

TEMat

Distribuciones tipo fase en un estudio de fiabilidad

✉ Christian José Acal González

Departamento de Estadística e Investigación Operativa, Universidad de Granada, e Instituto de Matemáticas de la Universidad de Granada (IEMath-GR)
chracal@correo.ugr.es

Juan Eloy Ruiz Castro

Departamento de Estadística e Investigación Operativa, Universidad de Granada, e Instituto de Matemáticas de la Universidad de Granada (IEMath-GR)
jeloy@ugr.es

Ana María Aguilera del Pino

Departamento de Estadística e Investigación Operativa, Universidad de Granada, e Instituto de Matemáticas de la Universidad de Granada (IEMath-GR)
aaguiler@ugr.es

Resumen: Hoy en día, el análisis de fiabilidad está presente en cualquier área de conocimiento donde se esté interesado en comprobar el tiempo de vida (o, análogamente, el tiempo de fallo) de un sistema dado. A modo de ejemplo, esta disciplina es ampliamente utilizada en estudios de ingeniería donde el objetivo fundamental es garantizar la calidad y el buen funcionamiento de los aparatos. La rama encargada de estudiar y analizar los tiempos de fallos de estos sistemas es la estadística. La estadística pretende ajustar los datos experimentales medidos en estos sistemas mediante el uso de distribuciones de probabilidad. Un enfoque relativamente nuevo, y que cada vez está teniendo más importancia en la actualidad, es a través de las distribuciones tipo fase, cuya flexibilidad y propiedades hacen que esta clase de distribuciones sea una buena candidata para obtener un ajuste más riguroso.

Abstract: Nowadays, the reliability analysis is present in any knowledge area where one is interested in checking the lifetime (or, similarly, the failure time) of a given system. As an example, this discipline is widely used in engineering studies where the main objective is to ensure the quality and the appropriate operation of the devices. The branch of knowledge responsible for studying and analyzing the failure times of these systems is statistics. The purpose of statistics is to fit the experimental data measured in these systems through the use of probability distributions. A relatively new approach, which is having more significance in currently, is through phase type distributions, whose flexibility and properties make them good candidates to obtain a better fit.

Palabras clave: fiabilidad, tiempos de fallo, distribuciones de probabilidad, análisis gráfico, distribución tipo fase, ajuste estadístico.

MSC2010: 62J05, 62N05, 62P99.

Recibido: 28 de noviembre de 2018.

Aceptado: 26 de abril de 2019.

Agradecimientos: Los autores agradecen el apoyo del Ministerio de Economía y Competitividad de España en el proyecto MTM2017-87708-P, también respaldado por el programa FEDER.

Referencia: ACAL GONZÁLEZ, Christian José; RUIZ CASTRO, Juan Eloy y AGUILERA DEL PINO, Ana María. «Distribuciones tipo fase en un estudio de fiabilidad». En: *TEMat*, 3 (2019), págs. 63-74. ISSN: 2530-9633. URL: <https://temat.es/articulo/2019-p63>.

© Este trabajo se distribuye bajo una licencia Creative Commons Reconocimiento 4.0 Internacional
<https://creativecommons.org/licenses/by/4.0/>

1. Introducción

El análisis de fiabilidad es una materia multidisciplinar que ha sido desarrollada en diferentes contextos dentro de la ingeniería (eléctrica, mecánica, química, civil) y que tiene múltiples aplicaciones, entre otras áreas, en supervivencia y medicina. Desde el punto de vista general, la fiabilidad comprende el conjunto de operaciones utilizadas para el buen funcionamiento y seguridad de los sistemas (un sistema puede ser una red eléctrica, cualquier electrodoméstico, un coche, un ser vivo, etc.). Estos sistemas funcionan un determinado tiempo influenciados por ciertas condiciones ambientales específicas que no se pueden controlar y sometidos a un continuo desgaste. El conjunto de todas estas condiciones o variables afecta al funcionamiento de los sistemas, provocando que dichos sistemas fallen de manera aleatoria. A raíz de esto, y haciendo hincapié en que la ocurrencia de fallos se produce aleatoriamente, la teoría de la probabilidad juega un papel determinante en el cálculo de la fiabilidad de componentes y sistemas. Está claro que todas las unidades de un cierto tipo, fabricadas y operando en condiciones similares, no fallarán a la vez, sino que habrá diferencias entre los tiempos de fallo. Consecuentemente, estos tiempos de fallo obedecen a distribuciones de probabilidad que pueden ser o no conocidas, y que permiten calcular la probabilidad de fallo de las unidades. Se debe destacar que en un estudio de fiabilidad existen múltiples ocasiones en las que el objetivo no es estudiar los tiempos de fallo en sí, sino otra variable que, aunque esté altamente relacionada con los tiempos de fallo, trate un problema distinto. Por ejemplo, en los artículos de Long *et al.* [8] y Luo *et al.* [9] se realiza un estudio de los datos experimentales correspondientes al voltaje de fallo de un tipo de memorias concretas (estos aparatos dejan de funcionar tras aplicarles un cierto voltaje, pero paralelamente han estado funcionando un determinado tiempo).

