$$\left(\hat{\beta} - z_{\frac{\alpha}{2}} \sqrt{Var(\hat{\beta})}, \hat{\beta} + z_{\frac{\alpha}{2}} \sqrt{Var(\hat{\beta})}\right),$$

respectivamente y donde $z_{\frac{\alpha}{2}}$ es el cuantil de la distribución normal estándar, que está tabulado y $\sqrt{Var(\hat{\lambda})}$ y $\sqrt{Var(\hat{\beta})}$ son las raíces cuadradas de las estimaciones de los elementos diagonales de la matriz (1.9).

3.4. Contrastes de hipótesis basados en el test de razón de verosimilitudes

Al realizar un contraste de hipótesis, la idea básica es trabajar con nuestra muestra suponiendo que la hipótesis nula H_0 es cierta. Si la conclusión es

que hay una gran diferencia entre la muestra y esta misma bajo H_0 es que rechazaremos H_0 . Con esto se sigue que hay dos tipos de errores, rechazar H_0 cuando es cierta, al que llamaríamos error tipo I, y aceptar H_0 cuando es falsa, al que llamaríamos error tipo II. El nivel de significación, denotado por α (suele utilizarse $\alpha = 0.05$ ó $\alpha = 0.01$), es la máxima probabilidad de cometer error de tipo I y la potencia del test es la probabilidad de no cometer error de tipo II. Estos contrastes están basados en determinados estadísticos, que resumen las características del conjunto de datos a estudio. Los contrastes comparan estos estadísticos con los valores de los estadísticos tabulados, que dependen del nivel de significación y si los estadísticos calculados superan a los tabulados rechazaremos H_0 . Por último, definiremos el p-valor, que es la probabilidad calculada al asumir que la hipótesis nula H_0 es cierta, de que el estadístico de contraste tome valores tan o más extremos que los calculados con la muestra sin restricciones. Si esta probabilidad es menor que nuestro nivel de significación, rechazaremos H_0 .

Por otra parte, el test de razón de verosimilitudes es un contraste de bondad de ajuste entre dos modelos y que nos sirve de herramienta para comprobar si un modelo reducido proporciona la misma información que el modelo sin restricciones. El estadístico del test de razón de verosimilitudes es

$$\Lambda = -2l(\tilde{\theta}) + 2l(\hat{\theta}),$$

donde $l(\tilde{\theta})$ es la función de log-verosimilitud del modelo reducido y $l(\hat{\theta})$ es la función de log-verosimilitud para el modelo sin restricciones. Ahora veremos cómo se comporta el test de razón de verosimilitudes con los parámetros de la distribución Weibull tanto para muestras completas como para muestras ordenadas.

3.4.1. Test de razón de verosimilitudes para el parámetro de forma β

Consideremos la hipótesis nula

$$H_0 : \beta = \beta_0,$$

frente a la cual se pueden presentar distintas hipótesis alternativas: alternativa bilateral

$$H_1 : \beta \neq \beta_0,$$

o bien alternativas unilaterales izquierda y derecha

$$H_1 : \beta < \beta_0 \quad \text{ó} \quad H_1 : \beta > \beta_0$$

Para un contraste de razón de verosimilitud de la hipótesis nula, H_0 , frente a la hipótesis alternativa, H_1 , necesitaremos el estimador de máxima verosimilitud de λ y β bajo H_0 . El estimador de máxima verosimilitud de β bajo H_0 es $\tilde{\beta} = \beta_0$, y el estimador de máxima verosimilitud de λ bajo H_0 , $\tilde{\lambda}$, se obtiene al calcular la derivada parcial de $l(\lambda, \beta_0)$ respecto a λ e igualándola a cero, entonces

$$\tilde{\lambda} = \left(\frac{n}{\sum_{i=1}^n T_i^{\beta_0}} \right)^{\frac{1}{\beta_0}}.$$

El estadístico del test de razón de verosimilitudes para contrastar H_0 frente a H_1 es entonces

$$\Lambda = -2l(\tilde{\lambda}, \beta_0) + 2l(\hat{\lambda}, \hat{\beta}),$$

donde $\hat{\lambda}$ y $\hat{\beta}$ son los estimadores de máxima verosimilitud del modelo sin restricciones y que corresponden a las expresiones (3.5) y (3.4) respectivamente. Las zonas de rechazo dependen de la hipótesis alternativa considerada. Por ejemplo, en el caso del contraste unilateral derecho, rechazaremos H_0 para valores grandes de Λ , teniendo en cuenta que Λ tiene una distribución

aproximada χ_1^2 bajo H_0 . Para muestras con censura tipo II, tenemos que

$$\tilde{\lambda} = \left(\frac{r}{\sum_{i=1}^r T_i^{\beta_0} + (n-r)T_{(r)}^{\beta_0}} \right)^{\frac{1}{\beta_0}},$$

y una vez obtenido este valor, procederemos como en el caso de muestras completas.

