

Proyecto Fin de Máster
Máster Universitario en Ingeniería Aeronáutica

Análisis de la Influencia de los Datos Cen-
surados en el Cálculo del Intervalo Óptimo
entre Mantenimientos Preventivos y Estudio
Comparativo de Varios Métodos de Estima-
ción

Autor: Antonio Á. Cuenca Ávalos

Tutor: Antonio J. Sánchez Herguedas

Dpto. Organización Industrial y Gestión de Empresas I
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2022



Proyecto Fin de Máster
Máster Universitario en Ingeniería Aeronáutica

Análisis de la Influencia de los Datos Censurados en el Cálculo del Intervalo Óptimo entre Mantenimientos Preventivos y Estudio Comparativo de Varios Métodos de Estimación

Autor:

Antonio Á. Cuenca Ávalos

Tutor:

Antonio J. Sánchez Herguedas

Profesor Contratado Doctor

Dpto. Organización Industrial y Gestión de Empresas I
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2022

Proyecto Fin de Máster: Análisis de la Influencia de los Datos Censurados en el Cálculo del Intervalo Óptimo entre Mantenimientos Preventivos y Estudio Comparativo de Varios Métodos de Estimación

Autor: Antonio Á. Cuenca Ávalos
Tutor: Antonio J. Sánchez Herguedas

El tribunal nombrado para juzgar el trabajo arriba indicado, compuesto por los siguientes profesores:

Presidente:

Vocal/es:

Secretario:

acuerdan otorgarle la calificación de:

El Secretario del Tribunal

Fecha:

Agradecimientos

En primer lugar, me gustaría agradecerle a mi tutor, Antonio J. Sánchez Herguedas, el haberme acompañado en este último tramo de mis estudios de máster, prestándome su ayuda cuando la necesitaba y descubriéndome todas estas herramientas matemáticas tan útiles que pueden materializarse en aplicaciones reales.

Una vez más, querría dale las gracias a mis padres, a mi hermano y al resto de mi familia, por todo el apoyo mostrado durante toda mi vida, en especial, en los momentos más duros. Gracias a ellos, he llegado a donde estoy y tengo un futuro prometedor por delante.

Finalmente, no puedo cerrar este apartado sin darle las gracias a todos mis amigos. Tanto a los de Córdoba, como a los de Cádiz, quienes siguen demostrando lo que es la amistad verdadera, a pesar de habernos ido a vivir a diferentes ciudades, y hacen parecer que no ha pasado el tiempo cuando nos volvemos a reunir. Además, no podían faltar los agradecimientos a mis nuevos amigos de Sevilla, que no solo han demostrado un gran compañerismo, ayudándome en todo momento, si no que le han dado el broche de oro a mi época universitaria.

A todos vosotros os dedico mi trabajo.

*Antonio Ángel Cuenca Ávalos
Sevilla, 2022*

Resumen

Una de las decisiones más importantes que debe tomar un ingeniero es cuándo realizar los mantenimientos preventivos de un equipo, con el fin de que no fallen debido al desgaste. Históricamente, se han definido los intervalos entre mantenimientos preventivos a partir de los datos de fallo recogidos de la observación de varios equipos en funcionamiento. No obstante, una estimación más precisa puede realizarse teniendo en cuenta los datos censurados producidos en esa misma observación.

El objetivo de este proyecto es estudiar cómo influye la inclusión de dichos datos censurados en el cálculo del intervalo óptimo entre mantenimientos preventivos. Para ello, se hace uso de un método extraído de la literatura existente, en el que se supone al equipo en un modelo de tres estados: operativo, correctivo y preventivo. En el primero se obtienen beneficios, en los otros dos, se producen costes. La fórmula que calcula el valor del intervalo óptimo entre mantenimientos preventivos hace uso tanto de estos retornos, como de los parámetros de forma, escala y umbral resultantes de ajustar la muestra de datos de fallo a una distribución Weibull.

Para obtener los parámetros de Weibull, se emplean tres métodos de estimación diferentes: la estimación de máxima verosimilitud, el método de Newton-Raphson y la regresión del rango de medianas. El primero de ellos es uno de los métodos con mejor desempeño según la bibliografía revisada. El método de Newton-Raphson, por su parte, es una versión del método anterior que permite la convergencia de la función de verosimilitud a una buena velocidad en las situaciones en las que la estimación de máxima verosimilitud no funciona correctamente. Por último, la regresión del rango de medianas es uno de los métodos más utilizados históricamente por los ingenieros, debido a su sencillez. A lo largo de este proyecto, se incorporan los resultados obtenidos por cada uno de estos métodos y se establecen comparaciones.

Por último, cabe destacar que se incluye un capítulo en el que se recogen varios métodos que han sido mencionados a lo largo del documento y se explican brevemente, haciendo hincapié en los tres métodos de estimación usados para calcular los parámetros de Weibull, con el fin de facilitar la comprensión del lector.

Abstract

It is important that the engineer defines properly when to perform preventive maintenances on the assets in order to avoid failure. Historically, the intervals between preventive maintenance have been defined on the basis of failure data collected from the observation. Nevertheless, more accurate estimates can be made taking censored data into consideration.

In this project, the influence of censored data on the optimal interval between preventive maintenances is studied. For this purpose, a method in which the asset is assumed to be in a three-state model is used. Those three states are operational, corrective and preventive. The first one produces incomes, while the other two produce costs. To obtain the optimal interval desired value, it is used an expression that makes use of those returns, as well as the three Weibull parameters.

Three estimation methods are used to calculate those Weibull parameters: the maximum likelihood estimation, the Newton-Raphson method and the median rank regression. The MLE is one of the best performing methods according to the literature reviewed. The Newton-Raphson method is a version of the above method that allows faster convergence when the MLE does not work properly. Lastly, the MRR has been one of the most used methods by the engineers due to its simplicity. Throughout this project, the results obtained by each of these methods are included, establishing comparisons between them.

To facilitate the reader's understanding, there is a chapter that briefly explains several estimation methods mentioned throughout the document. Emphasis is placed on the three methods mentioned in the previous paragraph.

Índice

<i>Agradecimientos</i>	I
<i>Resumen</i>	III
<i>Abstract</i>	V
<i>Índice de Figuras</i>	IX
<i>Índice de Tablas</i>	XI
<i>Notación</i>	XIII
1 Estudio de la literatura existente	1
1.1 Tipos de datos censurados	1
1.2 Revisión de la literatura	3
1.3 Conclusiones extraídas	6
2 Descripción de los métodos	7
2.1 Estimación de máxima verosimilitud	7
2.1.1 Estimación de máxima verosimilitud aproximada	8
2.1.2 Algoritmo de Newton-Raphson	8
2.1.3 Algoritmo de esperanza-maximización	9
2.1.4 Algoritmo EM gradiente	9
2.1.5 Algoritmo EM estocástico	9
2.2 Método de los momentos	10
2.3 Método del percentil	10
2.4 Estimación <i>L-moment</i>	10
2.5 Estimación de la regresión del rango de medianas	11
2.6 Método de Monte Carlo basado en cadenas de Markov	12
2.7 Método BLUE	12
2.8 Método Gupta	13
2.9 Estimación de Bayes	13
2.10 Maximización de la restauración bayesiana	13
3 Antecedentes	15
3.1 Marco teórico	15
3.2 Cálculo del intervalo preventivo	16
3.3 Teoría de fallo	18
4 Aplicación matemática	21
4.1 Base común a los diferentes casos de estudio	21
4.2 Caso de estudio 1: Desgaste	22
4.2.1 Estudio teniendo en cuenta los datos censurados	23
4.2.2 Estudio sin tener en cuenta los datos censurados	27

4.3	Caso de estudio 2: Fallos aleatorios	34
4.4	Caso de estudio 3: Fallos infantiles	36
5	Conclusiones	41
	<i>Bibliografía</i>	43

Índice de Figuras

1.1	Representación gráfica de los tipos de censura por la derecha	2
2.1	Ejemplo de gráfica obtenida con los datos facilitados en [33]	12
3.1	Transición entre los tres estados del modelo.	15
3.2	Evolución del retorno medio acumulado.	16
3.3	Curva de bañera [43]	19
3.4	Comportamiento del retorno medio acumulado para diez transiciones [1]	20
3.5	Comportamiento teórico de τ_0 para diez transiciones [1]	20
4.1	Resultados obtenidos mediante MRR para el caso de desgaste con censura	23
4.2	Iteraciones del método de Newton con $\gamma = 113.36$, para el caso de desgaste con censura	25
4.3	Iteraciones del método de Newton con $\gamma = 95.66$, para el caso de desgaste con censura	25
4.4	Intervalos óptimos entre mantenimientos preventivos para el caso de desgaste con censura	26
4.5	Resultados obtenidos mediante MRR para el caso de desgaste sin censura	31
4.6	Iteraciones del método de Newton con $\gamma = 113.21$, para el caso de desgaste con censura	32
4.7	Iteraciones del método de Newton con $\gamma = 62.90$, para el caso de desgaste con censura	32
4.8	Intervalos óptimos entre mantenimientos preventivos para el caso de desgaste con censura	33
4.9	Resultados obtenidos mediante MRR para el caso de fallos aleatorios	35
4.10	Iteraciones del método de Newton para el caso de fallos aleatorios	35
4.11	Intervalo óptimo entre mantenimientos preventivos según MRR para el caso de fallos aleatorios	36
4.12	Resultados obtenidos mediante MRR para el caso de fallos infantiles	37
4.13	Iteraciones del método de Newton para el caso de fallos infantiles	38
4.14	Intervalo óptimo entre mantenimientos preventivos según MRR para el caso de fallos infantiles	39

Índice de Tablas

1.1	Ventajas del algoritmo EM	4
1.2	Desventajas del algoritmo EM	5
4.1	Parámetros comunes a todos los casos de estudio	22
4.2	Datos de fallo para el caso de desgaste	22
4.3	Datos censurados para el caso de desgaste	23
4.4	Parámetros de Weibull obtenidos mediante el MLE para el caso de desgaste con censura	23
4.5	Resultados del estudio de velocidad de convergencia para el desgaste con datos censurados	25
4.6	Resumen de los resultados obtenidos para el caso de desgaste con datos censurados	25
4.7	Parámetros de Weibull obtenidos mediante el MLE para el caso de desgaste sin censura	27
4.8	Incremento, número de orden ajustado y RM_i de los datos de fallo sin tener en cuenta la censura	28
4.9	Incremento, número de orden ajustado y RM_i de los datos de fallo teniendo en cuenta la censura	29
4.10	Resultados del estudio de velocidad de convergencia para el desgaste sin datos censurados	32
4.11	Resumen de los resultados obtenidos para el caso de desgaste sin datos censurados	32
4.12	Valores finales de τ_0	34
4.13	Datos de fallo para el caso de fallos aleatorios	34
4.14	Datos censurados para el caso de fallos aleatorios	34
4.15	Resumen de los resultados obtenidos para el caso de fallos aleatorios	36
4.16	Datos de fallo para el caso de fallos infantiles	37
4.17	Datos censurados para el caso de fallos infantiles	37
4.18	Resumen de los resultados obtenidos para el caso de fallos infantiles	38

Notación

A	Tiempo medio que el sistema permanece en el estado operativo antes de que se produzca el fallo	
B	Tiempo medio que el sistema permanece en el estado correctivo	
C	Tiempo medio que el sistema permanece en el estado preventivo	
C_i^n	Coficiente binomial empleado en la estimación <i>L-moment</i>	
d	Posición anterior a la posición r en la estimación <i>L-moment</i>	
$\mathbf{F}(\boldsymbol{\tau})$	Matriz de tiempo de permanencia	
$F(t)$	Función de distribución acumulativa de fallos del equipo	
$f(t)$	Función de densidad de probabilidad de fallos del equipo	
$G(t_c)$	Función de distribución acumulativa del tiempo en el que el equipo se encuentra bajo mantenimiento correctivo	
$g(t_c)$	Función de densidad de probabilidad del tiempo en el que el equipo se encuentra bajo mantenimiento correctivo	
$h(\boldsymbol{\theta})$	Derivada de la función de verosimilitud respecto a un estimador cualquiera	
$h'(\boldsymbol{\theta})$	Segunda derivada de la función de verosimilitud respecto a un estimador cualquiera	
$J(t_p)$	Función de distribución acumulativa del tiempo en el que el equipo se encuentra bajo el mantenimiento preventivo	
$j(t_p)$	Función de densidad de probabilidad del tiempo en el que el equipo se encuentra bajo el mantenimiento preventivo	
\mathbf{I}	Matriz unidad	
$L(\boldsymbol{\alpha}, \boldsymbol{\beta})$	Función de verosimilitud	
$\ln L(\boldsymbol{\alpha}, \boldsymbol{\beta})$	Función de verosimilitud logarítmica	
m	Número de transiciones	
m_r'	Muestra correspondiente a un momento TL	
m_1^0	Muestra simplificada para una posición 1	
m_2^0	Muestra simplificada para una posición 2	
n	Totalidad de observaciones tomadas, tamaño de la muestra	
N_1, N_2	Parámetros de cálculo de $h(\boldsymbol{\theta})$ y $h'(\boldsymbol{\theta})$	
N_3, N_4		
N_5		
\mathbf{P}		Matriz de probabilidad de transición entre estados
p_{ij}		Probabilidad ligada a un valor total de los retornos producidos en un estado i y durante su transición al estado j
$Q_i(1)$	Retorno en un solo paso	
$Q_i(m)$	Retorno acumulado desde el estado i en m transiciones	
q_i	Número de orden ajustado de un elemento i	
q_{i-1}	Número de orden del dato de fallo anterior al elemento i	

R_1	Ingresos obtenidos durante el estado operativo
R_2	Costes derivados del estado correctivo
R_3	Costes derivados del estado preventivo
R_{ij}	Gastos producidos de la transición de un estado i a un estado j
RM_i	Función de distribución observada, rango de la mediana del elemento i
r	Posición determinada en la estimación <i>L-moment</i>
r_{ij}	Valor total de los retornos producidos en un estado i y durante su transición al estado j
t_i	Valor de la observación i
$\mathbf{V}(m)$	Forma matricial del retorno medio acumulado
$v_i(m)$	Retorno medio acumulado
$X_{i:n}$	Estadístico i ésimo dentro del tamaño n de la muestra
α	Factor de forma de Weibull
$\hat{\alpha}_{LM}$	Estimador del factor de forma según la estimación <i>L-moment</i>
β	Factor de escala de Weibull
$\hat{\beta}_{LM}$	Estimador del factor de escala según la estimación <i>L-moment</i>
$\Gamma(\cdot)$	Función Gamma
γ	Factor de localización o umbral de Weibull
δ_i	Indicador de censura
ε	Error impuesto
θ_k	Valor inicial de un estimador genérico
θ_{k+1}	Valor de un estimador genérico en la siguiente iteración
μ_r'	Momento TL
$\pi(\theta)$	Distribución a priori
τ	Intervalo entre mantenimientos preventivos
τ_0	Intervalo óptimo entre mantenimientos preventivos

1 Estudio de la literatura existente

En este capítulo, se realiza una revisión de la literatura existente acerca del trato de los datos censurados en sus diferentes tipos, a la hora de realizar estudios de fiabilidad. El objetivo es analizar los diferentes métodos que han utilizado varios expertos a la hora de incluir en sus análisis la información que ofrecen los datos censurados.

1.1 Tipos de datos censurados

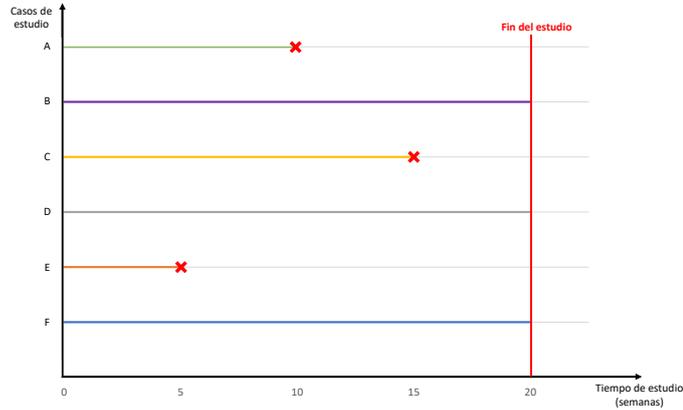
Los datos censurados hacen referencia a aquella información de la que no se dispone debido a que se desconoce cuando ocurre cierto evento clave. Particularizando en los análisis de fiabilidad, puede realizarse un estudio sobre el tiempo de fallo de una muestra de todos los activos de una planta. De dicho estudio, puede extraerse información de relevancia sobre cuándo pueden fallar dichos activos, lo cual es importante, por ejemplo, a la hora de establecer los intervalos de tiempo entre mantenimientos preventivos. La censura ocurre cuando no se puede conocer el momento exacto del fallo individual de todos los activos de la planta [47].

Existen diferentes tipos de censura [12]:

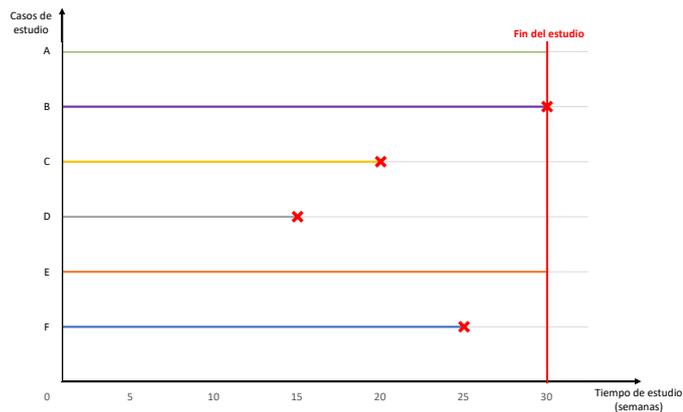
- **Censura por la izquierda.** También conocida como *left censoring*, en inglés. Este tipo de censura indica que el evento de relevancia ha ocurrido en algún instante antes de que el activo sea objeto de estudio, desconociéndose el momento exacto [20].
- **Censura por la derecha.** Por el contrario, cuando el evento relevante ocurre en un momento desconocido, posterior a cierto límite de tiempo definido, se dice que existe una censura por la derecha, o *right censoring*, como se conoce en inglés [47]. Por su parte, existen varios subtipos de censura por la derecha:
 - **Censura tipo 1.** En este caso, existe un tiempo fijo de censura, que puede ser el tiempo de estudio [19]. De este modo, el evento de relevancia quedará determinado en todos los sujetos de estudio que lo sufran durante el análisis. Todos aquellos que no lo experimenten en el tiempo fijado, puede que lo hagan tiempo después de que termine el estudio, aunque será indeterminado, quedando dicho dato censurado.
 - **Censura tipo 2.** Por su parte, la censura tipo 2 aparece en aquellos análisis que definen su duración según la proporción de activos que sufren el fallo [19]. Por ejemplo, si en cierto análisis se estudian seis equipos, se puede establecer que el estudio termina cuando cuatro de ellos registren el fallo, quedándose indeterminados los tiempos de fallo correspondientes a los otros dos equipos. Es por ello que en la literatura puede encontrarse textos que se refieren a este tipo de censura como “censura de fallo” (*failure censoring*) [47]. La figura 1.1 ilustra el ejemplo propuesto.
 - **Censura tipo 3.** Mientras que los dos tipos anteriores de censura se dan en aquellos análisis en los que los activos que son objeto de estudio son introducidos al principio del periodo experimental, este tercer tipo de censura tiene lugar en aquellos en los que cada activo presenta un instante inicial diferente. No obstante, también presentan un tiempo fijo de censura. Es por ello que este tipo de censura recibe el nombre de “censura tipo 1 generalizada” [19].

- **Censura de intervalo.** Cuando en un mismo estudio se dan los dos tipos de censuras anteriores, se dice que existe una censura en intervalo (*interval censoring*) [20].

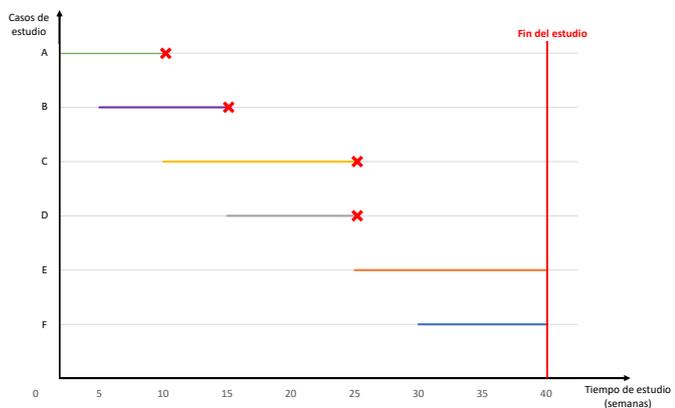
Por otro lado, hay que tener en cuenta que no son lo mismo los datos censurados que los datos truncados. En el truncamiento, solo se tienen en cuenta aquellos activos cuyo evento ocurre dentro de una ventana de observación. De este modo, no se dispone de información de todo activo cuyo evento quede fuera de dicha ventana. Esto es lo que contrasta con la censura, en la que hay al menos información parcial sobre cada activo de estudio [20]. Este documento se centra en los datos censurados, pero se han descrito los datos truncados ya que pueden ser mencionados en adelante.



(a) Censura tipo 1.



(b) Censura tipo 2.



(c) Censura tipo 3.

Figura 1.1 Representación gráfica de los tipos de censura por la derecha.