En consecuencia, la estadística y, más concretamente, la probabilidad proporcionan una serie de herramientas para el cálculo y la mejora de la fiabilidad y suministran una definición cuantitativa de la misma. Aunque existen distintas definiciones para aclarar este concepto, la definición usual es la que sigue: «Probabilidad de que un dispositivo efectuará la función para la que fue construido hasta un momento dado bajo condiciones específicas de uso» [4]. Por lo tanto, podemos concluir que la fiabilidad es la probabilidad de que un sistema se comporte adecuadamente durante un tiempo establecido y, en cierto sentido, podemos verla como una medida de calidad: cuanto más tiempo funcione, mejor. Sin embargo, dejando a un lado los conceptos típicos de ingeniería para centrarnos en la teoría de la probabilidad, que es lo que realmente nos compete, saber qué distribución probabilística siguen los tiempos de fallo nos permite, en primer lugar, determinar las principales características de operación de dicho sistema y, en segundo lugar, conocer cómo funcionará el sistema en el futuro.

En muchos campos donde se hace uso de la estadística aplicada, la distribución normal es el punto de partida natural para modelizar cualquier variable aleatoria de interés. La razón fundamental de su uso se debe a las buenas propiedades que presenta esta distribución (es simétrica, unimodal, asintótica, etc.) y por proporcionar la base para la estadística inferencial clásica por su relación con el teorema central del límite. Sin embargo, en el ámbito de la fiabilidad, donde es habitual trabajar con variables que toman valores positivos, la distribución normal tiene menor interés. En la década de los 50, Epstein y Sobel empezaron a trabajar con la distribución exponencial como modelo probabilístico para analizar el tiempo de vida de unos dispositivos [6]. Una razón fundamental de la popularidad e importancia de la distribución exponencial para su uso en el ámbito de la fiabilidad (probablemente la más usada) es su simplicidad y versatilidad. Por ejemplo, esta distribución es muy útil cuando los datos que se utilizan son escasos, cuando se estudian sistemas en régimen estacionario (el tiempo de funcionamiento crece indefinidamente) o cuando se tienen sistemas complejos en los cuales no es fácil aplicar técnicas analíticas. Sin embargo, la distribución exponencial se ha quedado «corta» en la actualidad y cada vez es más común utilizar otros tipos de distribuciones, como pueden ser la distribución de Weibull, la distribución gamma, la distribución de valores extremos, la distribución log-normal, etc. No obstante, y a pesar de que estas distribuciones son bien conocidas y ampliamente utilizadas de manera exitosa en diversas ramas de la ciencia, la ingeniería y la medicina (por ejemplo, y para seguir con los dispositivos mencionados por Long *et al.* [8] y Luo *et al.* [9], en el artículo de Pan *et al.* [12] se emplea la distribución de Weibull para modelizar los datos experimentales medidos en estos tipos de aparatos), a veces el ajuste que se obtiene no es del todo preciso y, en consecuencia, es necesario plantearse un enfoque diferente que mejore la rigurosidad del ajuste. Bajo este contexto, en los últimos años se ha venido utilizando con bastante frecuencia un enfoque basado en las distribuciones tipo fase, cuya flexibilidad y propiedades hacen que sean unas buenas candidatas para

obtener un mejor ajuste de los datos experimentales. A modo de ejemplo, y continuando con el tipo de memorias mencionadas a lo largo de la introducción, en el artículo de Acal *et al.* [1] se demuestra tras un estudio pormenorizado que el ajuste que se obtiene con las distribuciones tipo fase es más riguroso que el que se logra con la distribución de Weibull.

Las distribuciones tipo fase (PHD), las cuales fueron introducidas y analizadas en detalle por Neuts [10, 11], constituyen una clase de distribuciones no negativas que hacen posible modelizar problemas complejos con resultados bien estructurados gracias a su forma algebraico-matricial. Debido a sus valiosas propiedades, muchas variedades de esta clase de distribuciones han sido consideradas en diversas ramas del conocimiento y aplicadas en estudios de fiabilidad, procesos de renovación, teoría de colas y análisis de supervivencia [13, 14]. Casos particulares de las distribuciones tipo fase son las distribuciones exponencial, Erlang, Erlang generalizada, hipergeométrica y coxiana, entre otras. De hecho, no solo algunas distribuciones de probabilidad muy conocidas son distribuciones tipo fase, sino que cualquier distribución de probabilidad no negativa puede ser aproximada tanto como se desee mediante una PHD.

Finalmente, una vez que tenemos controlada la distribución, podemos extraer las principales propiedades del sistema a partir de las funciones que caracterizan a una distribución (función de densidad, función de distribución, etc.). En un análisis de fiabilidad habitual destaca el papel que juegan la función de fiabilidad y la función de riesgo (también conocida como función razón de fallo), y la aparición de datos censurados, que son aquellos tiempos de fallo de algunas unidades experimentales que se desconocen por algún motivo (por ejemplo, en un estudio de medicina donde estamos interesados en ver cómo afecta la dosis de un nuevo medicamento en el tiempo de vida de pacientes con cáncer, si un paciente deja el estudio, ya sea porque se ha mudado de ciudad o porque no quiere seguir perteneciendo a dicho estudio, este valor será desconocido y, por tanto, censurado). Aunque en el presente artículo no vamos a tratar los datos censurados, hay que tener en cuenta que es un concepto muy común en estos tipos de estudios y que deben ser tratados de manera correcta.