3.4.2. Test de razón de verosimilitudes para el parámetro escala λ

Consideremos las hipótesis nula

$$H_0 : \lambda = \lambda_0,$$

frente a la cual se pueden presentar distintas hipótesis alternativas: alternativa bilateral

$$H_1 : \lambda \neq \lambda_0,$$

o bien alternativas unilaterales izquierda y derecha

$$H_1 : \lambda < \lambda_0 \quad \text{ó} \quad H_1 : \lambda > \lambda_0$$

Para contrastar H_0 frente a H_1 tendremos que maximizar $l(\lambda, \beta)$ bajo H_0 . Calculando la derivada parcial de $l(\lambda_0, \beta)$ respecto a β e igualando esa expresión a cero tendremos la igualdad

$$\frac{n}{\beta} + n \log(\lambda_0) + \sum_{i=1}^n \log(T_i) - \sum_{i=1}^n (\lambda_0 T_i)^\beta \log(\lambda_0 T_i) = 0.$$

Esta ecuación puede resolverse utilizando métodos iterativos que nos proporcionarán el estimador de máxima verosimilitud de β bajo H_0 . El estadístico del test de razón de verosimilitudes para contrastar H_0 frente a H_1 es entonces

$$\Lambda = -2l(\lambda_0, \tilde{\beta}) + 2l(\hat{\lambda}, \hat{\beta}),$$

donde $\hat{\lambda}$ y $\hat{\beta}$ son los estimadores de máxima verosimilitud del modelo sin restricciones y que se corresponden con las expresiones (3.5) y (3.4) respectivamente. Las zonas de rechazo dependen de la hipótesis alternativa considerada. Por ejemplo, en el caso del contraste unilateral derecho, rechazaremos H_0 para valores grandes de Λ , teniendo en cuenta que Λ tiene una distribución aproximada χ_1^2 bajo H_0 . Para muestras con censura tipo II, obtendremos $\tilde{\beta}$ resolviendo la ecuación

$$\frac{r}{\beta} + \sum_{i=1}^r \log(\lambda_0 T_{(i)}) (1 - (\lambda_0 T_{(i)})^\beta) - (n - r) \log(\lambda_0 T_{(r)}) (\lambda_0 T_{(r)})^\beta = 0.$$

Para resolver la ecuación anterior utilizaremos métodos iterativos. Una vez obtenido este valor, procederemos como en el caso de muestras completas.

3.5. Pruebas de bondad de ajuste

Una prueba de bondad de ajuste trata de contrastar la hipótesis nula H_0 : El conjunto de datos sigue un modelo con función de distribución $F_0(t)$.

Se pueden considerar los siguientes casos

1. El parámetro (o parámetros de la distribución) es conocido. Se trata entonces de una hipótesis nula simple.
2. El parámetro (o parámetros) no es conocido y es necesario obtener su estimación. Se trata entonces de una hipótesis nula compuesta.

Consideraremos tres contrastes diferentes basados en la función de distribución empírica.

3.5.1. Contraste de Kolmogorov-Smirnov

El estadístico en el que se basa este contraste, D_n es

$$D_n = \sup_t |\hat{F}(t) - F_0(t)|,$$

donde $\hat{F}(t)$ es la función de distribución empírica, la cual viene dada por la expresión (3.3). Los valores críticos están tabulados y rechazaremos H_0 para valores grandes del estadístico D_n . Como este contraste está basado en el supremo, puede causar algún problema, por lo que consideramos los siguientes contrastes.

3.5.2. Contraste de Cramér-von Misses

Es un contraste que se basa en las diferencias entre F_0 y la función de distribución empírica (3.3). El estadístico de contraste es

$$W_n^2 = n \int_{-\infty}^{\infty} (\hat{F}(t) - F_0(t))^2 dF_0(t).$$

Si trabajamos con muestras ordenadas, el estadístico de contraste es

$$W_n^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^n \left[F_0(T_{(i)}) - \frac{2i-1}{2n} \right]^2.$$

Los valores críticos están tabulados y rechazaremos H_0 para valores grandes del estadístico W_n^2 .