1.2 Revisión de la literatura

Una vez definidos los diferentes tipos de datos censurados, se hace uso de la revisión de la literatura realizada para exponer las diferentes técnicas que han empleado numerosos autores a la hora de incluir los datos censurados en sus estudios, como se indica al principio del capítulo.

La distribución Weibull es un modelo de probabilidad ampliamente utilizado tanto en los estudios de vida útil, como en la teoría de fiabilidad [61]. Gran parte de la literatura revisada hace uso de ella para tratar las muestras con datos censurados. Cabe recordar que la distribución de Weibull depende de tres parámetros: el de forma (α), el de escala o vida característica (β) y el de localización (γ), también conocido como valor umbral o vida garantizada. Cuando este último tiene un valor nulo, se dispone de una distribución Weibull de dos parámetros [25].

El trabajo realizado por Teimouri et al. [26] establece una comparativa entre diferentes métodos de estimación propuestos para resolver una distribución Weibull. Estos son la estimación de máxima verosimilitud, el método del momento logarítmico, el método del percentil, el método de los momentos y el método de estimación *L-moment*, propuesto por los autores.

El estudio numérico consiste en un conjunto de 10 000 individuos con los que se hacen varios estudios con muestras de diferente tamaño. El modelo se ajusta a una distribución Weibull de dos parámetros y se utiliza el sesgo y el error de mínimos cuadrados como criterios de comparación. Finalmente, los autores concluyen que la estimación de máxima verosimilitud es la que mejor comportamiento presenta, para muestras lo suficientemente grandes. Le sigue la estimación *L-moment*, que actúa correctamente en la mayoría de casos estudiados. Se remarca el hecho de que, en ciertos casos, estos dos métodos alcanzan los mismos valores del error de mínimos cuadrados. El resto de métodos no alcanzan tan buenos resultados.

Según Genschen et al. [60], históricamente, el método de los mínimos cuadrados ordinarios ha sido una alternativa empleada por los ingenieros a la estimación de máxima verosimilitud, debido a su fácil programación. El método más común a la hora de implementar el enfoque de mínimos cuadrados ordinarios es la regresión del rango de medianas. En su artículo, establecen una comparativa entre ambos métodos para poner de manifiesto las ventajas que tiene cada uno de ellos. Antes de empezar el estudio, los autores parten de las siguientes ideas:

- Según la experiencia de los autores, es difícil encontrar un método de estimación que supere el buen funcionamiento del de máxima verosimilitud.
- Con el paso del tiempo, los métodos de mínimos cuadrados ordinarios han ido dejando de estar incluidos en los *softwares* comerciales más comunes, a diferencia del método de máxima verosimilitud. Por ello, su programación a día de hoy no es mucho más sencilla que la de la estimación de máxima verosimilitud.
- Tanto los estimadores de máxima verosimilitud, como los de la regresión del rango de medianas están sesgados. En el segundo caso, el sesgo suele estar dominado por la varianza; por lo que, cuando se compara la precisión global de los estimadores de ambos métodos, suelen dar mejores resultados los de máxima verosimilitud.
- La regresión del rango de medianas dibuja la línea atravesando los puntos trazados. El método de máxima verosimilitud lo hace cuando corresponde, pero la regresión del rango de medianas lo hace incluso cuando no debe. Eso puede llevar a resultados erróneos.
- La estimación de máxima verosimilitud presenta propiedades óptimas en muestras lo suficientemente grandes.
- La estimación de máxima verosimilitud puede aplicarse en situaciones de censura más complicada, como la de intervalo o truncamiento.

Como criterios de comparación, los autores emplean los parámetros de forma y de escala, así como varios cuantiles de Weibull. Se realizan estudios con muestras de datos con censura tipo 1 y 2. Finalmente, los autores llegan a la conclusión de que la estimación de máxima verosimilitud presenta un mejor comportamiento que la estimación por regresión del rango de medianas a lo largo de la mayoría del estudio. Referente a la censura tipo

1, la estimación de máxima verosimilitud tiene ciertas ventajas, ya que no ignora la información censurada. Igualmente, posee una mayor precisión que la regresión del rango de medianas ante datos censurados del tipo 2, aunque las diferencias son menores que las detectadas en las simulaciones con datos censurados del tipo 1.

Por su parte, Shen et al. proponen en su artículo [5] un nuevo método para llevar a cabo los estudios de fiabilidad con datos censurados. Dicho método hace uso de una serie de funciones de regresión, analizada mediante un enfoque Bayesiano y un algoritmo de Monte Carlo basado en Cadenas de Markov (MCMC, por sus siglas en inglés), para transformar los datos censurados en datos de fallo. Los autores aseguran que, de este modo, se simplifica el estudio de fiabilidad, manteniendo la precisión computacional. Este nuevo método tendría aplicación en los casos donde la muestra es reducida y la censura es alta.

Por otro lado, el uso del algoritmo de esperanza-maximización para resolver la estimación de máxima verosimilitud puede traer problemas de convergencia si el tamaño de la muestra no es lo suficientemente grande, lo cual es consistente con lo indicado en [60]. Igualmente, el algoritmo de Monte Carlo basado en Cadenas de Markov puede suponer una alternativa al algoritmo de esperanza-maximización si los cálculos se vuelven demasiado complejos y no se tiene una precisión adecuada.

Otros autores que comparan varios métodos de estimación de parámetros para una distribución Weibull son Avijit Joarder et al. En su artículo [6], indican que los estimadores de máxima verosimilitud no pueden obtenerse de forma explícita, haciendo uso de un proceso iterativo, por lo que proponen la llamada estimación de máxima verosimilitud aproximada, la cual presenta forma explícita en sus estimadores y, además, puede ser fácilmente computada. La contribución principal de este artículo es establecer una comparativa entre estas dos estimaciones y la estimación bayesiana. La comparación se lleva a cabo mediante numerosas simulaciones de Monte Carlo y se utilizan el sesgo y el error de mínimos cuadrados como criterio comparativo, al igual que en [26].

Como conclusión general, llegan a que todos los métodos funcionan correctamente menos cuando la muestra y el tiempo son demasiados pequeños. Por su parte, la estimación de máxima verosimilitud aproximada se comporta de forma muy similar a la estimación de máxima verosimilitud original, aunque presenta intervalos de confianza ligeramente más cortos. Por otro lado, la estimación bayesiana registra peores comportamientos, además, para muestras muy grandes. Finalmente, indican que si no se tienen ninguna información de los parámetros desconocidos, la estimación de máxima verosimilitud aproximada puede usarse para la mayoría de casos prácticos.

Los artículos expuestos hasta el momento coinciden en que el método que mejor comportamiento presenta, de forma generalizada, es el de máxima verosimilitud. Siguiendo este mismo hilo, el artículo escrito por Ferreira y Silva [56] tiene como objetivo validar la aplicación del algoritmo de esperanza-maximización como método para resolver la estimación de máxima verosimilitud de un conjunto de datos incompletos, censurados por la derecha. Además de lograr dicho objetivo, los autores incluyen una lista de ventajas y desventajas del algoritmo de esperanza-maximización que puede resultar interesante a la hora de decidir el algoritmo a utilizar en la parte práctica de este proyecto. Dicha lista se encuentra plasmada en las tablas 1.1 y 1.2

Tabla 1.1 Ventajas del algoritmo EM.

Ventajas
Converge en un gran número de condiciones, excepto si no se hace una buena elección de los valores iniciales del parámetro a maximizar.
El trabajo analítico requerido es más sencillo que el de otros métodos.
Es relativamente sencillo de programar e implementar.
Permite controlar la convergencia y los errores de programación durante las iteraciones.

Tabla 1.2 Desventajas del algoritmo EM.

Desventajas
Puede converger lentamente si la información sufre una fuerte censura o truncamiento.
No garantiza la convergencia global si existen varios puntos máximos. La estimación obtenida depende de la solución inicial.

Por otro lado, en el artículo desarrollado por Balakrishnan y Mitra [51], se dispone de ciertos datos truncados por la izquierda y censurados por la derecha. Los autores hacen uso de un algoritmo EM Gradiente para ajustar la distribución Weibull a dicho conjunto de datos. La razón por la que usan dicho algoritmo está relacionada con la primera desventaja plasmada en la tabla 1.2, ya que suele emplearse para acelerar la convergencia del modelo [13]. Además, con fines comparativos, se realiza otro ajuste mediante el método de Newton-Raphson, el cual también acelera la convergencia de la estimación de máxima verosimilitud y es más sencillo.

En dicho artículo, los autores llegaron a la conclusión de que el algoritmo EM Gradiente presenta algunos problemas de convergencia para muestras de alrededor de 100 individuos e inferiores, lo cual coincide con lo indicado por Shen et al. [5]. Igualmente, se detectan problemas del mismo tipo si se toman valores iniciales de los parámetros muy alejados de los reales. Por último, queda comentar que, aunque se alcanzan resultados muy similares con ambos métodos, se registran pequeñas diferencias en los intervalos de confianza y en las probabilidades de confianza. Mientras que los intervalos de confianza para el parámetro de localización, son similares para los dos algoritmos, en el caso del parámetro de escala, el algoritmo EM Gradiente presenta intervalos de confianza más amplios que el método de Newton-Raphson. Del mismo modo, para mayores porcentajes de truncamiento, mayores probabilidades de cobertura del parámetro de localización se registran.

Finalmente, cabe comentar que Robert Ross [38] [39] y Hideo Hirose [17], proponen en sus trabajos un método para reducir el sesgo de los estimadores de máxima verosimilitud, para muestras pequeñas, de forma que mejora el comportamiento de la estimación. Además, Zhang et al. [22] proponen una modificación de los métodos propuestos por Ross y Hirose aplicada a la estimación de mínimos cuadrados.

Respecto a los datos con censura del tipo 2, Riffat Jabeen et al. [36] realizan un estudio en el que comparan la eficiencia de dos métodos con los que obtienen los estimadores de la muestra: el BLUE (*Best Linear Unbiased Estimator*) y el método Gupta. Finalmente, llegan a la conclusión de que el método de Gupta puede reemplazarse por el BLUE, ya que se alcanzan valores de eficiencia relativa bastante alta. Esto es útil cuando se conocen las esperanzas, pero la matriz de varianza-covarianza es desconocida.

Por otro lado, como James K. Starling et al. comentan en su artículo [19], en muchas aplicaciones reales suele darse una censura de tipo 1 generalizada. Son conscientes de los problemas de sesgo que presenta la estimación de máxima verosimilitud y proponen un método mejorado combinando dicha estimación con un método de sobremuestreo. La conclusión del estudio es que los métodos propuestos reducen el sesgo en los estimadores del parámetro de forma, pero aumentan el sesgo en el de escala. No obstante, mediante una reducción en la divergencia de Kullback-Leibler, se puede solucionar el problema de aumento de sesgo del estimador del parámetro de escala.

En ciertas situaciones, los datos de la muestra pueden mostrar cierta heterogeneidad. Esto significa que la fiabilidad de los individuos de la muestra depende de parámetros muy diversos. Para tratar dicha heterogeneidad, se suele usar la distribución de mezcla Weibull de dos componentes. De este modo, para afrontar los problemas derivados de una alta censura en los datos, Ducros y Pamphile [57] proponen un método de *bootstrap* bayesiano llamado Maximización de la Restauración Bayesiana (BRM, por sus siglas en inglés). Las ventajas de este método son que presenta mejores resultados en términos de precisión respecto al algoritmo EM y a su variante, el algoritmo EM estocástico. Además, al no ser un método iterativo, requiere de tiempos menores de computación.

Por otro lado, Musleh y Helu, en su artículo [37], indican que la distribución Weibull no tiene un buen desempeño en aquellas muestras que disponen de tasas de fallo no monótonas, tal y como puede ser la unimodal. En estos casos, la distribución Weibull inversa es un modelo más adecuado. En dicho artículo,

los autores comparan varios modelos clásicos (estimación de máxima verosimilitud, estimación de máxima verosimilitud aproximada y estimación de mínimos cuadrados) y varios modelos bayesianos en la estimación de los parámetros de una distribución Weibull inversa, considerando una censura de tipo 2 progresiva. Se llega a la conclusión de que los estimadores de Bayes tienen un mejor desempeño trabajando con esta distribución, ya que presentan un sesgo menor, así como un error de mínimos cuadrados más bajo y una mejor probabilidad de proximidad de Pitman.

Por último, queda comentar que aunque la distribución Weibull es la más extendida, existen otros tipos de distribuciones con las que se tratan las muestras de datos. En particular, Ranjan et al. [34] realizan un análisis comparativo entre el modelo Weibull, el modelo Gamma y el modelo Lognormal, haciendo uso de un enfoque bayesiano, empleando los algoritmos Metrópolis y Monte Carlo Hamiltoniano. La comparación se realiza mediante los valores de la función de verosimilitud marginal. Concluyen el estudio indicando que el modelo Gamma es el que peores resultados presenta. Por otro lado, tanto la distribución Lognormal, como la Weibull, presentan buenos resultados, resultando la primera la más adecuada de las tres para tratar los datos considerados.

Si bien en el artículo anterior, la distribución Lognormal presentaba un mejor desempeño que la Weibull, cabe indicar que esto varía según la muestra considerada. Nanang, en su trabajo [28], compara también las distribuciones Weibull y Lognormal, añadiendo la distribución Normal al estudio. En este caso, de nuevo las dos primeras presentan un mejor comportamiento; aunque, en este caso, es la distribución Weibull la que mejor se ajusta a la muestra.

1.3 Conclusiones extraídas

Después de realizar la revisión literaria, se pueden llegar a ciertas conclusiones:

- La distribución de Weibull es la más ampliamente utilizada y de la que más material bibliográfico se dispone. No obstante, la distribución Lognormal parece presentar también muy buenos resultados según la muestra con la que se trate.
- En términos generales, el método que mejor funciona a la hora de tratar datos censurados es la estimación de máxima verosimilitud, en particular haciendo uso del algoritmo EM Gradiente, que optimiza la velocidad de convergencia del algoritmo EM original. Como alternativa en casos de muestras muy pequeñas o datos altamente censurados, se puede hacer uso del método propuesto por Shen et al. [5] o se puede hacer uso de las propuestas de Ross [38] [39] y Hirose [17] para reducir el sesgo de la estimación de máxima verosimilitud. Otro algoritmo al que se puede recurrir para acelerar la velocidad de convergencia es el de Newton-Raphson.
- La regresión del rango de medianas ha sido un método muy utilizado históricamente por los ingenieros, debido a su sencillez, aunque, con el desarrollo de la estimación de máxima verosimilitud, ha quedado atrás al mostrar un peor desempeño.
- Respecto a la censura tipo 2, se llega a la conclusión de que los métodos BLUE y Gupta son intercambiables entre sí.
- Para muestras de datos que sufran de cierta heterogeneidad, la distribución de Weibull no presenta un buen comportamiento. Se puede hacer uso de la distribución de mezcla Weibull de dos componentes y del método BRM, desarrollado por Ducros y Pampline [57], o de una distribución de Weibull inversa con algunos de los estimadores de Bayes expuestos en el trabajo de Musleh y Helu [37].

2 Descripción de los métodos

A lo largo de la sección 1.2, se hace referencia a un gran número de métodos de estimación y de algoritmos que se usan para ajustar el modelo de distribución pertinente a los conjuntos de datos correspondientes en cada uno de los estudios que se han revisado. En este capítulo, se incluye una breve descripción de cada uno de ellos, haciendo mayor hincapié en aquellos que se consideren más importantes para este proyecto, con el fin de facilitar su comprensión.

2.1 Estimación de máxima verosimilitud

La sección 1.3 concluye que, de entre todos los métodos mencionados a lo largo del capítulo anterior, la estimación de máxima verosimilitud es la que mejor funciona a la hora de tratar datos censurados. Podría decirse que este método es una herramienta muy extendida y empleada en el campo de la inferencia, presentando resultados óptimos cuando las muestras son lo suficientemente grandes [40]. Igualmente, como ya se comentó a lo largo del capítulo 1, su estudio está ya tan extendido que se han desarrollado varias propuestas para reducir el sesgo de los resultados para muestras pequeñas [17] [38] [39].

El objetivo de este método de estimación es el de encontrar ciertos valores de los parámetros de Weibull (α y β) que maximicen la probabilidad de obtener los datos observados. Esto se consigue optimizando la función de máxima verosimilitud. Para poder alcanzar dicha expresión, se parte de la función de densidad de probabilidad de Weibull [33]:

$$f(t) = \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{t}{\beta}\right)^\alpha\right] \quad (2.1)$$

Dicha función proviene de derivar respecto de t la función de distribución acumulativa $F(t)$ [33], que atiende a la siguiente expresión:

$$F(t) = 1 - \exp\left(-\left(\frac{t}{\beta}\right)^\alpha\right) \quad (2.2)$$

De este modo, la función de verosimilitud se calcula a partir de la función de densidad de probabilidad de Weibull [24]:

$$L(\alpha, \beta) = \prod_{i=1}^n f(t_i; \alpha, \beta) \quad (2.3)$$

$$L(\alpha, \beta) = \prod_{i=1}^n \left\{ \frac{\alpha}{\beta} \left(\frac{t_i}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{t_i}{\beta}\right)^\alpha\right] \right\} \quad (2.4)$$

A continuación, se obtiene la función de verosimilitud logarítmica a partir de la expresión 2.4, aplicando logaritmos neperianos [45]. El uso de esta función recién introducida está muy extendida en el mundo de la inferencia de probabilidad, debido a sus ventajosas propiedades. Al igual que la función de verosimilitud original, puede ser optimizada y presenta una optimización numérica más eficaz [40].

$$\ln L(\alpha, \beta) = n \ln \alpha - n\alpha \ln \beta + (\alpha - 1) \sum_{i=1}^n (\ln t_i) - \sum_{i=1}^n \left(\frac{t_i}{\beta} \right)^\alpha \quad (2.5)$$

Cabe destacar que la variable n hace referencia a la totalidad de observaciones tomadas. El último paso de la estimación consiste en maximizar la expresión 2.5, derivándola respecto a los parámetros de Weibull e igualando a 0; obteniéndose así el valor de los estimadores [13] [33].

Si bien el método expuesto resuelve los estimadores de máxima verosimilitud para una muestra de datos completos, cabe recordar que el proyecto trata sobre la influencia de los datos censurados. Para poder utilizar el método partiendo de unos datos incompletos, es necesario modificarlo levemente.

De esta forma, se define un indicador δ_i , ligado a cada dato de la muestra, que presenta dos únicos valores: 1 y 0. Cuando un dato se encuentre censurado, su indicador asignado tendrá un valor nulo. Por el contrario, si se detecta un fallo, el indicador del dato relacionado presentará un valor unidad [24]. En este caso, se habla de datos censurados por la derecha.

El método sigue los mismos pasos que se han expuesto anteriormente para obtener el valor de los estimadores. La única diferencia es que, en este caso, los indicadores δ_i hacen acto de presencia en la función de verosimilitud, así como en su variante logarítmica. De este modo, la expresión 2.5 aplicada a un caso con censura tiene la siguiente forma [45]:

$$\ln L(\alpha, \beta) = \sum_{i=1}^n \delta_i (\ln \alpha - \alpha \ln \beta) + (\alpha - 1) \sum_{i=1}^n (\delta_i \ln t_i) - \sum_{i=1}^n \left(\frac{t_i}{\beta} \right)^\alpha \quad (2.6)$$

Analizando brevemente la expresión, se puede observar cómo los dos primeros términos sólo van a tener en cuenta los datos de fallo, mientras que el tercero incluye a la totalidad de la muestra.

La estimación de máxima verosimilitud puede apoyarse en ciertos métodos iterativos para mejorar su desempeño en ciertas situaciones [24]. Algunos de ellos se exponen a continuación.

2.1.1 Estimación de máxima verosimilitud aproximada

Si bien la estimación de máxima verosimilitud es uno de los métodos más efectivos, en la mayoría de situaciones no facilita una forma explícita de los estimadores calculados [48]. Con el fin de alcanzar dichos estimadores explícitos, N. Balakrishnan desarrolló en numerosos trabajos, como [55] (en colaboración con J. Varadan) y [7], la llamada estimación de máxima verosimilitud aproximada.

En esta variación de la MLE, la función de verosimilitud se ve levemente modificada. Además, se emplea la función de Taylor para poder aproximar los términos complejos que conforman las complejas expresiones los estimadores y, así, obtener las formas explícitas que se buscaban [41]. Este no es un método iterativo.

2.1.2 Algoritmo de Newton-Raphson

El algoritmo de Newton-Raphson es un procedimiento utilizado para resolver ecuaciones no lineales que ha acabado por convertirse en uno de los métodos más utilizados de optimización estadística [46]. En aquellas situaciones en las que la función de verosimilitud logarítmica no puede maximizarse analíticamente, suelen emplearse métodos iterativos, como es el método de Newton-Raphson, cuya velocidad de convergencia es reconocida como uno de sus puntos fuertes [13].