Este artículo está estructurado como sigue. En la sección 2 se describen las funciones típicas de un estudio de fiabilidad y las expresiones que adoptan estas funciones en cada una de las distribuciones más comunes, además de detallar una técnica ampliamente utilizada para ajustar distribuciones de probabilidad a los tiempos de fallo. La sección 3 está centrada en detallar y en definir brevemente las propiedades más importantes de las distribuciones tipo fase. Un ejemplo de simulación con el paquete estadístico R, donde se discute la modelización de distintas distribuciones a los datos simulados, se puede ver en la sección 4. Finalmente, tras las secciones mencionadas, figura un apartado dedicado a las conclusiones obtenidas en el presente trabajo.

2. Metodología básica en fiabilidad

En esta sección se definen una serie de medidas que se utilizan en el contexto de la fiabilidad y análisis de supervivencia. Se debe notar que para estas definiciones se va a considerar el tiempo de fallo del sistema, pero son extensibles para cualquier otra variable que esté relacionada con el tiempo de fallo.

2.1. Función de distribución, fiabilidad y razón de fallo en fiabilidad

La *función de distribución* de una variable aleatoria T , definida sobre el eje real positivo, se define como la función $F(t) = \mathbb{P}[T \leq t]$. Asimismo, su *función de densidad* se define como

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}[t \leq T < t + \Delta t]}{\Delta t} = \frac{dF(t)}{dt}.$$

Dado que la variable aleatoria T representa el tiempo de fallo (tiempo de vida), la *función de fiabilidad* representa la probabilidad de que no ocurra un fallo en el intervalo $(0, t)$ o, lo que es lo mismo, la probabilidad de que un sistema sobreviva en el intervalo $(0, t)$, de ahí que también sea conocida como *función de supervivencia*. Se representa como sigue:

$$R(t) = \mathbb{P}[T > t] = 1 - F(t).$$

Cabe recalcar que la gráfica de esta función es decreciente, ya que es la complementaria de la función de distribución; se inicia en $R(0) = 1$, y tiende a cero cuando $t \rightarrow \infty$. Otra medida que es muy utilizada en este ámbito es el *tiempo medio de fallo*, que se define como la esperanza de la variable,

$$\mathbb{E}[T] = \int_0^{\infty} t f(t) dt = \int_0^{\infty} R(t) dt.$$

Esta relación se obtiene aplicando el teorema de Fubini, como se puede ver a continuación:

$$\int_0^{\infty} R(t) dt = \int_0^{\infty} \int_t^{\infty} f(u) du dt = \int_0^{\infty} \int_0^u f(u) dt du = \int_0^{\infty} f(u) \int_0^u dt du = \int_0^{\infty} u f(u) du.$$

Por otro lado, la *función razón de fallo* o *razón de riesgo* se interpreta como la razón instantánea de fallo, y se define a partir de la siguiente relación:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}[t \leq T < t + \Delta t \mid T > t]}{\Delta t} = \frac{f(t)}{R(t)}.$$

Asimismo, dado que la función de densidad se puede expresar en términos de la función de fiabilidad, la expresión anterior se puede escribir

$$h(t) = -\frac{dR(t)/dt}{R(t)},$$

e integrando esta ecuación diferencial en el intervalo $(0, t)$ con la condición inicial $R(0) = 1$ se tiene que

$$R(t) = e^{-H(t)},$$

siendo

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

la *tasa de fallo acumulada* hasta el tiempo t . Esta relación muestra que la función de fallo caracteriza la distribución.

2.2. Expresiones de algunas distribuciones en fiabilidad

En el cuadro 1 se pueden observar las expresiones de las medidas definidas en el apartado anterior que adoptan algunas de las distribuciones más comunes en un estudio de fiabilidad. Además, se indican los límites de cada uno de los parámetros de los que depende cada una de las distribuciones definidas, donde $t > 0$ en todas las distribuciones salvo en la log-normal, que satisface $t \in \mathbb{R}$.

Como se comentó en la introducción del presente artículo, además de estas distribuciones, existen otras distribuciones de probabilidad que tienen un rol importante en el campo de la fiabilidad. Sin embargo, como el objetivo no es estudiar todas las distribuciones que se emplean para modelizar los tiempos de fallo, simplemente se procede a mencionar el nombre de estas distribuciones: Erlang, gamma, chi-cuadrado, Pareto, etc.

2.3. Análisis gráfico

Cuando la estimación de los parámetros utilizando el método de máxima verosimilitud presenta serias dificultades de cálculo, como es el caso de la distribución de Weibull, se emplea un análisis gráfico que facilita el ajuste de un conjunto de datos observados. Esta técnica gráfica es una técnica paramétrica que se basa en el principio de mínimos cuadrados, y suele ser aplicada debido a su sencillez y porque permite una primera idea gráfica del ajuste.

Grosso modo, lo que se hace es construir una nube de puntos a partir de los tiempos de fallo observados y a este conjunto de datos se le ajusta una recta por el criterio de mínimos cuadrados. La forma final de la nube de puntos dependerá de la distribución de probabilidad que se considere. Si el ajuste es bueno, lo que se valora en función del coeficiente de determinación, se acepta que el conjunto de valores observados sigue la distribución considerada y se calculan los parámetros de dicha distribución a partir de la recta estimada por mínimos cuadrados.