3.5.3. Contraste de Anderson-Darling

Este contraste también está basado en las diferencias entre F_0 y la función de distribución empírica (3.3). Es una generalización del contraste de Cramér-von Misses. El estadístico de este contraste es

$$A_n^2 = \int_{-\infty}^{\infty} \frac{(\hat{F}(t) - F_0(t))^2}{F_0(t)(1 - F_0(t))} dF_0(t).$$

Los valores críticos están tabulados y rechazaremos H_0 para valores grandes del estadístico A_n^2 .

Estos tres contrastes solo son válidos cuando trabajamos con muestras completas. Para trabajar con datos censurados se pueden obtener los valores críticos utilizando técnicas de remuestreo.

Capítulo 4

Software R

Antes de empezar a describir los paquetes y las funciones relacionadas con la fiabilidad industrial disponibles en el software R hablaremos de las funciones relacionadas directamente con la distribución Weibull: `dweibull` proporciona la función de densidad, `pweibull` nos da la función de distribución, `qweibull` nos da el cuantil y `rweibull` genera números aleatorios con distribución Weibull con parámetros especificados. Estas funciones se pueden encontrar en el paquete `stats`. Tenemos que tener en cuenta que R utiliza como parámetro de escala para la distribución $Weibull(\lambda, \beta)$, $\alpha = \lambda^{-1}$ y que en todas las funciones relacionadas con esta distribución, se escribe primero el parámetro de forma β y a continuación es parámetro de escala λ^{-1} , que difiere de nuestra notación. En cada sección utilizaremos como ejemplo el conjunto de datos extraídos de la librería `FAmle`, `yarns`, (véase Gamerman y Lopes (2006)), en el que se miden cien ciclos hasta el tiempo de fallo en cristales de aviones. Para cargar los datos y poder trabajar con ellos en R utilizaremos el código

```
library(FAmle)
data(yarns)
datos=yarns$x
```

4.1. Paquete Renext

Antes de cargar este paquete tendremos que tener instalados en R otros paquetes como `survival`, `mgvc` o `splines`. El paquete `Renext`, mediante la función `fweibull`, nos permitirá calcular las estimaciones por máxima verosimilitud de los parámetros de la distribución $Weibull(\lambda, \beta)$ solo para muestras completas. Mediante la función `weibplot`, podremos representar un conjunto de datos sobre escalas logarítmicas, dibujando los pares de puntos $(\log(t), -\log(1 - F(t)))$. La sintaxis de la función `fweibull` es

```
fweibull(x, info.observed = FALSE, check.loglik = FALSE)
```

donde los argumentos son

<code>x</code>	vector de datos
<code>info.observed</code>	matriz de información observada
<code>check.loglik</code>	si es TRUE, la log-verosimilitud es recalculada utilizando <code>dweibull</code> con <code>log=TRUE</code> . El resultado es una lista.

Los valores que devuelve esta función son

<code>estimate</code>	estimación por máxima verosimilitud de los parámetros.
<code>sd</code>	desviación estándar(asintótica) de las estimaciones.
<code>cov</code>	matriz de covarianzas(asintótica) calculada a partir de la matriz de información teórica u observada.

Ejemplo 2. *Cálculo de las estimaciones por máxima verosimilitud para el conjunto de datos a estudio.*

```

library(Renext)
fweibull(datos)
$estimate
      shape      scale
1.604368 247.915708
$sd
      shape      scale
0.125092 16.270475
$loglik
[1] -625.199
$cov
      [,1]      [,2]
[1,] 0.01564802 0.6371975
[2,] 0.63719754 264.7283432

```

La sintaxis de la función `weibplot` es

```
weibplot(x, plot.pos= "exp", shape= NULL, scale= NULL,
labels= NULL, mono= TRUE, ...)
```

donde los argumentos son

<code>x</code>	vector de datos
<code>plot.pos</code>	posición respecto a la que se van a representar los puntos. Si es 'exp' entonces $F(x)$ en el punto i se toma como $i/n+1$ y si es 'med' se toma como $(i-0.3)/(n+0.4)$.
<code>shape</code>	parámetro forma para una o más líneas que queremos representar (puede ser un vector o un número).

<code>scale</code>	parámetro escala para una o más líneas que queramos representar (puede ser un vector o un número).
<code>labels</code>	leyenda del gráfico.
<code>mono</code>	para que el gráfico tenga color (<code>mono= FALSE</code>) o en escala de grises (<code>mono= TRUE</code>).
<code>...</code>	otros argumentos.