En este contexto, el método de Newton-Raphson se basa en la aproximación cuadrática de la función de verosimilitud logarítmica [40]. Dicha función viene definida por la expresión 2.5, en el caso en el que la muestra de datos no sufra censura, o por la expresión 2.6, en el caso de que exista algún tipo de censura. De este modo, es necesario derivar dicha función respecto a los estimadores correspondientes [29]. Se denomina $h(\theta)$ a la derivada de la función de verosimilitud respecto a un estimador cualquiera; mientras que $h'(\theta)$ representa la derivada segunda respecto del mismo estimador [58]. Para poder calcular el valor de los estimadores, se hace uso de la siguiente expresión [12]:

$$\theta_{k+1} = \theta_k - \frac{h(\theta)}{h'(\theta)} \quad (2.7)$$

La expresión 2.7 hace uso de un valor inicial del estimador para calcular uno nuevo. Dicha expresión se obtiene mediante la aplicación de una serie de Taylor sobre la función $h(\theta)$ [40]. El método convergerá cuando la diferencia entre los valores del estimador en dos iteraciones sucesivas sea inferior a un valor de error impuesto [58]. En este proyecto, $\varepsilon = 10^{-5}$.

$$|\theta_{k+1} - \theta_k| < \varepsilon \quad (2.8)$$

Por último, cabe indicar que aunque a lo largo de esta sección se haya empleado el estimador genérico θ , se recuerda que en este proyecto se emplea una distribución Weibull para tratar los conjuntos de datos, cuyos parámetros son α , β y γ .

A modo de conclusión, cabe indicar que puede fallar la convergencia del método si los valores iniciales que se emplean están muy alejados del valor real del estimador. De cualquier modo, cuando converge, este método es mucho más rápido que otros similares [29].

2.1.3 Algoritmo de esperanza-maximización

El algoritmo EM es ampliamente utilizado en aquellos problemas donde aparece cierta censura o truncamiento en los datos de la muestra a analizar. En dichas situaciones, el uso de otro tipo de algoritmos puede complicar el cálculo de los estimadores de máxima verosimilitud (MLE, por sus siglas en inglés). El objetivo es asociar el problema dado, que parte de unos datos incompletos, con un problema de datos completos, de forma que se pueda aprovechar la mayor simplicidad computacional del segundo a la hora de calcular los estimadores de máxima verosimilitud. Dicha asociación se logra mediante las funciones de verosimilitud de cada caso.

En cada iteración, el algoritmo presenta dos pasos: el de esperanza (*E-step*) y el de maximización (*M-step*), los cuales le dan el nombre. Por un lado, en el paso de esperanza se obtiene la función de verosimilitud para el caso de datos completos, partiendo de los datos incompletos del problema inicial y del valor de ciertos parámetros no censurados. Por su parte, en el *M-step* se maximiza la función de verosimilitud obtenida con el fin de alcanzar el próximo valor de los parámetros, del que partirá la siguiente iteración.

En resumen, el funcionamiento del algoritmo consiste en repetir la iteración entre los pasos de esperanza y de maximización hasta que converja, partiendo de unos valores iniciales de los parámetros [13].

2.1.4 Algoritmo EM gradiente

Esta es una versión alternativa del algoritmo de esperanza-maximización cuyas bases se sientan en el trabajo de Lange [23].

En este caso, se hace uso del algoritmo de Newton-Raphson para llevar a cabo el paso de maximización [13]. Esto permite acelerar la convergencia del algoritmo [51] e incluso resolverlo cuando sería imposible con el método original [23].

2.1.5 Algoritmo EM estocástico

Esta es otra variación del algoritmo de esperanza-maximización que permite la convergencia del método para conjunto de datos altamente censurados o demasiado pequeños [57].

El trabajo de Celeux y Diebolt [54] sentó las bases de este método. La diferencia respecto al algoritmo original se centra en la sustitución del paso de esperanza por el llamado paso estocástico [15], capaz de completar el conjunto de datos a partir de la muestra altamente censurada dada [30]. Tras ello, el paso de maximización se encarga de obtener los estimadores para el modelo mixto de estudio [13].

2.2 Método de los momentos

El método de los momentos suele usarse con poca frecuencia y, la mayoría de las veces, con el fin de comparar varios métodos analíticos para la estimación de los parámetros de Weibull [53]. No obstante, algunos autores indican que es una alternativa común al método de estimación de máxima verosimilitud en algunos campos de investigación, como el forestal, ofreciendo una mayor velocidad y simplicidad de cálculo, a costa de sufrir una disminución en la precisión [42].

El método realiza las estimaciones haciendo uso de la media y la varianza de la muestra [28]. Además, la resolución requiere que se conozca el valor del parámetro de localización, con el objetivo de obtener los valores de los parámetros de forma y escala. Para ello, se ajusta la media de la muestra a la distribución Weibull haciendo uso del parámetro de localización y se divide entre la desviación estándar de la distribución, alcanzando una expresión en la que el ratio creado depende del parámetro de forma. Por otro lado, existe una relación entre el parámetro de escala y el de forma, donde también aparece la media ajustada. Haciendo uso de esas dos expresiones, se alcanza el valor de los parámetros deseados [42].

2.3 Método del percentil

Al igual que el método de los momentos, la estimación mediante percentiles supone una alternativa al método de máxima verosimilitud en ciertos ámbitos de estudio (forestal), ya que presenta unos cálculos más simples y económicos, debido a que no requiere de algoritmos iterativos para su resolución [59]. De ese modo, la velocidad de convergencia es mayor [62].

Para llevar a cabo la resolución del método, es necesario conocer el valor de las variables de la muestra, según el percentil con el que se está trabajando. Para este método, es importante uso del parámetro p , cuyo valor está entre 0 y 1 y representa el valor de la función de distribución, sustituyendo el valor que presenta la muestra para el percentil $100p$ con el que se está trabajando [9]. Con ello, se obtienen ciertas expresiones con las que se obtienen los valores de los estimadores, relacionándolos entre ellos [62] [59].

2.4 Estimación *L-moment*

Este es un método propuesto por M. Teimouri y sus colaboradores en su artículo [26].

Dicho método está basado en el concepto de los momentos L y TL, comprendidos como secuencias de estadísticos, teniendo en cuenta que los momentos TL son más modernos y, por lo tanto, más robustos y aplicables a distribuciones de muestras infinitas, aunque con una aplicación menos estudiada, por el momento. A su vez, el desarrollo matemático de la estimación propuesta por los autores se basa en el trabajo de Hosking [18].

Todo momento L o TL está ligado a una muestra. El desarrollo mostrado en [26] se centra en el segundo tipo. De este modo, las expresiones del momento TL (μ'_r) y de su muestra (m'_r) correspondiente son las siguientes:

$$\mu'_r = \frac{\beta}{r} \Gamma\left(\frac{1}{\alpha} + 1\right) \sum_{k=0}^{r-1} (-1)^k \cdot C_k^{r-1} \cdot (r+d-k) \cdot C_{r+d-k}^{r+2t} \sum_{j=0}^{r+d-k-1} C_j^{r+d-k-1} \frac{(-1)^j}{(d+k+j+1)^{1/\alpha+1}} \quad (2.9)$$

$$m'_r = \frac{1}{r} \sum_{i=t+1}^{n-t} \frac{\sum_{k=0}^{r-1} (-1)^k \cdot C_k^{r-1} \cdot C_{d+k}^{n-i} \cdot C_{r+d-k-1}^{i-1}}{C_{r+2d}^n} X_{i:n} \quad (2.10)$$

En las expresiones 2.9 y 2.10 aparecen multitud de parámetros nuevos. Por su parte, C_i^n representa un coeficiente binomial que atiende a la siguiente expresión:

$$C_i^n = \frac{n!}{i!(n-i)!} \quad (2.11)$$

Por otro lado, n hace referencia al tamaño de la muestra, mientras que r hace referencia a una posición determinada y d , a la posición anterior. Para los momentos L, $d = 0$. Finalmente, $X_{i:n}$ hace referencia al

estadístico i ésimo dentro del tamaño n de la muestra. Esto se debe a que esta estimación se basa en la idea de obtener el momento r ésimo mediante el estadístico i ésimo.

El proceso matemático pasa por obtener el valor del momento TL y de la muestra para dos posiciones determinadas dentro de n y simplificarlo para los momentos L (m_1^0 y m_2^0). Combinando las expresiones alcanzadas, se obtienen los estimadores:

$$\hat{\alpha}_{LM} = -\frac{\ln(2)}{\ln\left(1 - \frac{m_2^0}{m_1^0}\right)} \quad (2.12)$$

$$\hat{\beta}_{LM} = -\frac{m_1^0}{\Gamma\left(\frac{1}{\alpha + 1}\right)} \quad (2.13)$$

2.5 Estimación de la regresión del rango de medianas

La regresión del rango de medianas es un método que permite resolver una distribución de forma gráfica, enfrentando en ella a la función de distribución observada y al tiempo de operación. Cuando se trabaja con conjuntos de datos incompletos, la mecánica de este método se centra en que los datos censurados no sean representados en dicha gráfica, pero que afecten al rango de los de fallo, que sí se grafican [49].

Para poder llevar a cabo los cálculos, es necesario ordenar los datos obtenidos del análisis de menor a mayor horas de funcionamiento registradas, mezclando tanto los datos de fallo, como los censurados. De este modo, se puede calcular el nuevo incremento que hay entre dos datos de fallo consecutivos, teniendo en cuenta los datos censurados que se encuentran entre ellos [33]. Para ello, se emplea la siguiente expresión [49]:

$$\text{Nuevo incremento} = \frac{(n + 1) - q_{i-1}}{(n + 1) - \text{Número de elementos anteriores}} \quad (2.14)$$

Como ya se ha mencionado anteriormente, el parámetro n hace referencia al número total de elementos. Por otro lado, el número de orden ajustado de un elemento i (q_i) se obtiene de la suma del número de orden del dato de fallo anterior (q_{i-1}) y del incremento recién calculado [49]. Para el primer dato de fallo, el número de orden ajustado anterior es 0, por lo que el número de orden de dicho elemento es igual al incremento.

$$q_i = \text{Nuevo incremento} + q_{i-1} \quad (2.15)$$

A continuación, se hace uso de la expresión de Bernard para obtener una buena aproximación del rango de la mediana de cada elemento y, así, de la función de distribución, que más tarde se empleará para graficar los elementos [21].

$$RM_i = \frac{q_i - 0.3}{n + 0.4} \quad (2.16)$$

Llegados a este punto, se dispone de todos los datos necesarios para plasmar los datos de fallo en una gráfica, en la que el eje X representa la transformación logarítmica de las horas de funcionamiento registradas y el eje Y se define como $\ln(\ln(1/(1 - RM_i)))$. A continuación, se dibuja una línea de tendencia lineal para unir los puntos. El método se apoya en el de mínimos cuadrados para encontrar el mejor ajuste [33]. La figura 2.1 ilustra cómo quedaría la gráfica con los datos del ejemplo mostrado en el apéndice C del libro escrito por Abernethy y sus colaboradores [33].

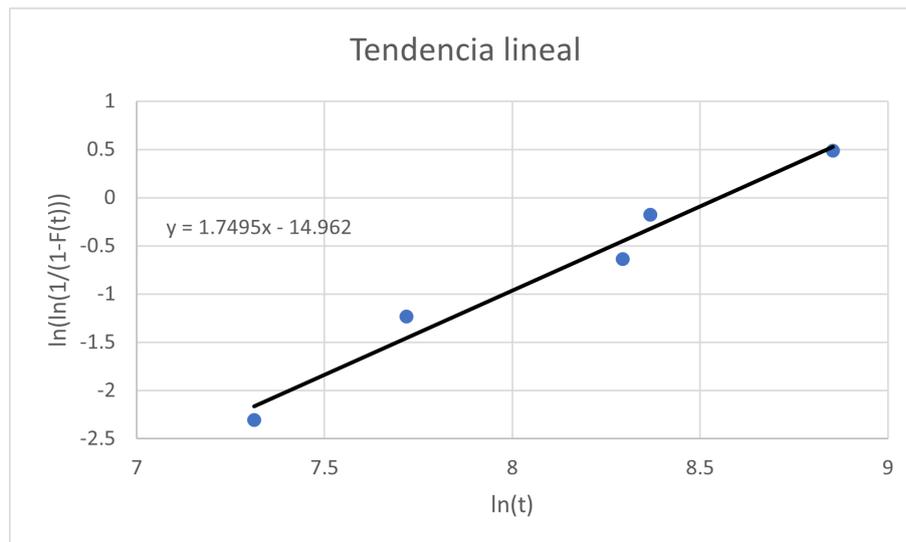


Figura 2.1 Ejemplo de gráfica obtenida con los datos facilitados en [33].

Los parámetros α y β se pueden obtener fácilmente de la gráfica. Mientras que el primero comparte valor con la pendiente de la línea de tendencia, el factor de escala se calcula mediante la exponencial de la división entre la ordenada en el origen (b) y la pendiente de la línea de tendencia [4].

$$\beta = \exp\left(\frac{-b}{\alpha}\right) \quad (2.17)$$

Por su parte, para calcular el valor de γ es necesario incluir otra línea de tendencia, en este caso, cuadrática. Para obtener el valor adecuado de γ , hay que incluirlo en la expresión del eje X, de forma que ahora se vea representado por $\ln(t_i - \gamma)$, indicando t_i las horas de funcionamiento de cierto elemento i antes de producirse el fallo. El γ óptimo es aquel que consiga hacer 0 el término cuadrático de la expresión de la nueva línea de tendencia cuadrática. Cabe indicar que para poder alcanzar dicho valor de γ , es necesario que el término cuadrático de la expresión sea negativo [4].

2.6 Método de Monte Carlo basado en cadenas de Markov

Existen multitud de métodos de Monte Carlo. En particular el método de Monte Carlo basado en cadenas de Markov es uno de los más usados en la estadística bayesiana, con el fin de obtener muestras de distribuciones de probabilidad específicas [35].

Este método presenta un buen desempeño, respecto al método de Monte Carlo ordinario, ante funciones más complicadas gracias al uso de las cadenas de Markov. Para poder generar dichas cadenas, se debe hacer uso de un algoritmo [35]. El más popular es el de Metrópolis-Hastings [14].

Una vez que las cadenas converjan, se dispondrá de una nueva muestra, ajustada a la función de densidad de la distribución uniforme correspondiente [14], y se podrán realizar los cálculos deseados [35].

2.7 Método BLUE

Este método parte de la regresión de mínimos cuadrados ordinarios [44]. Se habla del método BLUE cuando los estimadores obtenidos por mínimos cuadrados cumplen el teorema de Gauss-Markov, que indica que cuando un estimador presente determinadas condiciones (homocedasticidad, combinación lineal de variables observadas y no observadas, esperanza condicionada de las perturbaciones nula...) es el estimador lineal e

insesgado óptimo [50] y presenta una varianza mínima [27]. Precisamente, BLUE son las siglas en inglés para *Best Linear Unbiased Estimators*.

2.8 Método Gupta

Este método fue propuesto por A. K. Gupta en su trabajo [16], donde presentó unos estimadores lineales alternativos.

El método BLUE requiere que los valores exactos de la media, varianza y covarianza sean conocidos. Esto dificulta su uso en aquellas situaciones en las que se complique el cálculo de la covarianza o en las que muestra sea demasiado grande, dificultando la inversión de la matriz de varianza-covarianza [48]. En estas situaciones se puede recurrir al método de Gupta.

Los estimadores lineales alternativos se obtienen mediante la sustitución de la matriz de varianza-covarianza por una matriz unidad, \mathbf{I} [36], facilitando el cálculo de los estimadores cuando no se disponga de los parámetros requeridos para obtener los estimadores lineales e insesgados óptimos.

2.9 Estimación de Bayes

Existen una gran variedad de métodos bayesianos, empleados por los analistas en los estudios de fiabilidad [29].

Esta estimación consiste en combinar los datos reales de una muestra con las incertidumbres propias de las distribuciones *a priori* que se tengan en cuenta, con el fin de obtener nuevas distribuciones de parámetros conocidas como *a posteriori*, de las que se obtiene los estimadores. Esto provoca que la validez de los resultados esté ligada tanto a la validez del modelo, como de la distribución *a priori* [10] [20] [29].

2.10 Maximización de la restauración bayesiana

La maximización de la restauración bayesiana es un método desarrollado por M. Bacha y G. Celeux [52] basado en el concepto del *bootstrap* propuesto por B. Efron [11].

Este método es capaz de sobrellevar fuertes censuras en la muestra, ya que se calculan los estimadores a partir de una nueva muestra obtenida de la simulación de la muestra original y la distribución *a priori* [57]. En eso radica la ventaja de este método, en que tiene en cuenta la información *a priori* desde el primer paso [52].

Su funcionamiento es sencillo, el algoritmo realiza tres pasos [52]:

- Toma una muestra θ de la distribución *a priori* $\pi(\theta)$.
- Obtiene la información que falta, conociendo la muestra observada y el parámetro θ obtenido.
- Finalmente, obtiene el estimador correspondiente maximizando la verosimilitud de la muestra completa.

3 Antecedentes

El objetivo de este proyecto es estudiar cómo influyen los datos censurados en el cálculo del tiempo óptimo entre mantenimientos preventivos, así como en el comportamiento de la distribución Weibull. Esta sección pretende tanto presentar el marco en el que se desarrolla el modelo, como exponer el trasfondo matemático que posee.

3.1 Marco teórico

En el supuesto sobre el que se desarrolla el proyecto, los equipos pueden transitar entre tres estados diferentes, que representan el comportamiento del sistema ante el fallo. Estos son operativo, correctivo y preventivo.

El estado operativo supone que el equipo está operando en sus condiciones de diseño, produciendo ciertos ingresos. Los otros dos estados se dan dependiendo del momento en el que se dé el fallo del activo. De este modo, si transcurre un intervalo de tiempo predeterminado τ , conocido como intervalo preventivo, sin que se produzca el fallo, el equipo pasa al estado preventivo, caracterizado por los costes de la mano de obra y de las piezas de recambio derivados del mantenimiento preventivo, además de los originados por la inactividad del equipo. Por otro lado, si el fallo tiene lugar antes de alcanzar el tiempo τ , el equipo pasará al estado correctivo, donde quedará inoperativo hasta que sea reparado. Una vez finalizadas las acciones de mantenimiento, el equipo vuelve al estado operacional.

Cada vez que el equipo transita entre dos estados, se acumulan los costes o los ingresos en forma de retornos. El cálculo del intervalo preventivo óptimo se lleva a cabo maximizando el retorno medio acumulado. La figura 3.1 representa la transición entre los tres estados expuestos.

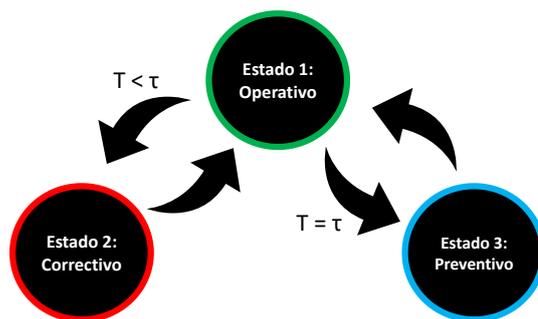


Figura 3.1 Transición entre los tres estados del modelo..

Como se ha comentado anteriormente, cada estado supone unos ingresos o unos costes. En este proyecto, se le denomina R_1 a los ingresos obtenidos durante el estado operativo. Por su parte, R_2 y R_3 representan los costes derivados de los estados correctivo y preventivo, respectivamente. Finalmente, cabe comentar que la transición entre estados supone también unos costes. Se representan como R_{ij} los gastos producidos de la transición de un estado i a un estado j .

3.2 Cálculo del intervalo preventivo

Una vez definido el modelo en el que se basa este proyecto, llega el momento de exponer la base matemática que existe detrás del cálculo del parámetro τ .

Como se ha comentado anteriormente, un equipo puede oscilar entre tres estados diferentes. En un estado i , y en la posterior transición a un estado j , se producen ciertos costes o ingresos. En un total de m transiciones, se genera un retorno acumulado desde el estado i denominado $Q_i(m)$. Cabe destacar que un sistema evoluciona de forma impredecible desde un estado inicial definido. Por ello, el parámetro de interés a la hora de realizar el desarrollo matemático es el retorno medio acumulado $v_i(m)$, el cual se encuentra representado por la figura 3.2. Dicho parámetro atiende a la siguiente expresión:

$$v_i(m) = v_i(1) + \sum_{j=1}^n v_j(m-1)p_{ij} \quad (3.1)$$

Se puede observar que la expresión 3.1 dispone de dos términos. El primero representa la transición desde el estado inicial i hasta el estado siguiente, j . El retorno en un solo paso $Q_i(1)$ es una variable aleatoria, que puede alcanzar diversos valores ($r_{i1}(1), r_{i2}(1), \dots, r_{in}(1)$). De este modo, el retorno medio en un solo paso es producto de la multiplicación de dichos valores por sus respectivas probabilidades ($p_{i1}, p_{i2}, \dots, p_{in}$).

$$v_i(1) = \sum_{j=1}^n r_{ij}(1) \cdot p_{ij} \quad (3.2)$$

Por su parte, el segundo término de la expresión 3.1 hace referencia al valor del retorno medio acumulado de las $m-1$ transiciones restantes donde, igualmente, se multiplican los valores aleatorios por sus probabilidades.