Cuadro 1: Expresiones que adoptan las distribuciones consideradas.

| | Exponencial | Weibull | Valores extremos (Gumbel) | Log-normal |
|--------|--------------------------|---|---|--|
| $F(t)$ | $1 - e^{-\lambda t}$ | $1 - e^{-(\lambda t)^\beta}$ | $1 - e^{-\gamma e^{\beta t}}$ | $\Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)$, con ϕ f. distr. $N(0, 1)$ |
| $f(t)$ | $\lambda e^{-\lambda t}$ | $\beta \lambda (\lambda t)^{\beta-1} e^{-(\lambda t)^\beta}$ | $(\beta \gamma) e^{\beta t - \gamma e^{\beta t}}$ | $\frac{1}{\sigma t \sqrt{2\pi}} e^{-\frac{(\ln(t) - \mu)^2}{2\sigma^2}}$ |
| $R(t)$ | $e^{-\lambda t}$ | $e^{-(\lambda t)^\beta}$ | $e^{-\gamma e^{\beta t}}$ | $1 - \Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)$ |
| $E[T]$ | $1/\lambda$ | $\frac{1}{\lambda} \Gamma\left(\frac{1}{\beta} + 1\right)$, con Γ f. gamma | $\int_0^\infty e^{-\gamma e^{\beta t}} dt$ | $e^{\mu + \frac{\sigma^2}{2}}$ |
| $h(t)$ | λ | $(\beta \lambda) (\lambda t)^{\beta-1}$ | $\beta \gamma e^{\beta t}$ | $f(\ln(t))/R(\ln(t))$ |
| Parám. | $\lambda > 0$ | $\lambda > 0, \beta > 0$ | $\gamma = \lambda^\beta, \lambda > 0, \beta > 0$ | $\mu \in \mathbb{R}, \sigma > 0$ |

2.3.1. Análisis gráfico en la distribución exponencial

Puesto que el análisis estadístico habitual que se utiliza en datos experimentales correspondientes a los tiempos de fallo de un sistema es mediante la distribución exponencial, se va a explicar cómo se aplica esta técnica y cómo se obtienen los parámetros de esta distribución en la práctica. Esto se puede extender para el caso de considerar otra distribución cualquiera.

Se parte de la función de fiabilidad de la distribución exponencial,

$$R(t) = e^{-\lambda t}, \quad t \geq 0, \quad \lambda > 0,$$

y se opera en esta expresión hasta conseguir una relación lineal. En consecuencia, si se toman logaritmos neperianos, queda

$$\ln(R(t)) = -\lambda t, \quad t \geq 0, \quad \lambda > 0,$$

o lo que es lo mismo,

$$\ln(1 - F(t)) = -\lambda t, \quad t \geq 0, \quad \lambda > 0.$$

Teniendo en cuenta esta última expresión, si despejamos el signo negativo que precede a t , la nube de puntos que hay que representar tiene la siguiente forma:

$$\{(t_i, -\ln(1 - p_i)), i = 1, 2, \dots, n\},$$

siendo t_i los tiempos de fallo observados (ordenados de menor a mayor) y p_i los valores que toma la función de distribución empírica en los t_i . En la literatura se han introducido varias elecciones posibles para estos valores p_i , como, por ejemplo, cualquiera de las tres aproximaciones siguientes, siendo la segunda opción la más utilizada:

$$p_i = \frac{i}{n}, \quad p_i = \frac{i - 0,5}{n}, \quad p_i = \frac{i}{n + 1}.$$

A este conjunto de datos se le ajusta una recta del tipo $y = ax + b$ por el principio de mínimos cuadrados y se estudia el ajuste. Si el ajuste es bueno, se acepta que el conjunto de valores observados sigue una distribución exponencial y, finalmente, a partir de la ecuación de la recta, se calcula la estimación del parámetro de la distribución, el cual en este caso coincide con la pendiente de la recta, es decir,

$$\hat{\lambda} = a.$$

2.3.2. Análisis gráfico en otras distribuciones

Extrapolando la teoría explicada en el anterior apartado, se pueden estimar los parámetros de cualquier distribución que se desee. En la cuadro 2 figuran las estimaciones de los parámetros de las distribuciones detalladas en el cuadro 1 aplicando la técnica del *análisis gráfico*, así como las nubes de puntos que hay que suponer en cada caso para calcular dichas estimaciones.

Cuadro 2: Estimaciones y nubes de puntos para cada distribución considerada.

| | Nube de puntos | Estimación de parámetros a partir de $y = ax + b$ |
|------------------|--------------------------------------|--|
| Weibull | $\{(\ln(t_i), \ln(-\ln(1 - p_i)))\}$ | $\hat{\beta} = a; \hat{\lambda} = e^{b/a}$ |
| Valores extremos | $\{(t_i, \ln(-\ln(1 - p_i)))\}$ | $\hat{\beta} = a; \hat{\gamma} = e^b$ |
| Log-normal | $\{(\ln(t_i), \phi^{-1}(p_i))\}$ | $\hat{\sigma} = \frac{1}{a}; \hat{\mu} = -\frac{b}{a}$ |

3. Distribuciones tipo fase

Como se ha comentado en la introducción del presente artículo, las distribuciones tipo fase fueron introducidas por Neuts en 1975. Estas distribuciones pueden ser definidas en el ámbito de cadenas de Markov y poseen propiedades interesantes (como, por ejemplo, la falta de memoria parcial o las propiedades de clausura) que hacen que dichas distribuciones sean consideradas en diversas ramas de la ciencia o ingeniería.