A continuación construiremos el gráfico Weibull para nuestro conjunto de datos valiéndonos de las estimaciones por máxima verosimilitud de los parámetros de la distribución $Weibull(\lambda, \beta)$ calculados utilizando la función `fweibull`. En este caso $1/\hat{\lambda} = 247.9$ y $\hat{\beta} = 1.6$.

Ejemplo 3. *Obtención del gráfico Weibull para el conjunto de datos a estudio.*

```
library(Renext)
weibplot(datos, shape=1.6, scale=247.9)
```

Y de este gráfico podemos deducir que los datos siguen una distribución Weibull y que por tanto la estimación del parámetro de escala es $\hat{\lambda} = 0.004$ y la del parámetro de forma es $\hat{\beta} = 1.6$.

4.2. Paquete STAR

Antes de cargar este paquete tendremos que tener instalados en R otros paquetes como `survival`, `mgvc` o `splines`. Este paquete nos permitirá calcular mediante la función `weibullMLE` las estimaciones por el método de máxima verosimilitud de los parámetros de la distribución $Weibull(\lambda, \beta)$ y nos permitirá hacer estas estimaciones tanto para muestras completas como para muestras con censura tipo II. Su sintaxis es

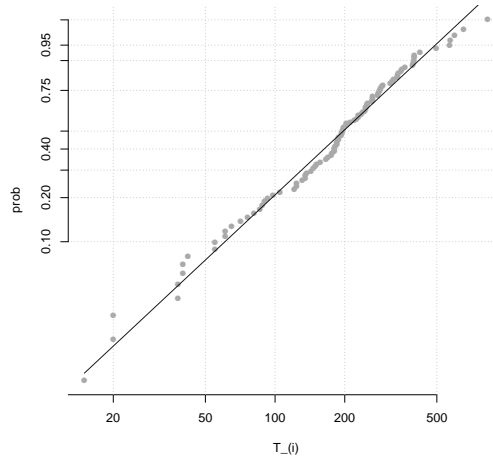


Figura 4.1: Gráfico Weibull

```
weibullMLE(yi,ni=numeric(length(yi))+1, si=numeric(length(yi))+1,
shape.min= 0.05, shape.max= 5)
```

y sus argumentos son

<code>yi</code>	vector de observaciones
<code>ni</code>	vector que cuenta cada valor de <code>yi</code> . Por defecto es <code>numeric(length(yi))+1</code> .
<code>si</code>	vector de datos no censurados. Mirar comentarios.
<code>shape.min</code>	el menor valor posible para el parámetro forma.
<code>shape.max</code>	el mayor valor posible para el parámetro forma.

Comentarios:

En el caso de muestras completas se definirá `si` como `numeric(length(yi))+1`,

que es el valor por defecto, pero en el caso de muestras censuradas (en nuestro caso censura tipo II), reordenaremos la muestra con el comando `sort` y definiremos `si` como `c(rep(1,r),rep(0,n-r))`, por lo que tendremos una muestra ordenada con r datos no censurados y $n - r$ datos censurados. Los valores que devuelve esta función son

<code>estimate</code>	los parámetros estimados.
<code>se</code>	errores estándar de cada parámetro.
<code>logLik</code>	el máximo del logaritmo de la función de máxima verosimilitud.
<code>r</code>	código R de la función que devuelve el logaritmo de la función de verosimilitud relativa.
<code>mll</code>	código R de la función que devuelve el opuesto de la función de log-verosimilitud utilizando los logaritmos de los parámetros.

En cuanto a la aplicación de esta función con muestras que presentan censura tipo II, hemos detectado que si el porcentaje de datos censurados es elevado, dicha función da errores y no calcula las estimaciones. Si la muestra es completa funciona correctamente.

Ejemplo 4. *Cálculo de las estimaciones de máxima verosimilitud para el conjunto de datos a estudio.*

```
library(STAR)
```

```
weibullMLE(datos,shape.min=1,shape.max=5)
$estimate
      shape      scale
1.604369 247.915477
$se
      shape      scale
0.1218647 16.2860248
$logLik
[1] -625.199
```

4.3. Paquete Sim.DiffProc

Antes de cargar este paquete tendremos que tener instalados en R los paquetes `rgl`, `xlsx`, `xlsx` y `rJava`. Dentro de este paquete podremos encontrar la función `Ajdweibull`, que nos proporcionará intervalos de confianza para el modelo $Weibull(\lambda, \beta)$ cuando trabajamos con muestras de datos no censuradas y la función `testksdweibull`, que es el contraste Kolmogorov-Smirnov aplicado a la distribución Weibull. La sintaxis para `Ajdweibull` es

```
Ajdweibull(x, starts= list(shape= 1, scale= 1), leve= 0.95)
```

donde sus argumentos son

```
x          vector de datos observados.
starts     lista con los valores iniciales
           de los parámetros de la
           distribución para su estimación.
leve       nivel de confianza.
```

Los valores que devuelva esta función son

```
coef       Coeficientes extraídos del modelo.
```

AIC	Valor numérico correspondiente con el criterio AIC.
vcov	Matriz de de varianzas-covarianzas estimada.
confint	Una matriz(o vector) donde las columnas proporcionan los límites de confianza superior e inferior para cada parámetro. Éstos son $(1-\text{leve})/2$ y $1 - (1-\text{leve})/2$ respectivamente.