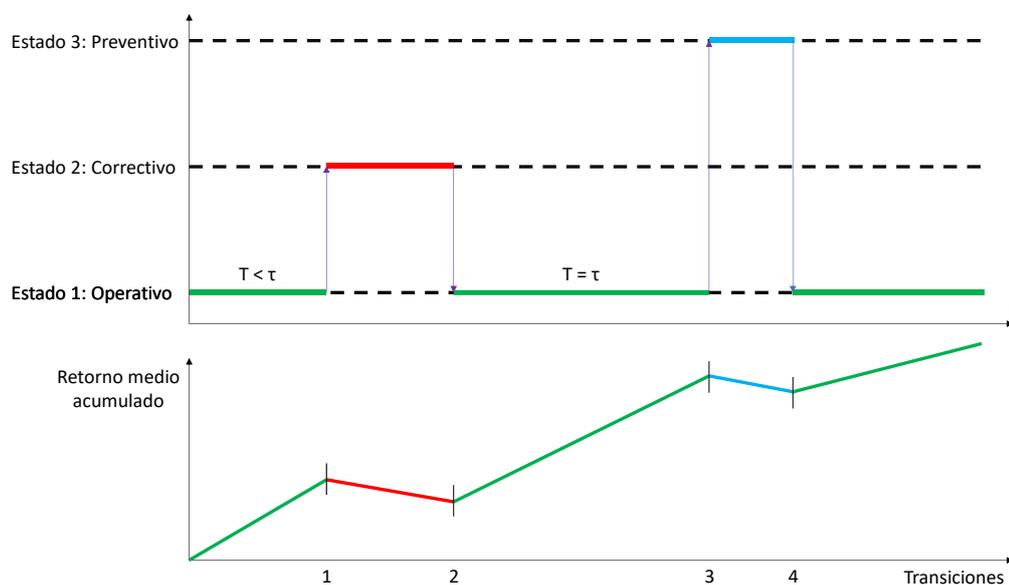


Figura 3.2 Evolución del retorno medio acumulado..

Finalmente, para poder calcular el retorno medio acumulado en m pasos desde cualquier estado i inicial, se puede expresar la ecuación 3.1 de forma matricial:

$$\mathbf{V}(m) = \mathbf{V}(1) + \mathbf{P}\mathbf{V}(m-1) \quad (3.3)$$

El parámetro \mathbf{P} es la matriz de probabilidad de transición entre estados, la cual atiende a la siguiente expresión:

$$\mathbf{P} = \begin{bmatrix} 0 & \mathbf{F}(\tau) & 1 - \mathbf{F}(\tau) \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad (3.4)$$

Al definir \mathbf{P} , aparece otra nueva matriz que definir: la de tiempo de permanencia \mathbf{F} . Cada componente F_{ij} representa el tiempo medio que el sistema permanece en el estado i , antes de pasar al j . De este modo, se tiene que:

$$\mathbf{F} = \begin{bmatrix} 0 & A & \tau \\ B & 0 & 0 \\ C & 0 & 0 \end{bmatrix} \quad (3.5)$$

De igual modo, en la expresión 3.5 aparecen otros tres términos en los que cabría profundizar. Dichos parámetros atienden a las siguientes expresiones:

$$A = \frac{1}{F(\tau)} \int_0^{\tau} t \cdot f(t) dt \quad (3.6)$$

$$B = \int_0^{\infty} t_c \cdot g(t_c) dt_c \quad (3.7)$$

$$C = \int_0^{\infty} t_p \cdot j(t_p) dt_p \quad (3.8)$$

Como se puede observar, las tres expresiones anteriores dependen de tres funciones de densidad de probabilidad. El parámetro A depende de la función $f(t)$, que está relacionada con la función de distribución acumulativa $F(t)$ de fallos del equipo. Por otro lado, el parámetro B depende de la función $g(t_c)$ que, a su vez, depende del tiempo necesario para finalizar la reparación, el cual está relacionado con la función de distribución $G(t_c)$. Por último, $j(t_p)$ es la función de densidad de probabilidad que hace acto de presencia en la expresión 3.8, la cual está relacionada con la función de distribución del tiempo en el que el equipo se encuentra bajo mantenimiento preventivo.

Disponiendo de todos estos parámetros, también se puede definir la matriz de retornos \mathbf{R} .

$$\mathbf{R} = \begin{bmatrix} 0 & r_{12} & r_{13} \\ r_{21} & 0 & 0 \\ r_{31} & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & R_1 A + R_{12} & R_1 \tau + R_{13} \\ R_2 B + R_{12} & 0 & 0 \\ R_3 C + R_{12} & 0 & 0 \end{bmatrix} \quad (3.9)$$

A continuación, se procede a resolver el sistema de ecuaciones diferencial que supone la expresión 3.3. Para ello, se aplica la transformada Z a ambos lados de la expresión. Por definición, la transformada Z de una secuencia cualquiera $x(y)$, es una función de variable compleja que se define mediante la siguiente serie de Laurent:

$$\mathcal{Z}[x(m)] = \sum_{n=0}^{\infty} x(n) z^{-n}; \quad z \in \mathbb{C} \setminus \{0\} \quad (3.10)$$

Después de aplicar la transformada Z a la expresión 3.3 y de hacer los cambios pertinentes, se alcanza la siguiente expresión:

$$\mathcal{Z}[\mathbf{V}(m)] = \frac{1}{z-1} (\mathbf{I} - z^{-1} \mathbf{P})^{-1} \mathbf{V}(1) \quad (3.11)$$

Llegados a este punto, para obtener el retorno medio acumulado de m transiciones solo quedaría revertir la transformación. Teniendo en cuenta que el modelo en el que se basa este proyecto oscila entre tres estados, el vector $\mathbf{V}(1)$ se puede expresar de la siguiente manera:

$$\mathbf{V}(1) = \begin{bmatrix} v_1(1) \\ v_2(1) \\ v_3(1) \end{bmatrix} = \begin{bmatrix} r_{12} F(\tau) + r_{13} [1 - F(\tau)] \\ v_2(1) \\ v_3(1) \end{bmatrix} \quad (3.12)$$

Como se observa la evolución del sistema desde el estado operacional, se realizan los cálculos con el primer componente del vector $\mathbf{V}(1)$. De este modo, para alcanzar definitivamente el valor óptimo de τ , hay

que obtener la expresión explícita de $v_1(1)$, derivarla respecto de τ e igualarla a cero. Por un lado, la expresión explícita de τ es la siguiente:

$$v_1(m) = \frac{1}{4} \left\{ (2m+1 + (-1)^{m-1}) \left[R_1 \int_0^\tau t f(t) dt + R_{12} F(\tau) + (R_1 \tau + R_{13}) [1 - F(\tau)] \right] + (2m-1 - (-1)^{m-1}) \left[\left(R_2 \int_0^\infty t_c g(t_c) dt_c + R_{21} \right) F(\tau) + \left(R_3 \int_0^\infty t_p j(t_p) dt_p + R_{31} \right) [1 - F(\tau)] \right] \right\} \quad (3.13)$$

Una vez definida la expresión 3.13, se procede a derivarla respecto de τ :

$$\frac{dv_1(m)}{d\tau} = \frac{1}{4} \left\{ f(\tau) [(2m+1 + (-1)^{m-1})(R_{12} - R_{13}) + (2m-1 - (-1)^{m-1})(R_2 B + R_{21} - R_3 C - R_{31})] + \frac{R_1}{4} [2m+1 + (-1)^{m-1}] [1 - F(\tau)] \right\} \quad (3.14)$$

Como se ha comentado al principio del capítulo, se hace uso de una distribución Weibull de tres parámetros para estudiar el modelo propuesto. La función de densidad de probabilidad $f(t)$ y la función de distribución acumulativa $F(t)$ atienden a las siguientes expresiones:

$$f(t) = \frac{\alpha}{\beta} \left[\left(\frac{t-\gamma}{\beta} \right)^{\alpha-1} \right] \exp \left[- \left(\frac{t-\gamma}{\beta} \right)^\alpha \right] \quad (3.15)$$

$$F(t) = 1 - \exp \left[- \left(\frac{t-\gamma}{\beta} \right)^\alpha \right] \quad (3.16)$$

Finalmente, teniendo en cuenta que los parámetros α , β y γ son mayores que 0, igualando la expresión 3.14 a cero y sustituyendo en ella las expresiones 3.15 y 3.16 se alcanza la expresión con la que se obtiene el valor óptimo del intervalo preventivo.

$$(\tau_0 - \gamma)^{\alpha-1} = \frac{\beta^\alpha}{\alpha} \cdot \frac{-R_1}{R_{12} - R_{13} + \frac{2m-1 - (-1)^{m-1}}{2m-1 + (-1)^{m-1}} \cdot (R_2 B + R_{21} - R_3 C - R_{31})} \quad (3.17)$$

Cabe destacar que la expresión 3.17 es independiente del número de transiciones para un valor de m par. No obstante, cuando m presenta un valor impar, el valor de τ_0 va descendiendo hasta alcanzar el valor independiente de m cuando $m \rightarrow \infty$. De igual modo, para que la expresión tenga una solución válida es necesario que el denominador de la fracción cuyo numerador es $-R_1$ sea negativo. De este modo, α y β han de ser positivos.

Este desarrollo matemático ha sido extraído de [1], [2], [3] y [4].

3.3 Teoría de fallo

Como se acaba de indicar, la expresión 3.17 permite calcular el intervalo óptimo entre mantenimientos preventivos. No obstante, el comportamiento de dicha expresión puede verse alterado por la naturaleza de la tasa de fallos que muestren los equipos que conformen el conjunto de datos de estudio [1].

La tasa de fallo se define como la probabilidad de que cierto elemento o equipo falle [32]. Una de las formas más comunes de representar el comportamiento de dicha tasa a lo largo del tiempo es la conocida curva de bañera, la cual se puede dividir en tres tramos: fallos infantiles, vida útil y desgaste [31].

Como se puede observar en la figura 3.3, la curva de bañera es, en realidad, resultado de la superposición de tres distribuciones de fallos, las cuales se corresponden con las tres regiones indicadas anteriormente [43].

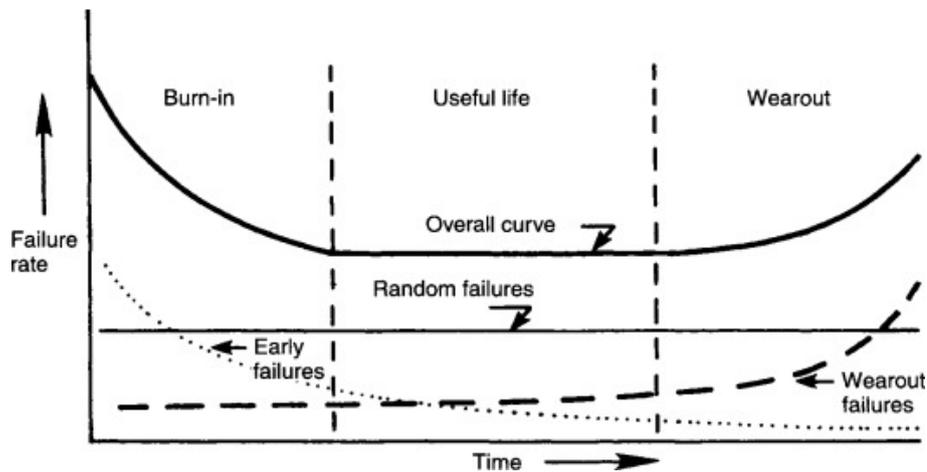


Figura 3.3 Curva de bañera [43].

Mientras que una tasa de fallos decreciente denota un aumento de la fiabilidad de los equipos, debido a la sustitución de los elementos más débiles que fallaron demasiado pronto, una tasa de fallos creciente supone que empiezan a predominar los efectos del desgaste en los equipos [49]. Por su parte, la tasa de fallo constante está relacionada con el desempeño de vida útil del equipo, donde los fallos ocurren de forma aleatoria [31].

La región de la curva de bañera en la que se encuentre el conjunto de datos influye directamente en los valores de los parámetros de Weibull, de ahí que afecte a los resultados de la expresión 3.17. En particular, es el parámetro de forma el que indica la región de la curva de bañera en la que se encuentra el conjunto de datos. Esto tiene sentido, ya que α está relacionada con la pendiente de la curva. [49]. De esta forma, se tiene que:

- $\alpha < 1$ representa una tasa de fallo decreciente, relacionada con fallos infantiles, de forma que una gran cantidad de fallos se dan al principio del marco temporal. Suelen estar relacionados con defectos de fabricación [49]. Como se ha comentado anteriormente, la tasa de fallo disminuye al sustituir dichos elementos defectuosos, demostrando los equipos un correcto desempeño tras la sustitución [32].
- $\alpha = 1$ está relacionado con una tasa de fallo constante. De este modo, los fallos que se dan lugar son aleatorios [49], a menudo relacionados con fluctuaciones fortuitas de tensión que superan la resistencia del componente [43].
- $\alpha > 1$ indica que la tasa de fallo es creciente, de modo que el desgaste en el elemento se está haciendo notar, dando lugar al final de la vida útil del mismo [49]. De este modo, valores de α superiores a la unidad indican el momento de realizar un mantenimiento preventivo, con el fin de alargar la vida útil del activo.

En lo referente al cálculo del intervalo óptimo entre mantenimientos preventivos, cabe destacar que se descarta el caso en el que $\alpha = 1$, ya que representa el caso en el que los fallos son aleatorios, por lo que no se podría hacer una estimación adecuada. Además, esto se encuentra matemáticamente reflejado en la expresión 3.17; ya que, si $\alpha = 1$, cualquier número puede ser solución o, por el contrario, ninguno, dependiendo del segundo término de la expresión [1].

Por otro lado, un valor de α inferior a la unidad supone que el máximo del retorno medio acumulado para m transiciones se obtiene para un τ_0 de valor infinito. De este modo, se concluye que no procede realizar mantenimientos preventivos en esta región de la curva de bañera [1]. Esto, además, tiene sentido, ya que, los fallos infantiles están relacionados con defectos de fabricación que provocan el fallo del equipo a las pocas horas de funcionamiento. En estas situaciones, corresponde sustituir el elemento defectuoso, lo cual está relacionado con el mantenimiento correctivo.

De este modo, el último caso que queda es aquel en el que $\alpha > 1$, donde la expresión 3.17 presenta un buen desempeño. De este modo, si se tiene un valor τ inferior al valor óptimo calculado, la función $v_1(m)$

presenta una tendencia creciente, que alcanza su máximo cuando $\tau = \tau_0$ y empieza a decrecer cuando $\tau > \tau_0$ [1]. Este comportamiento se encuentra ilustrado en la figura 3.4, mediante un ejemplo de diez transiciones.

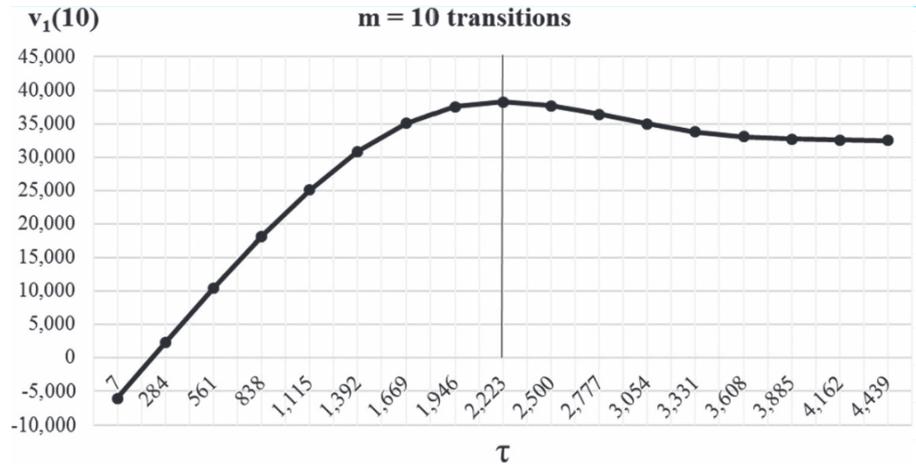


Figura 3.4 Comportamiento del retorno medio acumulado para diez transiciones [1].

Por lo tanto, los valores de α superiores a la unidad son los que permiten observar el comportamiento teórico de τ_0 según el valor de m , el cual fue expuesto al final de la sección 3.2.

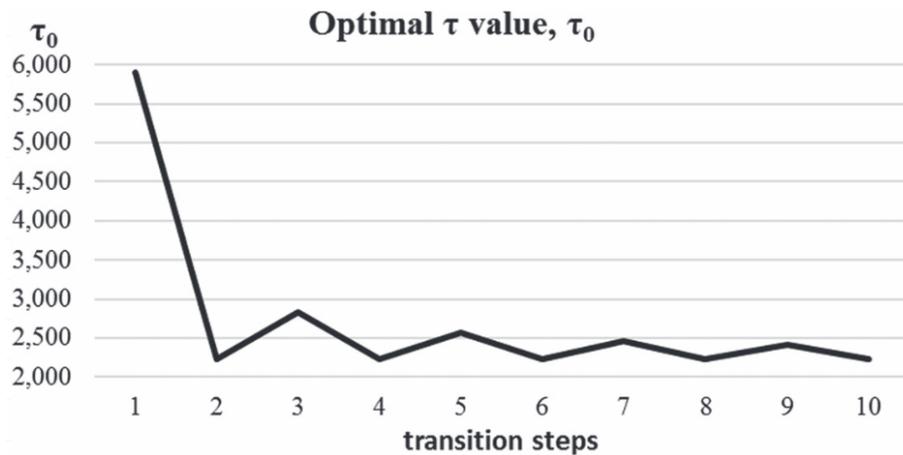


Figura 3.5 Comportamiento teórico de τ_0 para diez transiciones [1].

La figura 3.5 sigue con el ejemplo de las diez transiciones presentado en la figura 3.4. Como se puede observar, el valor óptimo de τ es constante para los valores pares de m , mientras que los impares presentan un intervalo óptimo superior, cuyo valor se va aproximando al del número de transiciones pares en cuanto más transiciones se llevan a cabo.

Por último, cabe comentar que tiene sentido que este sea el caso en el que se puede calcular el intervalo óptimo para el mantenimiento preventivo, ya que los equipos se encuentran en la región de desgaste y el objetivo del mantenimiento preventivo es poner a punto los equipos antes de que se produzca el fallo por desgaste.

A continuación, en el capítulo 4, se incluye una aplicación matemática para cada uno de los tres casos, con el fin de obtener una serie de conclusiones.

4 Aplicación matemática

En este capítulo, se desarrollan tres casos de estudio que demuestran la aplicabilidad real del desarrollo matemático y la expresión 3.17, basado tanto en la teoría de fallo, como en el marco teórico expuestos a lo largo del capítulo 3.

4.1 Base común a los diferentes casos de estudio

Todos los casos de estudios están ligados al marco expuesto en la sección 3.1. De este modo, se dispone de un conjunto de datos diferente para cada uno de los estudios realizados. Dichos datos agrupan información sobre la operatividad de ciertos equipos que se someten a un determinado periodo de observación, en el que se impone un intervalo entre mantenimientos preventivos de 2 000 h ($\tau = 2\,000$). Durante la toma de datos, todos los activos empezaron a ser observados en el estado de operación. Si registraban un fallo antes de llegar a las 2 000 h, se enviaban al mantenimiento correctivo. Por el contrario, si alcanzaban el valor de τ impuesto, se enviaban a mantenimiento preventivo, obteniendo así datos censurados por la derecha, ya que se desconoce cuándo podría haberse producido el fallo. Otro caso en el que pueden obtenerse datos censurados es aquel en el que el equipo estuviese funcionando hasta el fin del tiempo de análisis, sin registrar ningún fallo ni alcanzar las 2 000 h de funcionamiento. Los tres conjuntos de datos presentan 83 datos de fallo y 38 censurados.

El objetivo es obtener el intervalo óptimo entre mantenimientos preventivos y, para ello, se ajustan los conjuntos de datos a una distribución Weibull, tanto de dos, como de tres parámetros. Como se indicaba en la expresión 3.17, el valor de τ_0 depende de los parámetros de Weibull. Por ello, se hace uso de tres métodos matemáticos para encontrar el valor de α , β y, en el caso de tres parámetros, γ . Estos son la Estimación de Máxima Verosimilitud, el algoritmo de Newton-Raphson y el Método del Rango de Medianas.

La selección de métodos realizada se debe a lo expuesto a lo largo de la sección 1.2. De la revisión de la literatura se extrajo que la estimación de máxima verosimilitud es uno de los métodos más extendidos y de los que mejor comportamiento se registra en general. Por su parte, el método de Newton-Raphson presenta un método iterativo que mejora el comportamiento de la estimación de máxima verosimilitud cuando la función de verosimilitud logarítmica no puede maximizarse de forma analítica. Por último, el Método de Rango de Medianas ha sido históricamente muy utilizado por los ingenieros debido a su fácil programación. Todo esto es por lo que se considera interesante emplear estos tres métodos para obtener los parámetros de Weibull, comparando desempeños y resultados entre ellos.

Además de las variables de Weibull, para obtener los intervalos óptimos es necesario definir aquellos parámetros relacionados con el retorno de cada transición y estado, así como los parámetros B y C relacionados con los mantenimientos (correctivo y preventivo) de los equipos. Todos ellos fueron presentados a lo largo de la sección 3.2 y sus valores se encuentran reflejados en la tabla 4.1. Finalmente, se analizan los resultados obtenidos.

Tabla 4.1 Parámetros comunes a todos los casos de estudio.

Parámetro	Valor
Número de fallos	83
Número de datos censurados	38
τ	2 000 h
R_1	6 €/h
R_{12}	-4320 €
R_{13}	-1 €
R_2	-95 €/h
R_{21}	-620 €
R_3	-82 €/h
R_{31}	-620 €
B	8
C	7
m	1-10 transiciones
Métodos empleados	
MLE, Newton-Raphson y MRR	

Aunque anteriormente se indicó que el único caso que es realmente de relevancia a la hora de estudiar el dato del intervalo óptimo entre mantenimientos preventivos es el de desgaste, se incluyen dos casos de estudio más, en los que los conjuntos de datos se encuentran en las dos regiones restantes de la curva de bañera. Esto se hace con el objetivo de sacar más conclusiones, comparando el desempeño de los tres métodos en diferentes situaciones.