Una PHD se define como la distribución del tiempo hasta la absorción en una cadena de Markov con un estado o clase absorbente. El espacio de estados viene dado por un número general m de estados transitorios, donde la probabilidad de estar inicialmente en el estado i es α_i , y un estado absorbente, $m + 1$ (se considera que la cadena inicialmente no está en el estado absorbente, $\alpha_{m+1} = 0$). Además, la intensidad de transición del estado i al estado j viene dada por q_{ij} para $i \neq j$, y si $i = j$ entonces $q_{ii} = -\sum_{j=1, i \neq j}^{m+1} q_{ij}$. Una distribución tipo fase se representa, considerando solo los estados transitorios de la cadena de Markov asociada, a través del par (α, T) , siendo $\alpha = (\alpha_1, \dots, \alpha_m)$ y $T = (q_{ij})_{i,j=1,\dots,m}$.

La función de distribución de una PHD está dada por

$$F(t) = 1 - \alpha e^{Tt} \mathbf{e}, \quad t \geq 0,$$

donde \mathbf{e} es un vector columna de unos con el orden apropiado. A partir de esta definición se pueden obtener de forma inmediata las distintas expresiones definidas en la sección 2.1. Sin embargo, y como se utilizarán en la siguiente sección, se procede a mostrar seguidamente la expresión de la función de fiabilidad y de la función razón de fallo, respectivamente, para mayor claridad:

$$R(t) = \alpha e^{Tt} \mathbf{e}, \quad t \geq 0, \quad \text{y} \quad h(t) = \frac{\alpha \mathbf{e}^{Tt} T^0}{\alpha \mathbf{e}^{Tt} \mathbf{e}}, \quad t \geq 0,$$

siendo $T^0 = -T\mathbf{e}$ el vector de intensidad de transición desde un estado transitorio hasta un estado de absorción.

Las distribuciones tipo fase son importantes no solo por su estructura, sino también por las buenas propiedades que permiten la aplicabilidad e interpretación de los resultados de manera sencilla. Las propiedades de estas distribuciones han sido estudiadas exhaustivamente en los últimos años. De hecho, en el libro de He [7] puede verse un estudio actual en el que se revisan las propiedades más esenciales e importantes de las PHD. Sin embargo, la razón fundamental por la cual esta clase de distribuciones es tan atractiva en estos tipos de datos es debido a que la clase PHD es densa en el conjunto de distribuciones de probabilidad no negativas [2]. Este resultado implica que cualquier distribución de probabilidad no negativa puede ajustarse tanto como se desee mediante una PHD, por lo que dicha flexibilidad hace que las PHD sean unas buenas candidatas para obtener un mejor ajuste.

No obstante, a continuación se enumeran otras propiedades interesantes e importantes:

1. Constituyen una clase de distribuciones no negativas que permite describir las principales medidas asociadas en una forma algorítmica.
2. La clase de PHD es cerrada bajo una serie de operaciones: mínimo, máximo, suma, etc.
3. Generalizan un gran número de distribuciones conocidas:
 - Distribución exponencial: $F(t) = 1 - e^{-\lambda t}$ para $t \geq 0$, $\alpha = 1$, $T = -\lambda$ y $m = 1$.
 - Distribución de Erlang: $F(t) = 1 - \sum_{j=0}^{m-1} e^{-\lambda t} (\lambda t)^j / j!$ para $t \geq 0$, $m \geq 1$ y $\lambda > 0$,

$$\alpha = (1, 0, \dots, 0, 0), \quad T = \begin{pmatrix} -\lambda & \lambda & & \\ & -\lambda & \ddots & \\ & & \ddots & \lambda \\ & & & -\lambda \end{pmatrix}_{m \times m}.$$

- Distribución hipoexponencial: $F(t) = 1 - \sum_{x=0}^t \sum_{i=1}^m e^{-\lambda_i x} \left(\prod_{j=1; j \neq i}^m \frac{\lambda_j}{\lambda_j - \lambda_i} \right)$ para $t \geq 0$, $\lambda_i \neq \lambda_j$ con $i \neq j$,

$$\alpha = (1, 0, \dots, 0, 0), \quad T = \begin{pmatrix} -\lambda_1 & \lambda_1 & & \\ & -\lambda_2 & \ddots & \\ & & \ddots & \lambda_{m-1} \\ & & & -\lambda_m \end{pmatrix}_{m \times m}.$$

- Distribución hiperexponencial: $F(t) = 1 - \sum_{i=1}^m \alpha_i (1 - e^{-\lambda_i t})$ para $t \geq 0$,

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m), \quad T = \begin{pmatrix} -\lambda_1 & & & \\ & -\lambda_2 & & \\ & & \ddots & \\ & & & -\lambda_m \end{pmatrix}_{m \times m}.$$

- Distribución coxiana: $F(t) = 1 - \sum_{x=0}^t \sum_{i=1}^m e^{-\lambda_i x} \left(\prod_{j=1; j \neq i}^m \frac{g_j \lambda_j}{\lambda_j - \lambda_i} \right)$ para $t \geq 0$, $\lambda_i \neq \lambda_j$ con $i \neq j$,

$$\alpha = (1, 0, \dots, 0, 0), \quad T = \begin{pmatrix} -\lambda_1 & g_1 \lambda_1 & & \\ & -\lambda_2 & \ddots & \\ & & \ddots & g_{m-1} \lambda_{m-1} \\ & & & -\lambda_m \end{pmatrix}_{m \times m}.$$