Ejemplo 5. *Obtención de los intervalos de confianza basados en el test de razón de verosimilitudes para los parámetros de la distribución del conjunto de datos a estudio.*

```
library(Sim.DiffProc)
Ajdweibull(datos$yarns, starts=list(shape=1, scale=1), leve=0.95)
Coefficients:
      Estimate Std. Error
shape 5.809887 0.31892806
scale 5.420565 0.06978147
-2 log L: 556.7529
$coef
      shape      scale
5.809887 5.420565
$AIC
[1] 560.7529
$vcov
      shape      scale
shape 0.101715109 0.007086121
```



```
scale 0.007086121 0.004869454
$confint
      2.5 %   97.5 %
shape 5.201176 6.451196
scale 5.283088 5.557944
```

La función `testksdweibull`, que proporciona el test de Kolmogorov-Smirnov para contrastar si un conjunto de datos sigue o no un modelo de distribución $Weibull(\lambda, \beta)$. Su sintaxis es

```
test_ks_dweibull(x, shape, scale)
```

donde los argumentos son

<code>x</code>	vector de datos
<code>shape</code>	parámetro de forma. Por defecto es 1.
<code>scale</code>	parámetro de escala. Por defecto es 1.

Los valores que devuelve son

<code>statistic</code>	el valor del estadístico del contraste.
<code>p.value</code>	el p-valor del test
<code>alternative</code>	describe la hipótesis alternativa.
<code>data.name</code>	vector de caracteres queda el nombre del conjunto de datos a estudio.

Ejemplo 6. *Bondad de ajuste del conjunto `yarns` a la distribución $Weibull(1.6, 1/247.9)$.*

```
library(Sim.DiffProc)
test_ks_dweibull(datos,1.6,247.9)
  One-sample Kolmogorov-Smirnov test
data:  X
D = 0.0761, p-value = 0.6092.
```

Donde el p-valor es 0.5813, valor que supera el nivel de significación, 0.05. Por tanto, podremos afirmar que nuestro conjunto de datos sigue una distribución *Weibull*(1.6, 0.004).

Capítulo 5

Simulaciones y aplicación a datos reales

5.1. Estudio de simulación

En esta sección realizaremos un estudio de simulación sobre los intervalos de confianza obtenidos a partir de la normalidad asintótica de los estimadores de máxima verosimilitud de los parámetros de la distribución $Weibull(\lambda, \beta)$. Calcularemos la longitud, que es la diferencia entre extremo superior y el extremo inferior del correspondiente intervalo, y la cobertura, que es el número de veces que el parámetro real de la distribución cae dentro del intervalo. Para ello, fijaremos en 1000 el número de iteraciones de nuestro estudio. A continuación, generamos muestras Weibull de tamaño n (con n igual 25, 50 y 100), con r datos no censurados (consideraremos $r = 0.75n$ y $r = 0.50n$, redondeando al entero más próximo si fuese necesario), y elegimos los valores $\lambda = 1$ y $\beta = 0.5, 1$ y 2 para los parámetros de la Weibull. A continuación, calculamos la matriz de varianzas-covarianzas y mediante las raíces cuadradas de los elementos diagonales obtendremos los errores estándar de los parámetros λ y β , respectivamente. Después de este proceso, hemos obtenido los

resultados que se muestran en las tablas (5.1) y (5.2). Hemos visto en el capítulo 3, que los intervalos de confianza $(1 - \alpha)\%$ bilaterales basados en los pivotes (3.7) y (3.8) son de la forma

$$\left(\hat{\lambda} - z_{\frac{\alpha}{2}} \sqrt{\widehat{Var}(\hat{\lambda})}, \hat{\lambda} + z_{\frac{\alpha}{2}} \sqrt{\widehat{Var}(\hat{\lambda})} \right)$$