4.2 Caso de estudio 1: Desgaste

El primer caso que se expone es aquel que trata el tema principal de este proyecto: el cálculo de intervalos óptimos entre mantenimientos preventivos.

El conjunto de datos empleado para llevar a cabo este estudio se encuentra plasmado en las tablas 4.2 y 4.3. La primera de ellas contiene los datos de fallo de los equipos observados. Como se puede observar, una gran cantidad de fallos se registró pasadas las 1 000 h, con poca densidad de fallos a las pocas horas de actividad, lo que da a entender que dichos fallos son producto del desgaste del equipo debido a su uso. Por su parte, en la tabla 4.3 se pone de manifiesto que dos equipos consiguieron operar durante 2 000 h sin fallar, de forma que se les envió a mantenimiento preventivo.

Tabla 4.2 Datos de fallo para el caso de desgaste.

Horas de funcionamiento antes del fallo										
190	276	296	409	429	430	437	454	481	492	498
499	543	552	552	552	552	577	603	604	604	612
619	658	675	683	696	702	742	754	773	797	812
836	881	889	912	913	942	974	994	994	1 014	1 015
1 024	1 025	1 041	1 105	1 183	1 203	1 211	1 236	1 238	1 240	1 249
1 274	1 295	1 304	1 312	1 343	1 345	1 407	1 413	1 421	1 442	1 447
1 458	1 479	1 499	1 523	1 556	1 577	1 600	1 621	1 656	1 701	1 745
1 765	1 781	1 812	1 881	1 923	1 970					

Como se comentó en la sección 4.1, el estudio consiste en ajustar el conjunto de datos a una distribución Weibull y en hacer uso de varios métodos diferentes para calcular los parámetros α , β y γ . Estos tres métodos son la estimación de máxima verosimilitud, el método de Newton-Raphson y la regresión del rango de medianas. Los tres han sido implementados en Excel.

Tabla 4.3 Datos censurados para el caso de desgaste.

Horas de funcionamiento antes del fin de la observación										
45	84	84	103	121	176	199	203	207	211	215
219	223	508	514	519	638	655	661	735	760	764
766	769	790	867	909	986	1 011	1 111	1 167	1 188	1 188
1 395	1 396	1 505	2 000	2 000						

4.2.1 Estudio teniendo en cuenta los datos censurados

Para poder analizar correctamente la influencia de los datos censurados en el cálculo del intervalo óptimo entre mantenimientos preventivos, es necesario hacer dos estudios: uno que los tenga en cuenta y otro que no. En primer lugar, se incluye aquel que sí los considera.

Para resolver el análisis con la estimación de máxima verosimilitud, se implementa la expresión de la función de verosimilitud logarítmica y se maximiza haciendo uso del complemento *Solver*, que facilita Excel. Además, dicho complemento permite establecer restricciones en los valores de los parámetros, de forma que se impone que α y β son mayores que cero tanto en el ajuste de dos, como de tres parámetros. Referente a γ , se impone que su valor sea cero en el primer caso y que sea mayor que este en el segundo. Finalmente, en un breve espacio de tiempo de computación se alcanzan los siguientes resultados:

Tabla 4.4 Parámetros de Weibull obtenidos mediante el MLE para el caso de desgaste con censura.

Estimación de máxima verosimilitud			
Ajuste de Weibull	α	β	γ
Dos parámetros	2.48	1 281.35	0
Tres parámetros	2.16	1 156.11	113.36

Por otro lado, la figura 4.1 muestra los resultados obtenidos para el análisis mediante la regresión del rango de medianas. Recuperando la información plasmada en la sección 2.5, el valor de los parámetros de Weibull se obtienen a partir de una gráfica que muestra los datos de fallo, cuyo orden se encuentra influenciado por los datos censurados.

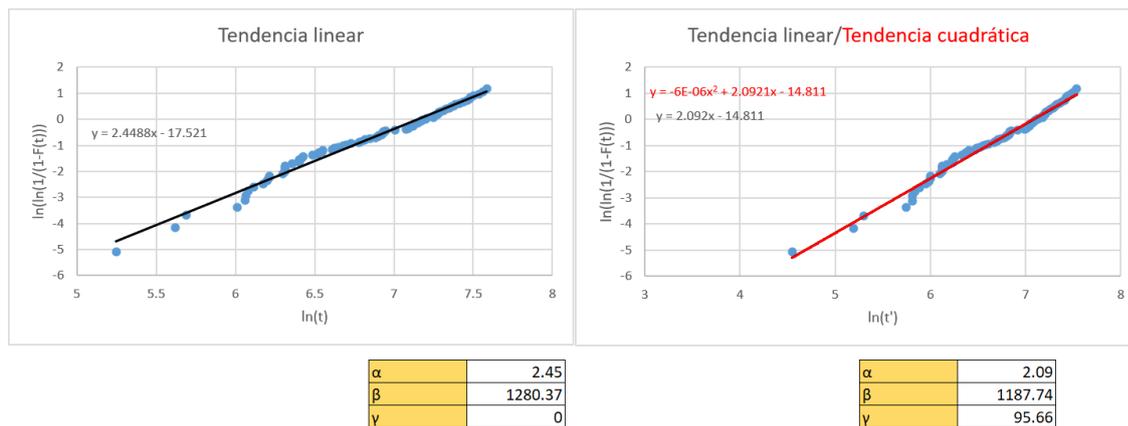


Figura 4.1 Resultados obtenidos mediante MRR para el caso de desgaste con censura.

La gráfica izquierda de la figura 4.1 representa el ajuste realizado para una distribución Weibull de dos parámetros: la única tendencia presente es lineal, cuya pendiente facilita el valor de α , mientras que el factor de escala se calcula con la expresión 2.17. Ambos valores se encuentran recogidos en una tabla debajo de la gráfica y se ha añadido la expresión de la tendencia lineal, para corroborar resultados.

A su derecha, aparecen representados los resultados de la distribución de tres parámetros. Como es de esperar, hace acto de presencia el parámetro γ , cuyo valor ya no es nulo, lo cual afecta directamente al valor de los otros dos parámetros. De igual modo, la expresión de la tendencia cuadrática pone de manifiesto que dicho valor de γ es el adecuado, ya que hace prácticamente nulo el término cuadrático.

Finalmente, queda comentar los resultados obtenidos con el método de Newton-Raphson. Como se ha comentado con anterioridad, este método no es más que una variante de la estimación de máxima verosimilitud que hace uso de un proceso iterativo para obtener los valores de los parámetros de Weibull a partir de la función de verosimilitud logarítmica. Por ello, haciendo uso del mismo conjunto de datos, se deberían alcanzar los mismos resultados o, en su defecto, unos muy similares.

El programa desarrollado al implementar este método, al contrario que con los otros dos, no calcula un valor de γ , en el caso del ajuste de tres parámetros, si no que calcula α y β a partir del método iterativo y de un valor de γ impuesto. En el caso del ajuste para dos parámetros, es tan fácil como imponer un γ de valor nulo. Por otro lado, lejos de suponer un problema en el ajuste de tres parámetros, esto presenta un nuevo estudio. Llegados a este punto, se tienen los resultados de dos algoritmos diferentes. Eso supone que se pueden introducir los dos γ obtenidos por los métodos anteriores para así estudiar cuál de ellos devuelve mejores resultados. De este modo, se pueden llegar incluso a combinar métodos para calcular el intervalo óptimo entre mantenimientos preventivos.

Según lo indicado en la sección 2.1.2, en cada paso, se calcula el nuevo valor del parámetro a calcular con la expresión 2.7. En este caso, el parámetro α sustituye al θ genérico de la expresión. Se considera que el algoritmo converge cuando se cumple la desigualdad de la expresión 2.8. Eso quiere decir que el algoritmo converge cuando se cumple lo siguiente:

$$\left| \frac{h(\alpha)}{h'(\alpha)} \right| < \varepsilon \quad (4.1)$$

De este modo, se puede comprobar qué valor de γ permite que el algoritmo converja antes. Las figuras 4.2 y 4.3 representan ciertas tablas que recogen los resultados de los cálculos realizados para cada iteración. Como se puede observar en ellas, en ambos casos la fracción $h(\alpha)/h'(\alpha)$ alcanza en la trigésima iteración un valor absoluto inferior al error impuesto ($\varepsilon = 10^{-5}$). Los parámetros N_1 , N_2 , N_3 , N_4 y N_5 son términos que se emplean para calcular $h(\alpha)$ y $h'(\alpha)$.

$$N_1 = \sum_{i=1}^n \delta_i t_i^\alpha \quad (4.2)$$

$$N_2 = \sum_{i=1}^n \delta_i t_i^\alpha \ln t_i \quad (4.3)$$

$$N_3 = \sum_{i=1}^n \delta_i t_i^\alpha \ln t_i \quad (4.4)$$

$$N_4 = \sum_{i=1}^n (1 - \delta_i) t_i^\alpha (\ln t_i)^2 \quad (4.5)$$

$$N_5 = \sum_{i=1}^n (1 - \delta_i) t_i^\alpha + N_1 \quad (4.6)$$

Como se puede observar en las expresiones anteriores [8] [58], los datos censurados no tienen influencia en las expresiones 4.2-4.4. Por el contrario, la expresión 4.5 excluye los datos de fallo. Finalmente, cabe comentar que la expresión 4.6 combina los dos tipos de datos.

Iteración	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
α	1	1.304756232	1.547113855	1.727646187	1.857823227	1.949985785	2.014470873	2.059234683	2.091046097	2.111399652	2.125985412	2.135975121	2.142808779	2.147479623	2.15067037
N_1	75812.76632	636028.8233	3495053.739	12509284.25	31468280.23	60551023.7	95783238.35	131728151.9	164167557.9	191021082.9	211951204.8	227597902	238963215.2	247057075.7	252743564.7
N_2	526762.1031	4457704.861	24640059.65	88533018.52	223284634.1	430375858.5	681572640.8	938070595.6	1169688843	1361499974	1511041539	1622854735	1704082874	1761934840	1802582175
N_3	3676390.318	31358346.76	174269107.3	628389599.3	1588564345	3066725450	4861786320	6696180061	8353534457	9726551292	10797263311	11597978292	12179735744	12594105514	12885261378
N_4	154981.0182	1265320.132	6841713.872	24254720.14	60668592.92	116336188.9	183636097.5	252211182.8	314050076	365215193.7	405081378.8	434877789.2	456518020.4	471927769.2	482753487.2
N_5	98671.83848	819698.2598	4478670.812	15975176.56	40103101.41	77065006.41	121803979.6	167422664.4	208577491	242637528	269180293.9	289020698.5	303431158.4	313693131.7	320902628.4
$h(\alpha)$	0.749786338	0.4435416	0.276072195	0.17617758	0.116666954	0.077639805	0.052103181	0.03515781	0.023809565	0.016163315	0.010990427	0.00748126	0.00509633	0.003473435	0.002368153
$h(\alpha)$	-2.46028222	-1.82996101	-1.52938136	-1.3644323	-1.26588234	-1.20399627	-1.16395768	-1.1375718	-1.11997812	-1.1081572	-1.10017489	-1.094766576	-1.091093982	-1.088596262	-1.086895821
$h(\alpha)/h'(\alpha)$	-0.30475623	-0.24237762	-0.18051233	-0.13017704	-0.09216256	-0.06448509	-0.04476381	-0.03090601	-0.02125896	-0.01458576	-0.009989709	-0.006833658	-0.004670844	-0.003190747	-0.002178823

Iteración	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
α	2.152849193	2.154336625	2.155351877	2.156044756	2.156517586	2.156840232	2.157060389	2.157216099	2.157313106	2.157383041	2.157430758	2.157466331	2.15748553	2.15750068	2.15751103
N_1	256701816.3	259439630.5	261325156.4	262619836.9	263507032	264114154.5	264529227.2	264812818.6	265006492.6	265138719.4	265228976	265290576	265332613	265361299	265380872
N_2	1830877105	1850448609	1863927467	1871182026	1879525225	1883865460	1886832771	1888660140	1890244701	1891189982	1891835221	1892275593	1892576116	1892781189	1892921121
N_3	13087944764	13228143691	13324700082	13391002034	13436436990	13467529241	13488786316	13503309926	13513228618	13520000411	13524622773	135288110	13531410	13531410	13531410
N_4	490288733.1	495500544.1	499089771.1	501554276.8	503243093.3	504398771.3	505188873.1	505728695.1	506097356.7	506349052.3	506520856.9	506638112	506771311	506772735	506809993
N_5	325920902	329391876.3	331782260.1	333423602.6	334548346.6	335318025	335844231.7	336203752.8	336449281.5	336616910.9	336731322.9	336809425	336836718	336899084	336923898
$h(\alpha)$	0.00161496	0.001101496	0.000751365	0.000512567	0.000349682	0.000238567	0.000162764	0.000111048	7.57652E-05	5.1693E-05	3.52693E-05	2.4064E-05	1.6418E-05	1.1202E-05	7.643E-06
$h(\alpha)$	-1.085737355	-1.08497749	-1.084402231	-1.08379181	-1.08326099	-1.08285045	-1.08242361	-1.082037013	-1.081693708	-1.08138046	-1.08109372	-1.08082826	-1.0805813901	-1.080342146	-1.080114651
$h(\alpha)/h'(\alpha)$	-0.001487432	-0.001015252	-0.000692879	-0.00047283	-0.00032265	-0.00022016	-0.00015022	-0.0001025	-6.99347E-05	-4.77166E-05	-3.2557E-05	-2.2213E-05	-1.5156E-05	-1.0341E-05	-7.0555E-06

Figura 4.2 Iteraciones del método de Newton con $\gamma = 113.36$, para el caso de desgaste con censura.

Iteración	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
α	1	1.313112666	1.56445196	1.753190012	1.89030951	1.988031962	2.056813089	2.104814473	2.138117676	2.161129352	2.176985031	2.187888751	2.195377011	2.200514912	2.204037928
N_1	77282.635	690258.638	4053876.492	15418700.37	40825375.5	81840030.08	133616829.6	188184630.2	238691266.3	281328479.8	315071949	340602851.6	359327663.3	372768713.3	382275226.2
N_2	538040.5485	4847234.406	28636716.67	109347557.7	290286416.8	582936693.5	952856834.8	1343062513	170446013	2009648744	2251257304	2434098028	2568214416	2664494629	2732595410
N_3	3761999.322	34161258.74	202919200.1	777643005.2	2069394600	4162342515	6811071957	9607339059	12198512638	14387700201	16121187867	17433264898	18395811195	19068687389	19575694227
N_4	159008.2387	1380231.504	7968818.048	30003637.8	78964951.69	157719170.1	256922951.2	361336259.5	457899459	539374837.7	603833570.5	65293477.5	688349644.9	714013648.5	732163954.8
N_5	100687.675	890220.3332	5197232.915	19697011.31	52039365.38	104176899.9	169936136	239199947.7	303284531.8	357370562.2	400167559.5	432544874.7	456289158.1	473332406.3	485386238.8
$h(\alpha)$	0.764086368	0.45309571	0.282894023	0.182618093	0.12036493	0.08037949	0.054130279	0.036653701	0.024909912	0.016969956	0.01157975	0.007910395	0.005407837	0.00369889	0.002530876
$h(\alpha)$	-2.44029211	-1.80270929	-1.4988892	-1.33181711	-1.2317019	-1.16862712	-1.12768163	-1.10606587	-1.08249007	-1.07027626	-1.061999968	-1.056372876	-1.052538299	-1.049921214	-1.048133198
$h(\alpha)/h'(\alpha)$	-0.31311267	-0.25134153	-0.18873582	-0.1371195	-0.09772245	-0.06878113	-0.04800138	-0.0333032	-0.02301168	-0.01585568	-0.01090372	-0.00748826	-0.005137901	-0.003523016	-0.002414651

Iteración	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
α	2.206452579	2.208107069	2.209240473	2.2100168	2.210548496	2.210912622	2.211161979	2.211332736	2.211449666	2.211529735	2.211584563	2.21162211	2.21164782	2.21166542	2.21167747
N_1	388930889.1	393558196.7	396759928.1	398968009.3	400487393.8	401531273.9	402247705.6	402739048.5	403075854.2	403306649.4	403464764.5	403573070	403647249	403698051	403732841
N_2	2780275894	2813426417	2836364418	2852183879	2863069379	2870548216	2875681085	2879201317	2881614371	2883267912	2884400735	2885176695	2885708154	2886072125	2886321379
N_3	19917954792	20155923908	20320586073	20434148697	20512292743	20565981603	20602829440	20628100558	20645423470	20657293982	20665426344	20671E+10	20675E+10	20677E+10	20679E+10
N_4	744870712.5	753704738.9	759817061.9	764032380.1	766932920.1	768925698.1	770293367.6	771231339.3	771874298.7	772314884.3	772616724	772823478	772965084	773062064	773128477
N_5	493825118.9	499692101	503751543.9	506551126.4	508477513.9	509801017.1	510709356.4	511333212.3	511759335.6	512051951.9	512252419.8	512389736	512483785	512548194	512592303
$h(\alpha)$	0.001732103	0.001185526	0.000811652	0.00055568	0.000380455	0.000260493	0.000178361	0.00012127	8.36235E-05	5.7296E-05	3.92076E-05	2.6847E-05	1.8383E-05	1.2588E-05	8.6193E-06
$h(\alpha)$	-1.046910743	-1.046074552	-1.045502387	-1.045110793	-1.04484274	-1.04465924	-1.04453361	-1.04444759	-1.044388693	-1.044348367	-1.04432076	-1.04430185	-1.0442899	-1.04428004	-1.0442797
$h(\alpha)/h'(\alpha)$	-0.00165449	-0.001133404	-0.000776327	-0.000531695	-0.00036413	-0.00024936	-0.00017076	-0.00011693	-8.00694E-05	-5.48281E-05	-3.7544E-05	-2.5708E-05	-1.7603E-05	-1.2054E-05	-8.2538E-06

Figura 4.3 Iteraciones del método de Newton con $\gamma = 95.66$, para el caso de desgaste con censura.

La tabla 4.5 agrupa los valores de los parámetros de Weibull obtenidos con el método de Newton-Raphson, para los dos valores de γ calculados anteriormente.

Tabla 4.5 Resultados del estudio de velocidad de convergencia de α con el desgaste con datos censurados.

Estudio de la velocidad de convergencia		
γ calculada con	MLE	MRR
Iteración de convergencia	30	30
α	2.16	2.21
β	1 156.11	1 175.92
γ	113.36	95.66

Como se puede observar, los dos valores de γ alcanzan la convergencia en la misma iteración. Además, al ser más altos los valores de los parámetros de Weibull, los valores de los intervalos óptimos serán más altos, por lo que se decide mantener el valor de γ proveniente del método de máxima verosimilitud, ya que ambos emplean la misma función para realizar los cálculos.

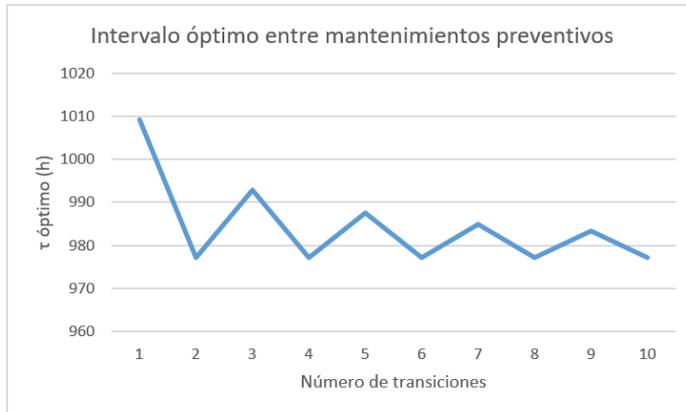
Tabla 4.6 Resumen de los resultados obtenidos para el caso de desgaste con datos censurados.