- Distribución coxiana generalizada: $F(t) = 1 - \sum_{i=1}^m \alpha_i \left(\sum_{x=0}^t \sum_{i=1}^m e^{-\lambda_i x} \left(\prod_{j=1; j \neq i}^m \frac{g_j \lambda_j}{\lambda_j - \lambda_i} \right) \right)$ para $t \geq 0$, $\lambda_i \neq \lambda_j$ con $i \neq j$,

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m), \quad T = \begin{pmatrix} -\lambda_1 & g_1 \lambda_1 & & \\ & -\lambda_2 & \ddots & \\ & & \ddots & g_{m-1} \lambda_{m-1} \\ & & & -\lambda_m \end{pmatrix}_{m \times m}.$$

Finalmente, en lo que a las PHD se refiere, cabe decir que para la estimación de los parámetros de estas distribuciones no se puede utilizar el análisis gráfico definido en la sección anterior, porque las PHD no se pueden linealizar. En consecuencia, se recurre a un método iterativo denominado *algoritmo esperanza-maximización*, más conocido como algoritmo EM (desarrollado por Asmussen, Nerman y Olsson [3] y presentado en el libro de Buchholz, Kriege y Felko [5]), que alterna dos pasos, esperanza y maximización, para obtener la estimación de los parámetros por máxima verosimilitud. En la actualidad existen paquetes implementados en Matlab o en R (véase el paquete `mapfit`), e incluso aplicaciones disponibles para ordenador (por ejemplo, `EMpht`), tanto para determinar la estructura de las distribuciones tipo fase como para la estimación de sus parámetros.

4. Simulación con R

El objetivo de la presente sección es demostrar el poder de ajuste de las distribuciones tipo fase. Para ello, se han simulado cien valores de una distribución no negativa (más concretamente, cien valores de una distribución uniforme en el intervalo $[0,3]$) y, seguidamente, se ha procedido a ajustar las distribuciones de Erlang, Weibull, valores extremos y log-normal a los datos simulados (ya ordenados) mediante la técnica del análisis gráfico. La bondad del ajuste de cada una de las distribuciones mencionadas se ha comparado por medio del coeficiente de determinación (R^2) obtenido en cada caso, eligiendo como óptima aquella cuyo valor R^2 esté más cercano a 1. Seleccionada la distribución más precisa, se procede, en primer lugar, a estimar la distribución tipo fase que mejor se ajusta a estos datos y, posteriormente, a comparar el ajuste obtenido en ambas distribuciones.

En la figura 1 aparece el ajuste de los datos simulados con las distribuciones descritas en la sección 2.2 utilizando el análisis gráfico: la gráfica (A) representa el ajuste mediante la distribución exponencial, la gráfica (B) denota el ajuste a través de la distribución de Weibull, la gráfica (C) muestra el ajuste por medio de la distribución de valores extremos, y la gráfica (D) revela el ajuste mediante la distribución log-normal.

A tenor del gráfico y apoyándose en el cuadro 3, en el cual figura la estimación de mínimos cuadrados de los parámetros de cada distribución y el coeficiente de determinación alcanzado en cada caso, se concluye que la distribución que mejor funciona para los datos simulados es la distribución de Weibull. En consecuencia, esta será la distribución elegida para compararla con la distribución tipo fase al final de la presente sección.

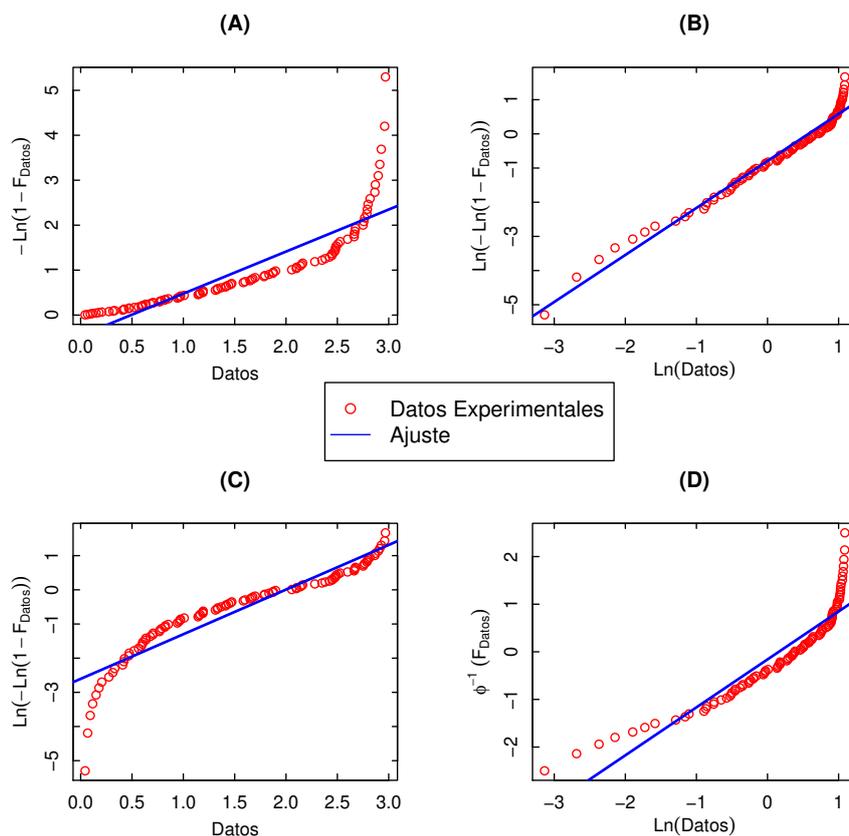


Figura 1: Ajuste por mínimos cuadrados de los datos simulados a través del método de análisis gráfico: (A) distribución exponencial, (B) distribución de Weibull, (C) distribución de valores extremos, (D) distribución log-normal. Para más claridad se indica la nube de puntos considerada en cada caso.