y

$$\left(\hat{\beta} - z_{\frac{\alpha}{2}} \sqrt{\widehat{Var}(\hat{\beta})}, \hat{\beta} + z_{\frac{\alpha}{2}} \sqrt{\widehat{Var}(\hat{\beta})} \right),$$

donde $\sqrt{\widehat{Var}(\hat{\lambda})}$ y $\sqrt{\widehat{Var}(\hat{\beta})}$ son las estimaciones de las raíces cuadradas de los elementos diagonales de la inversa de la matriz de información de Fisher, que viene dada por

$$I = \begin{pmatrix} -\frac{\partial^2 l(\lambda, \beta)}{\partial \lambda^2} & -\frac{\partial^2 l(\lambda, \beta)}{\partial \lambda \partial \beta} \\ -\frac{\partial^2 l(\lambda, \beta)}{\partial \beta \partial \lambda} & -\frac{\partial^2 l(\lambda, \beta)}{\partial \beta^2} \end{pmatrix}_{|(\lambda, \beta) = (\hat{\lambda}, \hat{\beta})},$$

donde las respectivas derivadas parciales para muestras con censura tipo II son (para muestras completas cambiamos r por n)

$$-\frac{\partial^2 l(\lambda, \beta)}{\partial \lambda^2} = \frac{\beta}{\lambda^2} \left[r + \lambda^\beta (\beta - 1) \left(\sum_{i=1}^r T_{(i)}^\beta + (n - r) T_{(r)}^\beta \right) \right].$$

Las expresiones de las derivadas parciales cruzadas coinciden y su expresión es

$$\begin{aligned} -\frac{\partial^2 l(\lambda, \beta)}{\partial \lambda \partial \beta} &= \frac{-r}{\lambda} + \sum_{i=1}^r (\lambda T_{(i)})^\beta [\lambda^{\beta-1} + \beta \lambda^{\beta-1} \log \lambda] \\ &+ \beta \lambda^{\beta-1} \sum_{i=1}^r \log T_{(i)} T_{(i)}^\beta + \frac{(n-r)}{\lambda} [(\lambda T_{(r)})^\beta \\ &+ \beta (\lambda T_{(r)})^\beta \log(\lambda T_{(r)}) + \beta \lambda^{\beta-1}]. \end{aligned}$$

Por último

$$-\frac{\partial^2 l(\lambda, \beta)}{\partial \beta^2} = \frac{r}{\beta^2} + \sum_{i=1}^r \log^2(\lambda T_{(i)}) (\lambda T_{(i)})^\beta + (n-r) \log^2(\lambda T_{(r)}) (\lambda T_{(r)})^\beta.$$

Después de obtener estas derivadas parciales, sustituiremos λ por $\hat{\lambda}$ y β por $\hat{\beta}$ y calcularemos la matriz inversa, que sería la matriz de varianzas-covarianzas

$$V = \begin{pmatrix} \widehat{Var}(\hat{\lambda}) & \widehat{Cov}(\hat{\lambda}, \hat{\beta}) \\ \widehat{Cov}(\hat{\beta}, \hat{\lambda}) & \widehat{Var}(\hat{\beta}) \end{pmatrix}.$$

Para obtener el errores estándar de $\hat{\lambda}$ y $\hat{\beta}$, calculamos la raíces cuadradas de los elementos diagonales de esta última matriz. A la vista de las tablas (5.1) y (5.2) podemos concluir lo siguiente:

- Las coberturas son buenas, se acercan a los niveles nominales, tanto para muestras completas como para muestras censuradas.
- A medida que aumenta n , las longitudes de los intervalos se van haciendo más pequeñas. Para un mismo n , las longitudes de los respectivos intervalos van creciendo a medida que aumenta el número de datos censurados.

Tabla 5.1: Longitud media y coberturas estimadas de los intervalos de confianza para β .

β	n	$1 - \alpha :$		0.950		0.900	
		r	cob.	long.	cob.	long.	
0.5	25	25	0.948	0.326	0.930	0.274	
		19	0.965	0.426	0.906	0.357	
		13	0.954	0.584	0.891	0.490	
	50	50	0.941	0.224	0.901	0.188	
		38	0.953	0.290	0.907	0.243	
		25	0.963	0.390	0.912	0.327	
	100	100	0.952	0.155	0.899	0.130	
		75	0.954	0.202	0.911	0.169	
		50	0.952	0.265	0.898	0.223	
1	25	25	0.949	1.313	0.906	1.102	
		19	0.955	1.703	0.900	1.429	
		13	0.961	2.358	0.891	1.979	
	50	50	0.956	0.447	0.918	0.376	
		38	0.946	0.582	0.905	0.489	
		25	0.941	0.780	0.891	0.655	
	100	100	0.950	0.311	0.905	0.261	
		75	0.940	0.404	0.895	0.339	
		50	0.952	0.532	0.908	0.447	
2	25	25	0.956	1.300	0.900	1.091	
		19	0.945	1.716	0.901	1.440	
		13	0.968	2.334	0.913	1.959	
	50	50	0.953	0.900	0.907	0.755	
		38	0.957	1.160	0.896	0.973	
		25	0.939	1.572	0.885	1.319	
	100	100	0.949	0.620	0.902	0.520	
		75	0.958	0.809	0.910	0.679	
		50	0.944	1.061	0.892	0.891	