Método	α	β	γ
MLE	2.16	1 156.11	113.36
Newton-Raphson	2.16	1 156.11	113.36
MRR	2.09	1 187.74	95.66

La tabla 4.6 ofrece un resumen de los resultados obtenidos con los tres métodos a lo largo de este estudio. Como se ha comentado anteriormente, la estimación de máxima verosimilitud y el método de Newton-

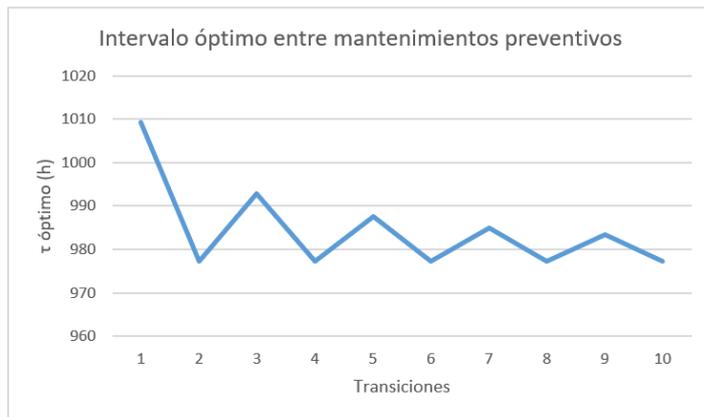
Raphson comparten resultados. La leve diferencia que hay entre los resultados de esos dos métodos y los de la regresión de rango de medianas va a dar lugar a que se obtengan resultados distintos a la hora de calcular el intervalo óptimo entre mantenimientos preventivos. Como este cálculo se realiza teniendo en cuenta el ajuste de 3 parámetros, la tabla 4.6 incluye únicamente los resultados de dicho ajuste.

m	τ óptimo (h)
1	1009.25507
2	977.208686
3	992.917107
4	977.208686
5	987.612881
6	977.208686
7	984.986555
8	977.208686
9	983.418912
10	977.208686



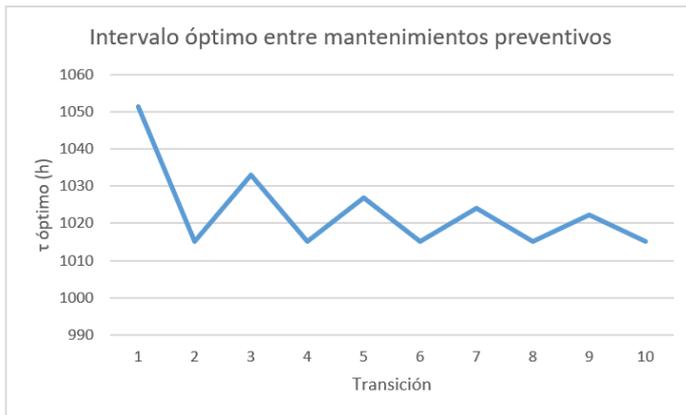
(a) Intervalo óptimo entre mantenimientos preventivos según la estimación de máxima verosimilitud.

m	τ óptimo (h)
1	1009.253165
2	977.2069512
3	992.9152873
4	977.2069512
5	987.6110903
6	977.2069512
7	984.9847778
8	977.2069512
9	983.4171435
10	977.2069512



(b) Intervalo óptimo entre mantenimientos preventivos según el método de Newton-Raphson.

m	τ óptimo (h)
1	1051.380553
2	1015.18247
3	1032.916075
4	1015.18247
5	1026.925869
6	1015.18247
7	1023.960693
8	1015.18247
9	1022.191047
10	1015.18247



(c) Intervalo óptimo entre mantenimientos preventivos según la regresión del rango de medianas.

Figura 4.4 Intervalos óptimos entre mantenimientos preventivos para el caso de desgaste con censura.

La figura 4.4 reúne los valores de intervalos óptimos entre mantenimientos preventivos para el conjunto de datos analizado, según cada uno de los métodos empleados. Como se indicó en la tabla 4.1, el estudio se hace para un intervalo de entre 1 y 10 transiciones.

Como se puede observar en los tres casos, las transiciones pares definen un valor de τ_0 mínimo, mientras que las transiciones impares van disminuyendo su valor, dándose lugar a una oscilación que se va amortiguando con cada transición. Esto coincide con el comportamiento teórico que se mostraba en la figura 3.5.

Referente a los valores obtenidos, se puede apreciar una despreciable diferencia entre los resultados que facilita la estimación de máxima verosimilitud y los del método de Newton-Raphson. Esto se debe a que, realmente, los parámetros de Weibull calculados en cada método diferían entre sí en decimales, del orden de diezmilésimas, lo que se traduce en discrepancias del orden de centésimas en algunos valores de τ_0 . De cualquier modo, estas diferencias se consideran irrelevantes.

Mayores diferencias se pueden notar al comparar con los resultados obtenidos en el análisis con la regresión del rango de medianas. En este caso, los valores mínimos, relacionados con las transiciones pares, presentan un valor superior que el valor más alto registrado en los otros dos estudios (el de $m = 1$).

Posiblemente, estas discrepancias vengan de la forma que tienen de tratar los datos cada uno de los métodos. Mientras que la estimación de máxima verosimilitud y, por supuesto, el método de Newton-Raphson integran los datos censurados en la función de verosimilitud y con ello calculan los parámetros, la regresión del rango de medianas calcula los parámetros según los datos de fallo, donde los datos censurados se tienen en cuenta sólo para modificar el rango de los datos de fallo.

De entre los tres casos, podría considerarse que los resultados obtenidos por los métodos que hacen uso de la función de verosimilitud son más adecuados, por varios motivos. El primero de ellos es que τ_0 tiene unos valores más bajos, lo cual da menos lugar a que puedan producirse fallos por desgaste, ya que se hacen antes los mantenimientos preventivos. Por otro lado, la información recogida y plasmada en la sección 1.2 indicaba que la regresión de rango de medianas es más imprecisa que la estimación de máxima verosimilitud.

4.2.2 Estudio sin tener en cuenta los datos censurados

Como se ha comentado anteriormente, es necesario complementar el estudio de la sección anterior con otro en el que no se consideren los datos censurados, con el fin de poder vislumbrar correctamente la influencia de este tipo de datos en el estudio del intervalo óptimo entre mantenimientos preventivos. De este modo, en esta sección sólo se hará uso de los datos plasmados en la tabla 4.2.

De nuevo, para obtener el valor de los parámetros de Weibull haciendo uso de la estimación de máxima verosimilitud se emplea el complemento *Solver* de Excel, imponiendo las restricciones correspondientes.

Tabla 4.7 Parámetros de Weibull obtenidos mediante el MLE para el caso de desgaste sin censura.

Estimación de máxima verosimilitud			
Ajuste de Weibull	α	β	γ
Dos parámetros	2.45	1 161.05	0
Tres parámetros	2.13	1 032.36	113.21

Por otro lado, la regresión del rango de medianas sigue el mismo procedimiento que el narrado anteriormente. La única diferencia es que al no disponer de datos censurados, los datos de fallo no ven su número de orden alterado, ya que el valor del incremento se mantiene constante en 1.

Con el fin de ilustrar las diferencias entre ambos estudios, se incluyen las tablas 4.8 y 4.9. Como se puede observar en la primera de ellas, el número de orden final del estudio que no tiene en cuenta la censura coincide con el número de fallos. En cambio, en la tabla 4.9 se refleja la diferencia existente entre números de orden debido al ajuste de la expresión 2.15.

Tabla 4.8 Incremento, número de orden ajustado y RM_i de los datos de fallo sin tener en cuenta la censura.

Orden	Horas de funcionamiento	Tipo	Incremento	Nº de orden	RM_i
1	190	F	1	1	0.011400651
2	276	F	1	2	0.027687296
3	296	F	1	3	0.043973941
4	409	F	1	4	0.060260586
5	429	F	1	5	0.076547231
6	430	F	1	6	0.092833876
7	437	F	1	7	0.109120521
8	454	F	1	8	0.125407166
9	481	F	1	9	0.141693811
10	492	F	1	10	0.157980456
11	498	F	1	11	0.174267101
12	499	F	1	12	0.190553746
13	543	F	1	13	0.206840391
14	552	F	1	14	0.223127036
15	552	F	1	15	0.239413681
16	552	F	1	16	0.255700326
17	552	F	1	17	0.271986971
18	577	F	1	18	0.288273616
19	603	F	1	19	0.304560261
20	604	F	1	20	0.320846906
21	604	F	1	21	0.33713355
22	612	F	1	22	0.353420195
23	619	F	1	23	0.36970684
24	658	F	1	24	0.385993485
25	675	F	1	25	0.40228013
26	683	F	1	26	0.418566775
27	696	F	1	27	0.43485342
28	702	F	1	28	0.451140065
29	742	F	1	29	0.46742671
30	754	F	1	30	0.483713355
31	773	F	1	31	0.5
32	797	F	1	32	0.516286645
33	812	F	1	33	0.53257329
34	836	F	1	34	0.548859935
35	881	F	1	35	0.56514658
36	889	F	1	36	0.581433225
37	912	F	1	37	0.59771987
38	913	F	1	38	0.614006515
39	942	F	1	39	0.63029316
40	974	F	1	40	0.646579805
41	994	F	1	41	0.66286645
42	994	F	1	42	0.679153094
43	1014	F	1	43	0.695439739
44	1015	F	1	44	0.711726384
45	1024	F	1	45	0.728013029
46	1025	F	1	46	0.744299674
47	1041	F	1	47	0.760586319
48	1105	F	1	48	0.776872964
49	1183	F	1	49	0.793159609
50	1203	F	1	50	0.809446254
51	1211	F	1	51	0.825732899

Continúa en la siguiente página

Tabla 4.8 – Continúa desde la página anterior

Orden	Horas de funcionamiento	Tipo	Incremento	Nº de orden	RM _i
52	1236	F	1	52	0.842019544
53	1238	F	1	53	0.858306189
54	1240	F	1	54	0.874592834
55	1249	F	1	55	0.890879479
56	1274	F	1	56	0.907166124
57	1295	F	1	57	0.923452769
58	1304	F	1	58	0.939739414
59	1312	F	1	59	0.956026059
60	1343	F	1	60	0.972312704
61	1345	F	1	61	0.988599349

Tabla 4.9 Incremento, número de orden ajustado y RM_i de los datos de fallo teniendo en cuenta la censura.

Orden	Horas de funcionamiento	Tipo	Incremento	Nº de orden	RM _i
1	45	R	-	-	-
2	84	R	-	-	-
3	84	R	-	-	-
4	103	R	-	-	-
5	121	R	-	-	-
6	176	R	-	-	-
7	190	F	1.0517	1.0517	0.00619
8	199	R	-	-	-
9	203	R	-	-	-
10	207	R	-	-	-
11	211	R	-	-	-
12	215	R	-	-	-
13	219	R	-	-	-
14	223	R	-	-	-
15	276	F	1.1198	2.1726	0.01541
16	296	F	1.1198	3.2915	0.02464
17	409	F	1.119891443	4.411398467	0.033866544
18	429	F	1.119891443	5.531289911	0.04309135
19	430	F	1.119891443	6.651181354	0.052316156
20	437	F	1.119891443	7.771072797	0.061540962
21	454	F	1.119891443	8.89096424	0.070765768
22	481	F	1.119891443	10.01085568	0.079990574
23	492	F	1.119891443	11.13074713	0.08921538
24	498	F	1.119891443	12.25063857	0.098440186
25	499	F	1.119891443	13.37053001	0.107664992
26	508	R	-	-	-
27	514	R	-	-	-
28	519	R	-	-	-
29	543	F	1.155632659	14.52616267	0.117184207
30	552	F	1.155632659	15.68179533	0.126703421
31	552	F	1.155632659	16.83742799	0.136222636
32	552	F	1.155632659	17.99306065	0.14574185
33	552	F	1.155632659	19.14869331	0.155261065
34	577	F	1.155632659	20.30432597	0.16478028
35	603	F	1.155632659	21.45995863	0.174299494
36	604	F	1.155632659	22.61559129	0.183818709
37	604	F	1.155632659	23.77122395	0.193337924

Continúa en la siguiente página

Tabla 4.9 – Continúa desde la página anterior

Orden	Horas de funcionamiento	Tipo	Incremento	Nº de orden	RM _i
38	612	F	1.155632659	24.92685661	0.202857138
39	619	F	1.155632659	26.08248927	0.212376353
40	638	R	-	-	-
41	655	R	-	-	-
42	658	F	1.184166799	27.26665607	0.22213061
43	661	R	-	-	-
44	675	F	1.199156252	28.46581232	0.232008339
45	683	F	1.199156252	29.66496857	0.241886067
46	696	F	1.199156252	30.86412482	0.251763796
47	702	F	1.199156252	32.06328108	0.261641525
48	735	R	-	-	-
49	742	F	1.215361067	33.27864214	0.271652736
50	754	F	1.215361067	34.49400321	0.281663947
51	760	R	-	-	-
52	764	R	-	-	-
53	766	R	-	-	-
54	769	R	-	-	-
55	73	F	1.286852894	35.7808561	0.292264054
56	790	R	-	-	-
57	797	F	1.306350665	37.08720677	0.303024767
58	812	F	1.306350665	38.39355743	0.313785481
59	836	F	1.306350665	39.6999081	0.324546195
60	867	R	-	-	-
61	881	F	1.327420837	41.02732893	0.335480469
62	889	F	1.327420837	42.35474977	0.346414743
63	909	R	-	-	-
64	912	F	1.349919495	43.70466927	0.357534343
65	913	F	1.349919495	45.05458876	0.368653944
66	942	F	1.349919495	46.40450826	0.379773544
67	974	F	1.349919495	47.75442775	0.390893145
68	986	R	-	-	-
69	994	F	1.374918005	49.12934576	0.402218664
70	994	F	1.374918005	50.50426376	0.413544183
71	1011	R	-	-	-
72	1014	F	1.401877181	51.90614094	0.425091771
73	1015	F	1.401877181	53.30801812	0.436639359
74	1024	F	1.401877181	54.70989531	0.448186947
75	1025	F	1.401877181	56.11177249	0.459734534
76	1041	F	1.401877181	57.51364967	0.471282122
77	1105	F	1.401877181	58.91552685	0.48282971
78	1111	R	-	-	-
79	1167	R	-	-	-
80	1183	F	1.467080771	60.38260762	0.494914396
81	1188	R	-	-	-
82	1188	R	-	-	-
83	1203	F	1.540434809	61.92304243	0.507603315
84	1211	F	1.540434809	63.46347724	0.520292234
85	1236	F	1.540434809	65.00391205	0.532981154
86	1238	F	1.540434809	66.54434686	0.545670073
87	1240	F	1.540434809	68.08478167	0.558358992
88	1249	F	1.540434809	69.62521648	0.571047912
89	1274	F	1.540434809	71.16565129	0.583736831

Continúa en la siguiente página

Tabla 4.9 – Continúa desde la página anterior

Orden	Horas de funcionamiento	Tipo	Incremento	Nº de orden	RM _i
90	1295	F	1.540434809	72.7060861	0.59642575
91	1304	F	1.540434809	74.24652091	0.60911467
92	1312	F	1.540434809	75.78695572	0.621803589
93	1343	F	1.540434809	77.32739052	0.634492508
94	1345	F	1.540434809	78.86782533	0.647181428
95	1395	R	-	-	-
96	1396	R	-	-	-
97	1407	F	1.658929795	80.52675513	0.660846418
98	1413	F	1.658929795	82.18568492	0.674511408
99	1421	F	1.658929795	83.84461472	0.688176398
100	1442	F	1.658929795	85.50354451	0.701841388
101	1447	F	1.658929795	87.16247431	0.715506378
102	1458	F	1.658929795	88.8214041	0.729171368
103	1479	F	1.658929795	90.4803339	0.742836358
104	1499	F	1.658929795	92.13926369	0.756501348
105	1505	R	-	-	-
106	1523	F	1.7565139	93.89577759	0.770970161
107	1556	F	1.7565139	95.65229149	0.785438974
108	1577	F	1.7565139	97.40880539	0.799907787
109	1600	F	1.7565139	99.16531929	0.8143766
110	1621	F	1.7565139	100.9218332	0.828845413
111	1656	F	1.7565139	102.6783471	0.843314226
112	1701	F	1.7565139	104.434861	0.85778304
113	1745	F	1.7565139	106.1913749	0.872251853
114	1765	F	1.7565139	107.9478888	0.886720666
115	1781	F	1.7565139	109.7044027	0.901189479
116	1812	F	1.7565139	111.4609166	0.915658292
117	1881	F	1.7565139	113.2174305	0.930127105
118	1923	F	1.7565139	114.9739444	0.944595918
119	1970	F	1.7565139	116.7304583	0.959064731
120	2000	R	-	-	-
121	2000	R	-	-	-

A continuación, al igual que con el estudio anterior, se incluye la figura 4.5, que plasma tanto los valores de Weibull obtenidos, así como las gráficas de donde provienen. De nuevo, se puede observar cómo el valor de γ hace cero el término cuadrático y que eso disminuye el valor de α y β .

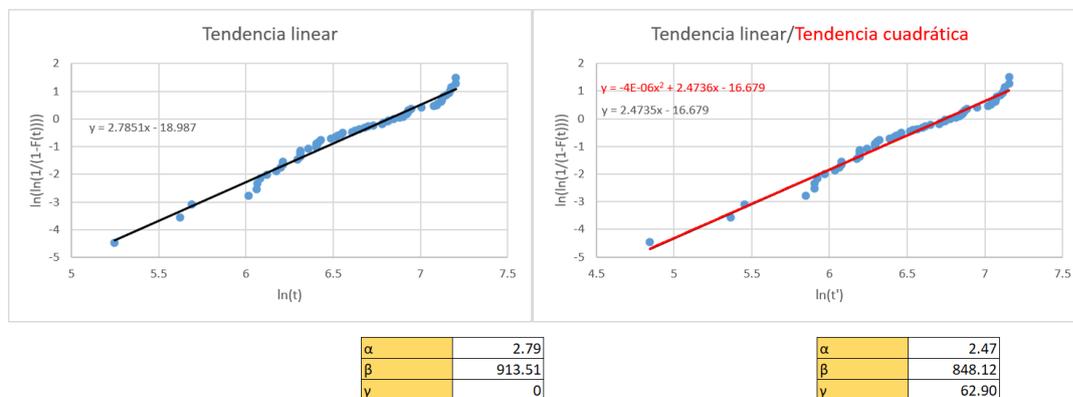


Figura 4.5 Resultados obtenidos mediante MRR para el caso de desgaste sin censura.

Finalmente, se añaden los resultados obtenidos con el método de Newton-Raphson. De nuevo, se realiza el estudio de velocidad de convergencia valiéndose de los valores de γ calculados con los otros dos métodos.

Iteración	1	2	3	4	5	6	7	8	9	10
α	1	1.58487442	2.007376139	2.121239078	2.126940304	2.126953311	2.126953311	2.126953311	2.126953311	2.126953311
N_1	75826.2114	4562063.66	91093929.71	204960136.4	213462232.8	213482028.4	213482028.6	213482028.6	213482028.6	213482028.6
N_2	526865.081	32190093.9	648132985.5	1461106259	1521857698	1521999153	1521999154	1521999154	1521999154	1521999154
N_3	3677170.7	227846824	4622788401	10439831588	10874847324	10875860267	10875860273	10875860273	10875860273	10875860273
N_4	0	0	0	0	0	0	0	0	0	0
N_5	75 826	4 562 064	91 093 930	204 960 136	213 462 233	213 482 028	213 482 029	213 482 029	213 482 029	213 482 029
$h(\alpha)$	0.71092176	0.23417266	0.042412014	0.001934459	4.39332E-06	2.27569E-11	0	0	0	0
$h'(\alpha)$	-1.21551181	-0.55425256	-0.37248304	-0.339305751	-0.337766375	-0.337762875	-0.337762875	-0.337762875	-0.337762875	-0.337762875
$h(\alpha)/h'(\alpha)$	-0.58487442	-0.42250172	-0.11386294	-0.005701225	-1.3007E-05	-6.73754E-11	0	0	0	0

Figura 4.6 Iteraciones del método de Newton con $\gamma = 113.21$, para el caso de desgaste con censura.

Iteración	1	2	3	4	5	6	7	8	9	10
α	1	1.62439055	2.115074459	2.266887756	2.276550163	2.276584979	2.27658498	2.27658498	2.27658498	2.27658498
N_1	80001.715	6501952.9	214657848.7	637050090.9	682776511.2	682947087.8	682947090	682947090	682947090	682947090
N_2	559008.729	46123493.5	1536301579	4569951736	4898670250	489896573	489896589	489896589	489896589	489896589
N_3	3921878.55	328134471	11019687348	32850322046	35217895780	35226728828	35226728942	35226728942	35226728942	35226728942
N_4	0	0	0	0	0	0	0	0	0	0
N_5	80 002	6 501 953	214 657 849	637 050 091	682 776 511	682 947 088	682 947 090	682 947 090	682 947 090	682 947 090
$h(\alpha)$	0.74792199	0.25720604	0.051199537	0.002900305	1.03761E-05	1.33726E-10	0	0	0	0
$h'(\alpha)$	-1.19784322	-0.52417867	-0.33725331	-0.300163798	-0.298020043	-0.298012361	-0.298012361	-0.298012361	-0.298012361	-0.298012361
$h(\alpha)/h'(\alpha)$	-0.62439055	-0.49068391	-0.1518133	-0.009662407	-3.48167E-05	-4.48726E-10	0	0	0	0

Figura 4.7 Iteraciones del método de Newton con $\gamma = 62.90$, para el caso de desgaste con censura.

Tabla 4.10 Resultados del estudio de velocidad de convergencia para el desgaste sin datos censurados.

Estudio de la velocidad de convergencia		
γ calculada con	MLE	MRR
Iteración de convergencia	6	6
α	2.13	2.27
β	1 032.36	1 090.39
γ	113.21	62.90

Como se puede observar en las figuras 4.6 y 4.7, el método converge en la sexta iteración en ambos estudios, mucho más rápido que en el caso con datos censurados. No obstante, la diferencia de los otros dos parámetros se hace notar mucho más que en la tabla 4.5.

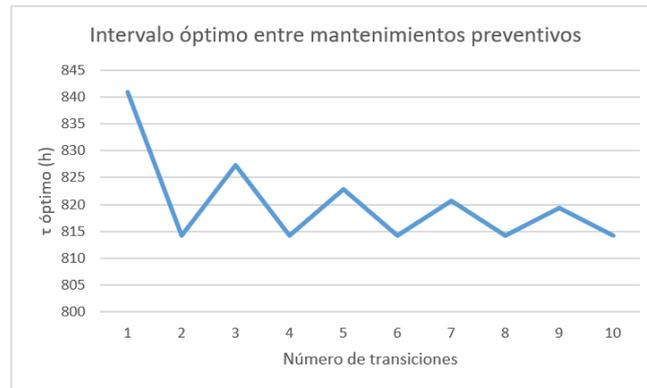
La conclusión es la misma que la de la sección anterior, se elige el valor de γ calculado mediante la estimación de máxima verosimilitud. Además, es remarcable el hecho de que el parámetro N_4 tiene un valor nulo, lo cual es congruente con la expresión 4.5, ya que solo se tienen en cuenta los datos censurados en su cálculo.