Seguidamente, se han ajustado distintas distribuciones tipo fase a los datos simulados usando el algoritmo EM, asumiendo que la matriz que contiene las intensidades transitorias hasta la absorción, es decir, la matriz T , tiene estructura interna general. Esto da lugar a que se estimen $m(m + 1) + m$ parámetros, siendo

Cuadro 3: Coeficiente de determinación alcanzado y parámetros estimados por mínimos cuadrados tras aplicar el método de análisis gráfico a los datos simulados considerando distintas distribuciones.

| | R^2 | Parámetros estimados |
|------------------|-------|--|
| Exponencial | 0,75 | $\hat{\lambda} = 0,935$ |
| Weibull | 0,97 | $\hat{\lambda} = 0,562; \hat{\beta} = 1,375$ |
| Valores extremos | 0,87 | $\hat{\gamma} = 0,07; \hat{\beta} = 1,31$ |
| Log-normal | 0,86 | $\hat{\mu} = 0,16; \hat{\sigma} = 0,99$ |

m el número de estados transitorios. Después de este análisis, se ha observado que la estructura interna y el vector que contiene las probabilidades iniciales no adoptan una expresión concreta para un m fijo, sino que van variando según el número de estados que se considere. En base a ello, no se puede asumir ninguna distribución tipo fase conocida a los datos.

Los parámetros de la distribución tipo fase han sido estimados aplicando el algoritmo EM con la aplicación EMpht y con el paquete `mapfit` de R. Se empieza con un número pequeño de estados y se va aumentando hasta que se consigue el ajuste óptimo, el cual es alcanzado para los datos simulados en veinte fases. La representación de dicha distribución tipo fase se puede apreciar en el apéndice A que figura al final del artículo. Una vez estimados los parámetros de la distribución tipo fase, se procede a observar gráficamente la precisión del ajuste de esta distribución a los datos simulados. La tasa de fallo acumulada experimental estimada por la distribución tipo fase se muestra en la figura 2. Cabe notar que en este caso, y con el objetivo de utilizar de nuevo la técnica del análisis gráfico, en el eje de ordenadas se representa $-\ln(1 - F)$ y en el eje de abscisas, los datos ordenados (esta transformación es una generalización de la realizada en el caso exponencial).

Finalmente, se procede a comparar los ajustes obtenidos a través de la distribución de Weibull y de la distribución tipo fase. Para ello, se utilizan la función de fiabilidad y la función tasa de riesgo. En la figura 3 se muestra, por un lado, la función de fiabilidad experimental, además del ajuste por Weibull y tipo fase (A) y, por otro lado, la tasa de riesgo experimental frente al ajuste por Weibull y tipo fase.

A raíz de los gráficos de la figura 3, el mejor resultado se logra cuando se considera la distribución tipo fase con veinte estados transitorios (fases) y puede apreciarse que la precisión es bastante notable.

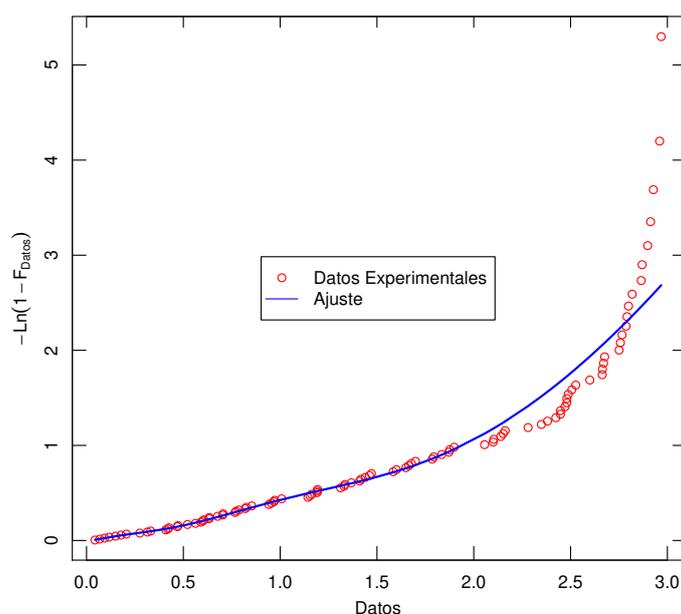


Figura 2: Tasa de fallo acumulada de los datos simulados y el correspondiente ajuste de la distribución tipo fase con veinte fases.