Tabla 5.2: Longitudes medias y coberturas estimadas para los intervalos de confianza para λ .

β	n	$1 - \alpha :$		0.950		0.900	
		r	cob.	long.	cob.	long.	
0.5	25	25	0.890	1.839	0.891	1.543	
		19	0.933	2.158	0.884	1.811	
		13	0.910	2.980	0.855	2.501	
	50	50	0.937	1.245	0.896	1.045	
		38	0.944	1.342	0.910	1.126	
		25	0.938	1.970	0.897	1.653	
	100	100	0.945	0.862	0.901	0.723	
		75	0.942	0.946	0.902	0.794	
		50	0.936	1.333	0.892	1.119	
1	25	25	0.935	0.410	0.878	0.344	
		19	0.934	0.441	0.883	0.370	
		13	0.900	0.571	0.836	0.480	
	50	50	0.954	0.588	0.907	0.494	
		38	0.927	0.634	0.875	0.532	
		25	0.940	0.894	0.888	0.751	
	100	100	0.932	0.415	0.886	0.348	
		75	0.940	0.453	0.882	0.380	
		50	0.951	0.624	0.896	0.524	
2	25	25	0.931	0.409	0.879	0.343	
		19	0.919	0.443	0.871	0.371	
		13	0.898	0.571	0.842	0.479	
	50	50	0.946	0.289	0.897	0.243	
		38	0.931	0.315	0.882	0.264	
		25	0.910	0.429	0.857	0.360	
	100	100	0.948	0.206	0.900	0.173	
		75	0.948	0.225	0.898	0.189	
		50	0.940	0.308	0.876	0.259	

5.2. Aplicación a datos reales

Los datos que analizaremos en esta sección proceden de una empresa de la automoción y han sido estudiados con más detalle en Hermida (2007). Obtendremos los estimadores de máxima verosimilitud y los intervalos de confianza obtenidos a partir de normalidad asintótica.

Pieza 1

En este caso, tenemos un vector de tiempos de fallo en el que el porcentaje de censura es enorme comparado con el de la no censura. En primer lugar, construiremos el gráfico Weibull correspondiente para ver si nuestro conjunto de datos sigue o no una distribución Weibull. El gráfico aparece en la figura (5.1). Viendo el gráfico, parece que los datos se pueden ajustar mediante una

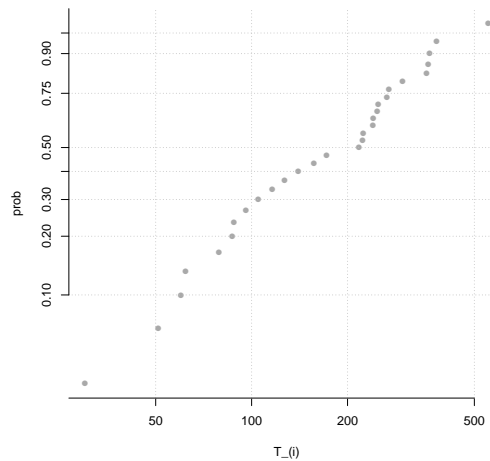


Figura 5.1: Gráfico Weibull para pieza 1.

recta, por lo que podemos afirmar que nuestro conjunto sigue una distribución Weibull. Los parámetros estimados por el método de máxima verosimilitud de λ y β son $3.58 \cdot 10^{-8}$ y 0.82, respectivamente. Ahora, nos queda calcular

Tabla 5.3: Intervalos de confianza para λ y β .

	$1 - \alpha = 0.95$	$1 - \alpha = 0.90$
β	(0.52, 1.12)	(0.57, 1.07)
λ	(0, $1.78 \cdot 10^{-7}$)	(0, $1.55 \cdot 10^{-7}$)

los intervalos de confianza, que en este caso son los correspondientes a la tabla (5.3).