A continuación, se incluye una tabla similar a la tabla 4.6, en la que se plasman los resultados obtenidos por los tres métodos. Igualmente, la figura 4.8 ofrece una comparativa entre las gráficas y los valores alcanzados del intervalo entre mantenimientos preventivos.

Tabla 4.11 Resumen de los resultados obtenidos para el caso de desgaste sin datos censurados.

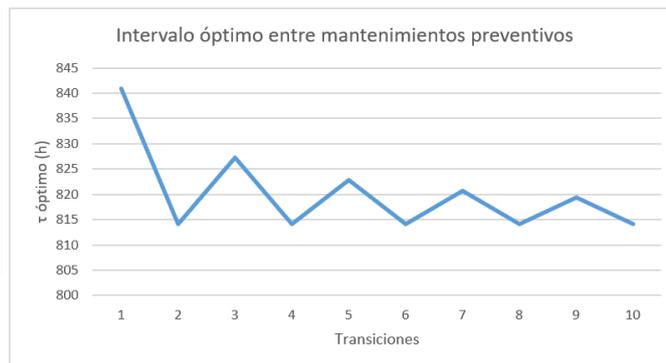
Método	α	β	γ
MLE	2.13	1 032.36	113.21
Newton-Raphson	2.13	1 032.36	113.21
MRR	2.47	848.12	62.895

m	τ óptimo (h)
1	840.909446
2	814.185864
3	827.281864
4	814.185864
5	822.859053
6	814.185864
7	820.669418
8	814.185864
9	819.36252
10	814.185864



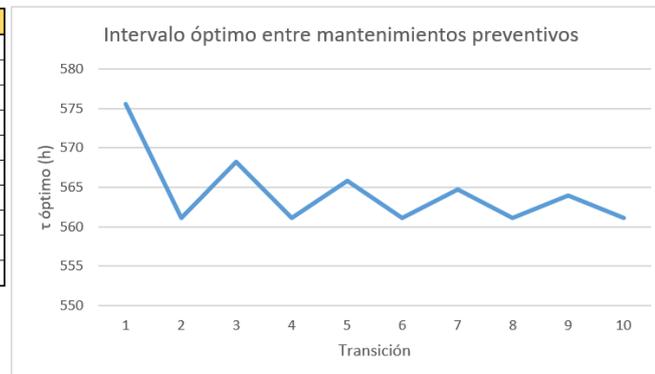
(a) Intervalo óptimo entre mantenimientos preventivos según la estimación de máxima verosimilitud.

m	τ óptimo (h)
1	840.9121063
2	814.1885011
3	827.284513
4	814.1885011
5	822.861698
6	814.1885011
7	820.672061
8	814.1885011
9	819.3651616
10	814.1885011



(b) Intervalo óptimo entre mantenimientos preventivos según el método de Newton-Raphson.

m	τ óptimo (h)
1	575.60884
2	561.1459259
3	568.2494407
4	561.1459259
5	565.8538797
6	561.1459259
7	564.6665921
8	561.1459259
9	563.9575397
10	561.1459259



(c) Intervalo óptimo entre mantenimientos preventivos según la regresión del rango de medianas.

Figura 4.8 Intervalos óptimos entre mantenimientos preventivos para el caso de desgaste con censura.

Llegados a este punto, se pueden alcanzar ciertas conclusiones comparando los valores de Weibull plamados en las tablas 4.6 y 4.11. Como se puede observar en ellas, los valores de los tres parámetros son inferiores en el caso en el que no se tiene en cuenta la censura. En particular, destaca la disparidad entre los valores de β . Esto provoca que los valores de τ_0 sean menores en este caso.

Como se puede comprobar en la figura 4.8, el valor óptimo del intervalo entre mantenimientos preventivos ha pasado de estar entre unas 1 009 h y unas 977 h a estar entre unas 840 h y unas 814 h. Si bien escoger unos intervalos de tiempo inferiores forman parte de un enfoque más conservador, ya que cuanto antes se lleven a cabo los mantenimientos menos probable es se produzca el fallo por desgaste, lo óptimo es estirar lo máximo posible el intervalo entre mantenimientos, con el fin de aumentar los beneficios percibidos. Por ello, se prefieren los resultados obtenidos al incluir los datos censurados, ya que se consideran más precisos y los intervalos son mayores.

Por otro lado, en el caso de la regresión de rango de medianas, se puede advertir que la diferencia entre los valores de los parámetros de Weibull es aún mayor, y en cada parámetro es diferente. Mientras que en el estudio con censura el valor de α es de 2.09, en este caso alcanza un valor de 2.47. Por su parte, el valor de β se ve disminuido en casi 340 unidades al desprestigiar los datos censurados. El parámetro γ también se ve disminuido. Esta enorme discrepancia entre los dos casos se debe a la valiosa información que se pierde al eludir los datos censurados. Aunque el método obtenga sus resultados de los datos de fallo, los datos censurados modifican su rango. Al no disponer de ellos, se pierde una gran cantidad de precisión. De este modo, el hecho de que se le sume mayor imprecisión a un método que de por sí es menos exacto que los otros dos, da lugar a que no se tomen en cuenta sus resultados.

Para concluir este estudio, puede decirse que los valores de τ_0 calculados mediante la estimación de máxima verosimilitud y el método de Newton-Raphson son los más adecuados, especialmente los obtenidos teniendo en cuenta los datos censurados.

Tabla 4.12 Valores finales de τ_0 .

m	1	2	3	4	5	6	7	8	9	10
τ_0 [h]	1 009.26	977.21	992.92	977.21	987.61	977.21	984.99	977.21	983.42	977.21

4.3 Caso de estudio 2: Fallos aleatorios

Por definición, el caso de fallos aleatorios, al igual que el de fallos infantiles, no da lugar al cálculo, ni mucho menos al estudio de los intervalos entre mantenimientos preventivos. No obstante, se incluyen dos estudios con conjuntos de datos situados en cada una de esas regiones de la curva de bañera, con el fin de estudiar el desempeño de los métodos de estimación de los parámetros de Weibull. En particular, esta sección incluye el estudio del caso de fallos aleatorios.

Como se puede observar en las tablas 4.13 y 4.14, los datos de fallo se encuentran repartidos a lo largo de la vida útil del equipo, registrándose fallos tanto al principio de su desempeño, como al final. Referente a los datos censurados, se observa que un equipo ha alcanzado τ y se ha enviado al mantenimiento preventivo.

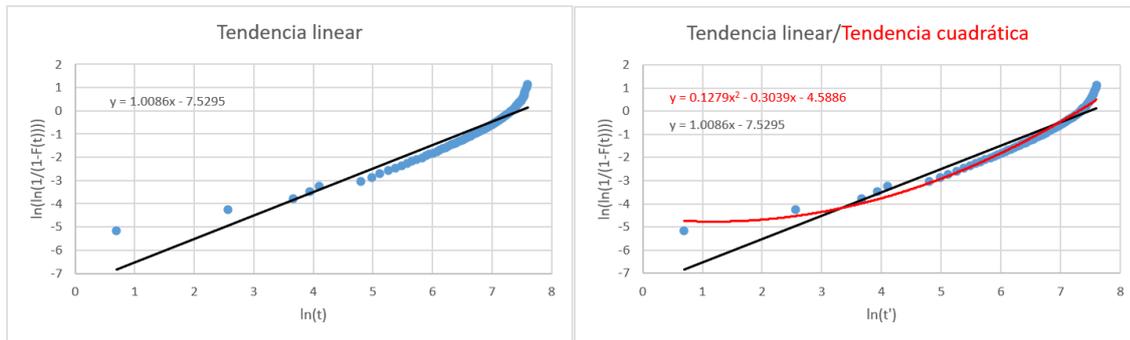
Tabla 4.13 Datos de fallo para el caso de fallos aleatorios.

Horas de funcionamiento antes del fallo										
2	13	39	51	60	121	145	167	193	217	239
265	287	309	337	361	383	409	433	455	477	505
527	553	573	601	621	647	673	697	721	743	765
793	817	843	865	887	909	937	959	985	1 009	1 033
1 058	1 081	1 105	1 129	1 153	1 177	1 201	1 225	1 249	1 273	1 297
1 319	1 345	1 369	1 389	1 417	1 441	1 465	1 489	1 513	1 537	1 561
1 585	1 609	1 633	1 657	1 681	1 703	1 727	1 754	1 756	1 771	1 781
1 831	1 881	1 916	1 956	1 992	1999					

Tabla 4.14 Datos censurados para el caso de fallos aleatorios.

Horas de funcionamiento antes del fin de la observación										
20	81	124	176	228	280	332	384	436	488	540
592	644	696	748	800	852	904	956	1 008	1 060	1 112
1 164	1 216	1 268	1 320	1 372	1 424	1 476	1 528	1 580	1 632	1 684
1 736	1 788	1 840	1 892	2 000						

Al analizar este conjunto de datos con el método la regresión de rango de medianas, se obtienen los resultados plasmados en la figura 4.9.



α	1.01
β	1746.06
γ	0

Figura 4.9 Resultados obtenidos mediante MRR para el caso de fallos aleatorios.

Como se puede observar, se ha obtenido un valor de α de 1.01, muy cercano a la unidad. En este caso, la parábola posee un término cuadrático positivo, por lo que el valor de γ que lo anula debe de ser negativo, lo que va en contra de las restricciones impuestas en este proyecto, ya que $\gamma > 0$. Por ello, no tiene sentido calcular γ para el caso de fallos aleatorios.

Respecto a la estimación de máxima verosimilitud, se ha añadido, en este caso, la restricción $\gamma = 0$ en el complemento *Solver*, debido a la conclusión alcanzada sobre la vida garantizada al analizar los resultados de la regresión del rango de medianas.

Iteración	1	2	3	4	5	6	7	8	9	10
α	1	1.19035826	1.31523531	1.39316078	1.4406952	1.4693657	1.48655499	1.49682651	1.50295265	1.50660237
N_1	81651	316411.092	772649.784	1350759.67	1900148.19	2334854.87	2641987.47	2844538.67	2972686.52	3051768.68
N_2	579216.87	2258184.55	5533166.38	9692036.86	13649402	16783089.2	18998134.7	20459331.7	21383933.8	21954575.6
N_3	4129091.06	16181726.5	39767616.9	69777052.2	98365310.1	121018730	137037788	147607614	154296845	158425626
N_4	264928.757	1031821.6	2526733.01	4424341.78	6229565.86	7658843.68	8669038.98	9335398.43	9757037.69	10017258.2
N_5	119032	461101.297	1125734.68	1967788.92	2767941.75	3401035.36	3848320.59	4143295.92	4329915.66	4445081.08
$h(\alpha)$	0.44034153	0.23707349	0.13273608	0.07616142	0.04434537	0.02604568	0.01537653	0.00910555	0.00540182	0.00320806
$h'(\alpha)$	-2.3132252	-1.89845524	-1.70337214	-1.60223726	-1.54672499	-1.51522712	-1.49700584	-1.4863445	-1.48006454	-1.47635069
$h(\alpha)/h'(\alpha)$	-0.19035826	-0.12487705	-0.07792547	-0.04753442	-0.0286705	-0.01718929	-0.01027152	-0.00612614	-0.00364972	-0.00217296

Iteración	11	12	13	14	15	16	17	18	19	20	21
α	1.50877533	1.51006857	1.51083808	1.51129588	1.51156822	1.51173023	1.5118266	1.51188392	1.51191802	1.5119383	1.51195037
N_1	3099851.16	3128827.45	3146197.61	3156577.62	3162768.87	3166457.6	3168653.86	3169961.01	3170738.79	3171201.53	3171476.81
N_2	22301549.2	22510655.5	22636009	22710918.3	22755599	22782219.7	22798069.7	22807503.1	22813116.2	22816455.7	22818442.3
N_3	160936222	162449297	163356361	163898414	164221731	164414364	164529057	164597320	164637937	164662103	164676478
N_4	10175481.2	10270834.9	10327996.6	10362155.5	10382530	10394669	10401896.7	10406198.3	10408757.9	10410280.7	10411186.6
N_5	4515102.15	4557299.36	4582594.89	4597710.93	4606727.03	4612098.77	4615297.12	4617200.66	4618333.32	4619007.18	4619408.06
$h(\alpha)$	0.00190643	0.00113335	0.00067392	0.00040078	0.00023837	0.00014178	8.4327E-05	5.0158E-05	2.9835E-05	1.7746E-05	1.0556E-05
$h'(\alpha)$	-1.4741492	-1.47284239	-1.47206602	-1.47160455	-1.47133018	-1.47116701	-1.47106998	-1.47101226	-1.47097794	-1.47095752	-1.47094537
$h(\alpha)/h'(\alpha)$	-0.00129324	-0.0007695	-0.0004578	-0.00027234	-0.00016201	-9.6369E-05	-5.7324E-05	-3.4098E-05	-2.0282E-05	-1.2064E-05	-7.1762E-06

Figura 4.10 Iteraciones del método de Newton para el caso de fallos aleatorios.

Por su parte, no puede llevarse a cabo el estudio de convergencia del algoritmo de Newton-Raphson al suponerse nulo el valor de la vida garantizada. De cualquier modo, se incluye la figura 4.10 para ilustrar el proceso iterativo que ha llevado a cabo el algoritmo. Como se puede observar en ella, el método converge en la iteración número 21.

Por otro lado, de los estudios de la sección 4.2 se sacó como conclusión que existe cierta diferencia entre los resultados de cada método debido a que cada uno trata los datos de forma diferente y que algunos son más precisos que otros. En la tabla 4.15 se agrupan los diferentes resultados obtenidos por cada método en

este estudio.

Tabla 4.15 Resumen de los resultados obtenidos para el caso de fallos aleatorios.

Método	α	β	γ
MLE	1.51	1 376.15	0
Newton-Raphson	1.51	1 376.15	0
MRR	1.01	1 746.06	0

Al igual que en el análisis de fallo por desgaste, se registran valores más altos de α y más bajos de β en los métodos que emplean la función de verosimilitud. No obstante, en esta situación la diferencia es más acusada, de forma que el valor de α según la estimación de máxima verosimilitud y el método de Newton superan el valor unidad, entrando en la fase de desgaste.

Esta mayor diferencia de cálculo puede deberse, a parte de a las diferencias relacionadas con el método de cálculo de cada método, al caso que se está estudiando. El caso de fallos aleatorios no suele darse en la vida real, donde predomina el desgaste. Por ello, al ser un caso especial, los métodos se muestran más sensibles, dando lugar a mayores diferencias. Por último, queda comentar que, al ser la regresión del rango de medianas un método más manejable, es más sencillo lograr ajustar una muestra de datos aleatoria con ella.

La figura 4.11 incluye los valores de τ_0 que se obtendrían con la expresión 3.17 al introducir los valores obtenidos con la regresión del rango de medianas, ya que es el método con el que se ha logrado alcanzar un $\alpha \approx 1$.

m	τ óptimo (h)
1	2.71475E+47
2	2.0433E+45
3	2.29531E+46
4	2.0433E+45
5	1.01913E+46
6	2.0433E+45
7	6.8053E+45
8	2.0433E+45
9	5.34454E+45
10	2.0433E+45

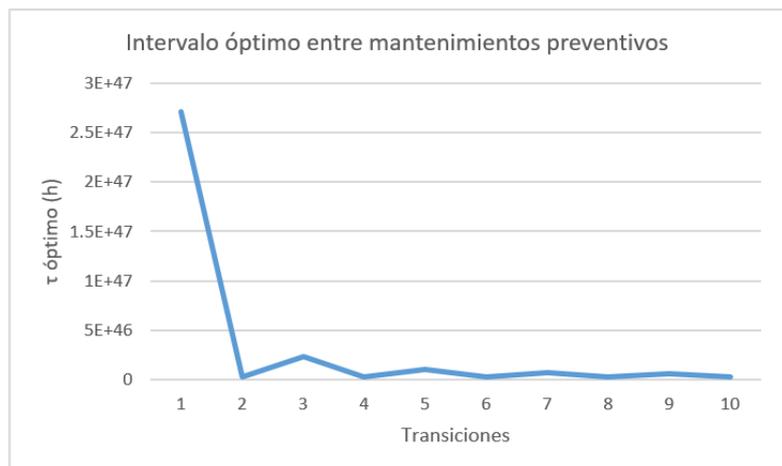


Figura 4.11 Intervalo óptimo entre mantenimientos preventivos según MRR para el caso de fallos aleatorios.

Como se puede observar, la expresión 3.17 devuelve unos resultados muy grandes, haciendo imposible hablar de intervalos entre mantenimientos preventivos, lo cual es coherente con la situación en la que se encuentra la muestra analizada.

4.4 Caso de estudio 3: Fallos infantiles

Finalmente, se presenta un conjunto de datos donde los fallos se dan a las pocas horas de comenzar a operar. Esto se puede ver reflejado en la tabla 4.16, donde más de la mitad de los fallos se producen antes de las 100 horas de funcionamiento.

Del mismo modo, la tabla 4.17 agrupa al conjunto de datos censurados de este estudio. En este caso, un equipo logró alcanzar el mantenimiento preventivo.

Tabla 4.16 Datos de fallo para el caso de fallos infantiles.

Horas de funcionamiento antes del fallo										
30	30	32	34	36	42	42	43	45	48	49
49	51	51	51	52	56	57	60	61	61	63
63	64	65	68	69	70	71	72	72	74	74
75	78	82	82	85	88	90	94	97	99	100
102	104	104	107	108	109	110	115	117	119	121
125	127	127	132	132	139	140	140	142	143	146
146	149	151	152	279	406	490	502	621	752	891
1 025	1 321	1 430	1 490	1 590	1 702					

Tabla 4.17 Datos censurados para el caso de fallos infantiles.

Horas de funcionamiento antes del fin de la observación										
80	90	93	109	131	134	135	140	143	150	156
170	178	188	198	201	208	212	218	500	512	649
842	900	1 021	1 054	1 056	1 125	1 143	1 256	1 271	1 321	1 562
1 632	1 689	1 703	1 900	2 000						

Según lo indicado en la sección 3.3, los conjuntos de datos con predominio de fallos prematuros deben de registrar un valor de α inferior a la unidad. Para comprobarlo, se hace uso de los mismos tres métodos que se han empleado a lo largo de este proyecto.

Por su parte, los resultados obtenidos por medio de la regresión del rango de medianas se ven representados en la figura 4.12.

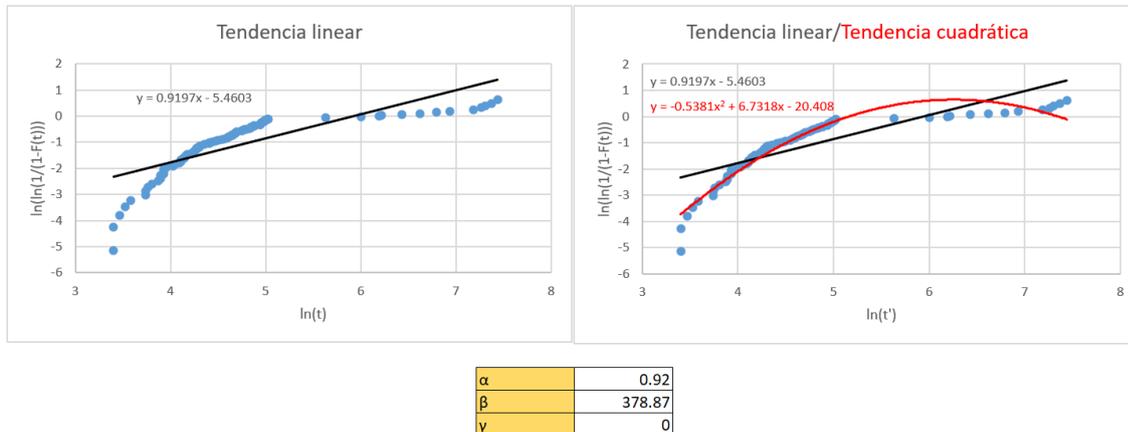


Figura 4.12 Resultados obtenidos mediante MRR para el caso de fallos infantiles.

Como se puede observar, se obtiene un valor de α de 0.92 y un valor de β de 378.87. En este caso, el valor de γ que anula el término cuadrático debe de ser muy pequeño para que no se produzcan errores en los cálculos. Como no se puede encontrar dicho valor, se asume nulo, siendo necesario volver a imponer, en este caso de estudio, la restricción $\gamma = 0$ en el complemento Solver de Excel. De nuevo, es imposible realizar el estudio de velocidad de convergencia de Newton-Raphson, convergiendo el algoritmo en la décimo tercera iteración.

Iteración	1	2	3	4	5	6	7	8	9	10	11	12	13
α	1	0.8009617	0.72844412	0.69980627	0.68813301	0.68330875	0.68130314	0.68046725	0.6801185	0.67997293	0.67991216	0.67988679	0.6798762
N_1	18581	5585.21419	3663.20799	3108.70551	2908.67855	2830.00883	2797.96494	2784.72268	2779.21729	2776.92272	2775.96538	2775.56578	2775.39895
N_2	115131.187	32824.2781	21083.2718	17742.1306	16543.4571	16073.0797	15881.6625	15802.5891	15769.7201	15756.0217	15750.3066	15747.9211	15746.9253
N_3	741529.96	202219.767	127504.779	106485.802	98979.9268	96040.1308	94844.7466	94351.1031	94145.935	94060.4345	94024.7636	94009.8748	94003.6591
N_4	180779.357	45188.6873	27414.1194	22523.9264	20793.517	20118.3318	19844.2184	19731.0952	19684.0918	19664.5061	19656.3353	19652.925	19651.5012
N_5	44651	12220.7078	7720.34734	6453.10811	6000.38983	5823.04437	5750.92658	5721.14405	5708.76567	5703.60715	5701.45501	5700.55672	5700.18171
$h(\alpha)$	-0.88801898	-0.39600239	-0.16980554	-0.07165738	-0.03004889	-0.01256878	-0.00525177	-0.00219347	-0.00091596	-0.00038247	-0.0001597	-6.6679E-05	-2.7841E-05
$h'(\alpha)$	-4.46154814	-5.46077836	-5.92940832	-6.13859557	-6.22869606	-6.26682705	-6.28283932	-6.28954106	-6.29234208	-6.29351209	-6.29400069	-6.29420472	-6.29428991
$h(\alpha)/h'(\alpha)$	0.1990383	0.07251757	0.02863786	0.01167325	0.00482427	0.0020056	0.00083589	0.00034875	0.00014557	6.0772E-05	2.5373E-05	1.0594E-05	4.4232E-06

Figura 4.13 Iteraciones del método de Newton para el caso de fallos infantiles.

Al igual que en el resto de casos de estudio, la diferencia de precisión entre los métodos, sumada al distinto trato que cada uno le da a los datos, produce diferencias entre los resultados obtenidos con los diferentes métodos. La tabla 4.18 agrupa todos los resultados obtenidos por los tres métodos, para este conjunto de datos.

Tabla 4.18 Resumen de los resultados obtenidos para el caso de fallos infantiles.

Método	α	β	γ
MLE	0.68	503.13	0
Newton-Raphson	0.68	503.13	0
MRR	0.92	378.87	0

En este caso, se invierte el comportamiento mostrado en la tabla 4.15, de forma que el valor de α obtenido por los métodos que hacen uso de la función de verosimilitud es menor que el obtenido mediante la regresión de rango de medianas. Al contrario que en el estudio de fallos aleatorios, esta vez todos los valores de α obtenidos se encuentran bajo la misma región de la curva de bañera. Esto se debe a que, en este caso, se habla de un intervalo de valores de α , en lugar de un valor puntual. Por su parte, β presenta un valor superior en la estimación de máxima verosimilitud y en el método de Newton-Raphson.

Al igual que el caso de fallos aleatorios, este es un caso especial, que no suele darse en la vida real. Si bien los defectos de fabricación son una realidad presente en la industria, pocas veces se da la situación en la que se acumulan de tal forma en una muestra de datos. Es por esto que se detecta una especial sensibilidad en la precisión de los métodos. De cualquier modo, sigue siendo más recomendable emplear la estimación de máxima verosimilitud o el método de Newton-Raphson, según la situación, ya que los datos fuera de la tendencia del conjunto tienen demasiada influencia sobre la línea de tendencia de la regresión del rango de medianas, perdiéndose precisión.

Como se había comentado en la sección 3.3, el caso de fallo infantiles tampoco contempla el estudio de los mantenimientos preventivos. De cualquier modo, se incluye la figura 4.14 para comprobar el desempeño de la expresión 3.17 en esta situación.

Los resultados obtenidos en esta situación revelan que la expresión 3.17 no presenta un buen comportamiento, dando resultados demasiado altos. De cualquier modo, es coherente con la teoría, ya que no se puede hablar de mantenimientos preventivos en fallos tan prematuros. Por último, cabe destacar que el comportamiento de la gráfica se invierte con respecto a los otros estudios. En este caso, es máximo el valor registrado para las transiciones pares y el valor de τ para transiciones impares va aumentando, amortiguándose con el tiempo.

m	τ óptimo (h)
1	394869.9122
2	667505.3525
3	514820.6831
4	667505.3525
5	561717.0648
6	667505.3525
7	586610.0295
8	667505.3525
9	602028.7118
10	667505.3525

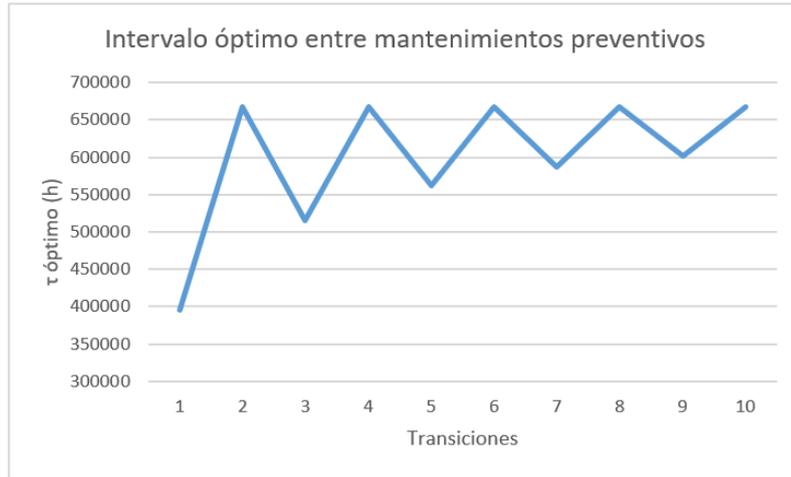


Figura 4.14 Intervalo óptimo entre mantenimientos preventivos según MRR para el caso de fallos infantiles.

Llegados a este punto, se ha analizado tanto el comportamiento de la expresión 3.17, como el de los métodos elegidos, en diferentes situaciones. Del mismo modo, se ha comparado el desempeño de dichos métodos entre sí, dando lugar a varias conclusiones. Todas ellas se encuentran recogidas en el capítulo 5.

5 Conclusiones

Los estudios llevados a cabo a lo largo de este proyecto han dado lugar a ciertos resultados con los que se pueden alcanzar varias conclusiones. Si bien dichas conclusiones han sido expuestas mientras se desarrollaba la aplicación matemática del capítulo anterior, se emplea este capítulo final para agrupar todas ellas y poder mostrárselas de forma clara al lector.

- Los valores de α y β disminuyen al aumentar γ . Esto produce una disminución en los valores del intervalo óptimo obtenidos.
- El método de Newton-Raphson converge a la misma velocidad haciendo uso de los dos valores de γ obtenidos mediante los otros métodos. Se concluye que es mejor emplear el γ calculado mediante la estimación de máxima verosimilitud, ya que ambos emplean la misma función para realizar los cálculos.
- Los resultados obtenidos con la estimación de máxima verosimilitud y el método de Newton-Raphson son prácticamente idénticos. Esto tiene sentido ya que el segundo método es una variación del primero.
- Se demuestra el comportamiento teórico de la evolución del intervalo óptimo para 10 transiciones (figura 3.5).
- Se consideran más adecuados los resultados obtenidos mediante la estimación de máxima verosimilitud y, por supuesto, el método de Newton-Raphson, ya que se consideran métodos más precisos por los autores revisados. Además, desde un punto de vista más conservador, los resultados obtenidos por estos métodos son preferibles, ya que presentan unos valores de τ_0 ligeramente inferiores, por lo que se darían menos fallos por desgaste.
- Al no tener en cuenta los datos censurados, se obtienen valores de τ_0 aún más pequeños. Se concluye que son más precisos los resultados obtenidos mediante la inclusión de los datos censurados en el estudio. Además, aunque cuanto menores sean los valores de τ_0 , menos probabilidad habrá de que se produzca un fallo por desgaste, económicamente interesa aumentar el intervalo entre mantenimientos lo máximo posible.
- En particular, la regresión del rango de medianas pierde una mayor precisión que los otros dos métodos al eludir la información de los datos censurados.
- Las discrepancias en los resultados entre métodos se deben a que tratan los datos de diferente forma.
- En el caso de fallos aleatorios, se obtiene un término cuadrático positivo, por lo que se requeriría de un valor negativo de γ óptimo. Esto choca con la restricción de que γ ha de ser mayor o igual a 0, por lo que se concluye que no tiene sentido calcular γ en este caso.
- El hecho de estudiar un caso tan particular como es el de fallos aleatorios aumenta la sensibilidad de los métodos, dando lugar a mayores diferencias entre los resultados calculados con cada uno. De este modo, mientras que la regresión del rango de medianas sitúa al conjunto en el rango de fallos aleatorios con una α de valor 1.01, los otros dos métodos registran un valor de α de 1.51 más propio de la región de desgaste.
- Se concluye que la manejabilidad de la regresión del rango de medianas hace más fácil poder ajustar una muestra de datos aleatorios.

- En el caso de fallos infantiles tampoco tiene sentido calcular la vida garantizada, ya que tendría un valor muy bajo y desde el punto de vista del mantenimiento, tiene poco sentido. De cualquier modo, el valor de γ que anula el término cuadrático en la regresión del rango de medianas, muy probablemente sería mayor al primer fallo registrado, lo cual no es posible, ya que es incongruente con la definición de la vida garantizada.
- En el caso de fallos infantiles, también se registra una discrepancia notable entre los resultados de los diferentes métodos, aunque todos los valores de α son inferiores a 1.
- Los enormes resultados de τ_0 obtenidos de los estudios de fallos aleatorios e infantiles, confirman que no tiene sentido hablar de mantenimientos preventivos en estos casos.
- En el caso de fallos infantiles, el comportamiento de la gráfica de intervalos óptimos según las transiciones se invierte.

Bibliografía

- [1] Antonio Sánchez Herguedas, Adolfo Crespo Márquez y Francisco Rodrigo Muñoz, *Optimizing preventive maintenance over a finite planning horizon in a semi-Markov framework*, IMA Journal of Management Mathematics **33** (2020), no. 1, 75–99.
- [2] Antonio Sánchez-Herguedas, Angel Mena-Nieto, Francisco Rodrigo-Muñoz, Javier Villalba-Diez y Joaquín Ordieres-Meré, *Aplicación de enfoques semimarkovianos a la mejora de políticas de mantenimiento predeterminado en activos industriales*, 11 2021, pp. 1–4.
- [3] ———, *Optimisation of maintenance policies based on right-censored failure data using a semi-Markovian approach*, Sensors **22** (2022), no. 4.
- [4] Antonio Sánchez-Herguedas, Angel Mena-Nieto y Francisco Rodrigo-Muñoz, *A new analytical method to optimise the preventive maintenance interval by using a semi-Markov process and Z-transform with an application to marine diesel engines*, Reliability Engineering System Safety **207** (2021), 107394.
- [5] Anwei Shen, Jilian Guo, Zhuojian Wang y Wentong Jia, *A novel reliability evaluation method on censored data*, Journal of Mechanical Science and Technology **31** (2017), 1105–1117.
- [6] Avijit Joarder, Hare Krishna y Debasis Kundu, *Inferences on Weibull parameters with conventional type-I censoring*, Comput. Stat. Data Anal. **55** (2011), 1–11.
- [7] N. Balakrishnan, *Approximate MLE of the scale parameter of the Rayleigh distribution with censoring*, IEEE Transactions on Reliability **38** (1989), no. 3, 355–357.
- [8] Charles Zaiantz, *Weibull with censored data*. Disponible en: <https://www.real-statistics.com/distribution-fitting/distribution-fitting-via-maximum-likelihood/weibull-censored-data/>, última consulta: 12 de septiembre del 2022.
- [9] Satya D. Dubey, *Some percentile estimators for Weibull parameters*, Technometrics **9** (1967), no. 1, 119–129.
- [10] Edwin J. Green, Francis A. Roesch, Adrian F. M. Smith y William E. Strawderman, *Bayesian estimation for the three-parameter Weibull distribution with tree diameter data*, Biometrics **50** (1994), no. 1, 254–269.
- [11] Bradley Efron, *Bootstrap methods: Another look at the Jackknife*, The Annals of Statistics **7** (1979), no. 1, 1 – 26.
- [12] Elisa T. Lee y John W. Wang, *Statistical methods for survival data analysis*, tercera ed., John Wiley & Sons, Inc, 2003.
- [13] Geoffrey J. MacLachlan y Thriyambakan Krishnan, *The EM algorithm and extensions*, segunda ed., John Wiley & Sons, Inc, 2008.
- [14] Charles J. Geyer, *Practical Markov Chain Monte Carlo*, Statistical Science **7** (1992), no. 4, 473 – 483.

- [15] Didier Chauveau y Jean Diebolt Gilles Celeux, *Stochastic versions of the EM algorithm: an experimental study in the mixture case*, Journal of Statistical Computation and Simulation **55** (1996), no. 4, 287–314.
- [16] A. K. Gupta, *Estimation of the mean and standard deviation of a normal population from a censored sample*, Biometrika **39** (1952), no. 3/4, 260–273.
- [17] Hideo Hirose, *Bias correction for the maximum likelihood estimates in the two-parameter Weibull distribution*, IEEE Transactions on Dielectrics and Electrical Insulation **6** (1999), 66–68.
- [18] J. R. M. Hosking, *L-moments: Analysis and estimation of distributions using linear combinations of order statistics*, Journal of the Royal Statistical Society. Series B (Methodological) **52** (1990), no. 1, 105–124.
- [19] James K. Starling, Chrsitina Mastrangelo y Youngjun Choe, *Improving Weibull distribution estimation for generalized type I censored data using modified SMOTE*, Reability Engineering and System Safety **221** (2021), 107505.
- [20] John P. Klein y Melvin L. Moeschberger, *Survival analysis: Techniques for censored and truncated data*, segunda ed., Springer-Verlag New York, Inc, 2003.
- [21] Kailash C. Kapur y Leonard R. Lamberson, *Reliability in engineering design*, primera ed., John Wiley & Sons, Inc, 1991.
- [22] M. Xie y L. C. Tang L. F. Zhang, *Bias correction for the least squares estimator of Weibull shape parameter with complete and censored data*, Reliability Engineering System Safety - RELIAB ENG SYST SAFETY **91** (2006), 930–939.
- [23] Kenneth L. Lange, *A gradient algorithm locally equivalent to the EM algorithm*, Journal of the Royal Statistical Society: Series B (Methodological) **57** (1995), 425–437.
- [24] Jerald F. Lawless, *Statistical models and methods for lifetime data*, segunda ed., John Wiley & Sons, Inc, 2003.
- [25] Maha Alqam, Richard M. Bennett y Abdul-Hamid Zureick, *Three-parameter vs. two-parameter Weibull distribution for pultruded composite material properties*, Composite Structures **58** (2002), no. 4, 497–503.
- [26] Mahdi Teimouri, Seyed M. Hoseini y Saralees Nadarajah, *Comparison of estimation methods for the Weibull distribution*, Statistics **47** (2013), no. 1, 93–109.
- [27] Barry Kurt Moser, *Linear models: A mean model approach (a volume in probability and mathematical statistics)*, primera ed., Academic Press, 1996.
- [28] David M. Nanang, *Suitability of the normal, log-normal and Weibull distributions for fitting diameter distributions of neem plantations in northern ghana*, Forest Ecology and Management **103** (1998), 1–7.
- [29] Wayne Nelson, *Applied life data analysis*, primera ed., John Wiley & Sons, Inc, 1982.
- [30] Søren Feodor Nielsen, *The stochastic EM algorithm: Estimation and asymptotic results*, Bernoulli **6** (2000), no. 3, 457–489.
- [31] Milton Ohring, *Engineering materials science*, primera ed., Academic Press, noviembre 1995.
- [32] Philip Kosky, Robert Balmer, William Keat y George Wise, *Exploring engineering. an introduction to engineering and design*, quinta ed., Academic Press, abril 2020.
- [33] R. B. Abernethy, J. E. Breneman, C. H. Medlin y G. L. Reinman, *Weibull analysis handbook*, Tech. report, Defense Technical Information Center, noviembre 1933.
- [34] Rakesh Ranjan, Rijji Sen y Satyanshu K. Upadhyay, *Bayes analysis of some important lifetime models using MCMC based approaches when the observations are left truncated and right censored*, Reliability Engineering and System Safety **214** (2021), 107747.

- [35] Ricardo Medel Esquivel, Isidro Gomez-Vargas, J. Alberto Vazquez y Ricardo García-Salcedo, *An introduction to Markov chain Monte Carlo*, Boletín de Estadística e Investigación Operativa **37** (2021), 47–84.
- [36] Riffat Jabeen, Azaz Ahmad, Navid Feroze y Ghausia M. Gilani, *Estimation of location and scale parameters of Weibull distribution using generalized order statistics under type II singly and doubly censored data*, International Journal of Advanced Science and Technology **55** (2013), 67–80.
- [37] Rola M. Musleh y Amal Helu, *Estimation of the inverse Weibull distribution based on progressively censored data: Comparative study*, Reliability Engineering & System Safety **131** (2014), 216–227.
- [38] Robert Ross, *Formulas to describe the bias and standard deviation of the ML-estimated Weibull shape parameter*, IEEE Transactions on Dielectrics and Electrical Insulation **1** (1994), 247–253.
- [39] _____, *Bias and standard deviation due to Weibull parameter estimation for small data sets*, IEEE Transactions on Dielectrics and Electrical Insulation **3** (1996), 28–42.
- [40] Russell B. Millar, *Maximum likelihood estimation and inference: with examples in R, SAS and ADMB*, primera ed., John Wiley & Sons, Inc, 2011.
- [41] S. B. Kang, Y. S. Cho y S. H. Choi, *Approximate maximum likelihood estimation for the three-parameter Weibull Distribution*, The Korean Communications in Statistics **8** (2001), no. 1, 209–217.
- [42] Stephen Shifley and Ellen Lentz, *Quick estimation of the three-parameter Weibull to describe tree size distributions*, Forest Ecology and Management **13** (1985), 195–203.
- [43] David Smith, *Reliability, maintainability and risk. practical methods for engineers including reliability centred maintenance and safety-related systems*, séptima ed., Butterworth-Heinemann, abril 2005.
- [44] Steven Brown, Beata Walczak y Romà Tauler, *Comprehensive chemometrics: Chemical and biochemical data analysis*, primera ed., Elsevier Science, Ltd, 2009.
- [45] Jeannet Gunneklev Støtvig, *Censored Weibull distributed data in experimental design*, Ph.D. thesis, Norwegian University of Science and Technology, 2014.
- [46] Switamy Angnitha Purba, Sutarman y Open D. Sembiring, *Maximum likelihood based on Newton-Raphson, Fisher scoring and expectation maximization algorithm application on accident data.*, International Journal of Advanced Research **6** (2018), 965–969.
- [47] William Q. Meeker y Luis A. Escobar, *Statistical methods for reliability data*, primera ed., John Wiley & Sons, Inc, 1998.
- [48] N. Balakrishnan y A. Clifford Cohen, *Order statistics and inference: Estimation methods*, primera ed., Academic Press, INC., 1991.
- [49] Patrick D. T. O'Connor y André Kleyner, *Practical reliability engineering*, quinta ed., John Wiley & Sons, Inc, 2012.
- [50] Kandethody M. Ramachandran y Chris P. Tsokos, *Mathematical statistics with applications in R*, tercera ed., Elsevier Inc, 2020.
- [51] N. Balakrishnan y Debanjan Mitra, *Left truncated and right censored Weibull data and likelihood inference with an illustration*, Computational Statistics and Data Analysis **56** (2012), no. 12, 4011–4025.
- [52] Mostafa Bacha y Gilles Celeux, *Bayesian estimation of a Weibull distribution in a highly censored and small sample setting*, Research Report RR-2993, INRIA, 1996, Projet ISIS.
- [53] Jan Merganič y Hubert Sterba, *Characterisation of diameter distribution using the Weibull function: Method of moments*, European Journal of Forest Research **125** (2006), 427–439.
- [54] G. Celeux y J. Diebolt, *The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem*, Computational Statistics Quarterly **2** (1986), 73–82.

- [55] N. Balakrishnan y J. Varadan, *Approximate MLEs for the location and scale parameters of the extreme value distribution with censoring*, IEEE Transactions on Reliability **40** (1991), no. 2, 146–151.
- [56] Luis Ferreira y José Silva, *Parameter estimation for Weibull distribution with right censored data using EM algorithm*, Eksploatacja i Niezawodność - Maintenance and Reliability **19** (2017), 310–315.
- [57] Florence Ducros y Patrick Pamphile, *Bayesian estimation of Weibull mixture in heavily censored data setting*, Reliability Engineering and System Safety **180** (2018), no. C, 453–462.
- [58] Takeshi Emura y Shau-Kai Shiu, *Estimation and model selection for left-truncated and right-censored lifetime data with application to electric power transformers analysis*, Communications in Statistics - Simulation and Computation **45** (2016), no. 9, 3171–3189.
- [59] Stanley J. Zarnoch y Tommy R. Dell, *An evaluation of percentile and maximum likelihood estimators of weibull parameters*, Forest Science **31** (1985), 260–268.
- [60] Ulrike Genschel y William Meeker, *A comparison of maximum likelihood and median-rank regression for Weibull estimation*, Quality Engineering **22** (2010), 236–255.
- [61] Anwar Hossain y William Zimmer, *Comparison of estimation methods for Weibull parameters: Complete and censored samples*, Journal of Statistical Computation and Simulation **73** (2003), no. 2, 145–153.
- [62] Stelios H. Zanakos, *A simulation study of some simple estimators for the three-parameter Weibull distribution*, Journal of Statistical Computation and Simulation **9** (1979), 101–116.