Observamos que los intervalos correspondientes a β contienen al uno, por lo que no podemos concluir si en nuestro conjunto de datos se ha producido un fallo de fabricación, fenómenos aleatorios o envejecimiento. La longitud de los intervalos de confianza correspondientes a λ es muy pequeña.

Pieza 2

Como en el apartado anterior, disponemos de un vector de datos donde el porcentaje de censura es enorme comparado con el de la no censura. Como estamos trabajando con tiempos de vida, solo consideraremos aquellos datos que sean estrictamente mayores que cero. En primer lugar, construiremos el gráfico Weibull correspondiente para ver si nuestro conjunto de datos sigue o no una distribución Weibull. El gráfico es el que aparece en la figura (5.2).

Parece que los datos se pueden ajustar mediante una recta, por lo que podemos afirmar que nuestro conjunto sigue una distribución Weibull. Ahora nos falta estimar los parámetros de la distribución. Los parámetros estimados por el método de máxima verosimilitud de λ y β son $2.90 \cdot 10^{-8}$ y 0.78, respectivamente. Ahora, nos queda calcular los intervalos de confianza, que en este caso son los correspondientes a la tabla (5.4).

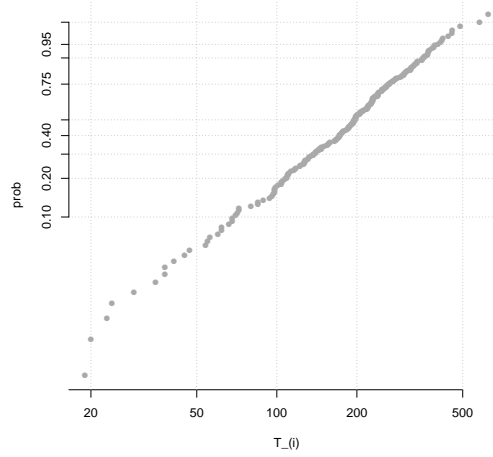


Figura 5.2: Gráfico Weibull para pieza 2.

Tabla 5.4: Intervalos de confianza para λ y β .

	$1 - \alpha = 0.95$	$1 - \alpha = 0.90$
β	(0.67, 0.88)	(0.69, 0.87)
λ	(0, 7.28 $\cdot 10^{-8}$)	(0, 6.57 $\cdot 10^{-8}$)

A la vista de los intervalos correspondientes a β , concluimos que ninguno de ellos contienen al 1, por lo que estaremos ante una situación de mortalidad infantil, debida a defectos de fabricación o a errores en el control de calidad. Observamos también que la longitud de los intervalos de λ es muy pequeña.

Bibliografía

- [1] Epstein, B. y Sobel, M. *Life Testing*. Journal of the American Statistical Association. Volume 48, No. 263(Sep. 1953), pp. 486-502.
- [2] Gamerman, D. y Lopes, H. F. *Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference (2nd Edition)*. Chapman & Hall, 2006.
- [3] Griful Ponsati, E. *Fiabilidad Industrial*. Ediciones UPC, S.L., 2004.
- [4] Hermida Diéguez, J. J. *Fiabilidad y Garantía en Procesos de Producción*. Proyecto Fin de Carrera presentado en la Escuela Técnica Superior de Ingeniería Industrial de la Universidad de Vigo, 2007.
- [5] Jiang, R. , Zuo, M.J. y Li, H.-X. *Weibull and Inverse Weibull Models allowing negative weights*. Reliability Engineering and System Safety, No. 66, 1999, pp. 227-234.
- [6] Kalbfleish, J. y Prentice, R. L. *The Statistical Analysis of Failure Lifetime Data*. John Wiley & Sons, Inc., 2002.
- [7] Klein, J. P. y Moeschberg, M. L. *Survival Analysis for Censored and Truncated Data*. Springer, 1982.
- [8] Lawless, J. F. *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, Inc., 2003.

- [9] Lee, E. T. y Wenyu Wang, J. *Statistical Methods for Survival Data Analysis*. Wiley, 2003.
- [10] Nachlas, J. A. *Fiabilidad* (1ª edición). Isdefe (Publicaciones de Ingeniería de Sistemas), 1995.
- [11] Nelson, W. *Applied Lifedata Analysis*. John Wiley & Sons, Inc., 1982.
- [12] Prabhakar Murphy , D.N. , Min, X. y Jiang, R. *Weibull Models*. John Wiley & Sons, Inc. , 2004.
- [13] Shortle, J. y Mendel, M.A. . *Predicting Dynamic imbalance in rotors*. Probability Engineering Mechanics, No. 11, pp. 31-35, 1996.