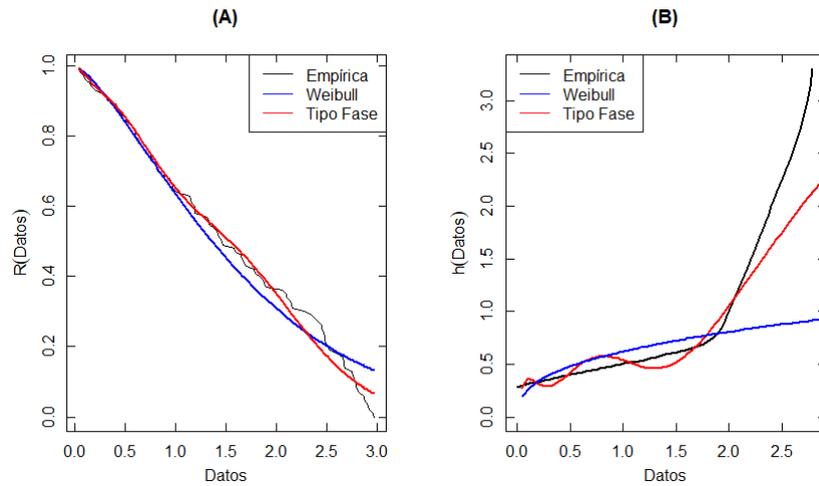


Figura 3: (A) Función de fiabilidad $R(t)$ frente a los datos simulados y las distribuciones de Weibull y tipo fase. (B) Tasa de riesgo $h(t)$ frente a los datos simulados y las distribuciones de Weibull y tipo fase.

5. Conclusiones

El análisis estadístico habitual realizado con datos experimentales en estudios de fiabilidad para caracterizar las principales razones de fallo de cualquier sistema es a través de distribuciones continuas, como pueden ser la distribución de Weibull, exponencial, gamma, log-normal, etc. Sin embargo, a veces el ajuste obtenido a los datos medidos no es del todo riguroso. Este hecho sugiere que otras distribuciones estadísticas podrían funcionar de una mejor manera y el ajuste sería más preciso. A este respecto, en los últimos años es cada vez más común utilizar un enfoque basado en las distribuciones tipo fase, ya que cualquier distribución no negativa puede aproximarse tanto como sea necesario mediante una distribución tipo fase. Cabe destacar también las buenas propiedades algebraico-matricial y de clausura que posee la clase de distribuciones tipo fase y su fácil algoritmización que permite obtener expresiones de las cantidades de interés del modelo en forma bien estructurada, de modo que puedan ser tratadas computacionalmente de manera rápida y eficaz, facilitando así la estimación de los parámetros.

Para poner de manifiesto el poder de precisión de estas distribuciones, en el presente trabajo se han simulado cien valores de una distribución uniforme en el intervalo $[0, 3]$, y se han ajustado y comparado distintas distribuciones. Después de un estudio exhaustivo y pormenorizado se ha concluido que la distribución que mejor funciona para los datos simulados es la distribución tipo fase. Por lo tanto, y en consecuencia de todo lo comentado a lo largo del artículo, parece evidente concluir que las distribuciones tipo fase deben ser tenidas en cuenta cuando se quiera analizar datos experimentales referidos al fallo de cualquier sistema.

Referencias

- [1] ACAL, Christian; RUIZ-CASTRO, Juan Eloy; AGUILERA, Ana María; JIMÉNEZ-MOLINOS, Francisco, y ROLDÁN, Juan B. «Phase-type distributions for studying variability in resistive memories». En: *Journal of Computational and Applied Mathematics* 345 (2019), págs. 23-32. ISSN: 0377-0427. <https://doi.org/10.1016/j.cam.2018.06.010>.
- [2] ASMUSSEN, Søren. *Ruin Probabilities*. Hong Kong: World Scientific, 2000. <https://doi.org/10.1142/7431>.
- [3] ASMUSSEN, Søren; NERMAN, Olle, y OLSSON, Marita. «Fitting Phase-Type Distributions via the EM Algorithm». En: *Scandinavian Journal of Statistics* 23.4 (1996), págs. 419-441. ISSN: 03036898. URL: <http://www.jstor.org/stable/4616418>.
- [4] BAZOVSKY, Igor. *Reliability theory and practice*. Prentice-Hall, 1961.

- [5] BUCHHOLZ, Peter; KRIEGE, Jan, y FELKO, Iryna. *Input Modeling with Phase-Type Distributions and Markov Models. Theory and Applications*. 1.^a ed. SpringerBriefs in Mathematics. An optional note. Cham: Springer, 2014. <https://doi.org/10.1007/978-3-319-06674-5>.
- [6] EPSTEIN, Benjamin y SOBEL, Milton. «Life Testing». En: *Journal of the American Statistical Association* 48.263 (1953), págs. 486-502. <https://doi.org/10.1080/01621459.1953.10483488>.
- [7] HE, Qi-Ming. *Fundamentals of Matrix-Analytic Methods*. New York: Springer, 2014. <https://doi.org/10.1007/978-1-4614-7330-5>.
- [8] LONG, Shibing; CAGLI, Carlo; IELMINI, Daniele; LIU, Ming, y SUÑÉ, Jordi. «Analysis and modeling of resistive switching statistics». En: *Journal of Applied Physics* 111.7, 074508 (2012). <https://doi.org/10.1063/1.3699369>.
- [9] LUO, Wun-Cheng; LIU, Jen-Chieh; FENG, Hsien-Tsung; LIN, Yen-Chuan; HUANG, Jiun-Jia; LIN, Kuan-Liang, y HOU, Tuo-Hung. «RRAM SET speed-disturb dilemma and rapid statistical prediction methodology». En: *2012 International Electron Devices Meeting*. Dic. de 2012, págs. 9.5.1-9.5.4. <https://doi.org/10.1109/IEDM.2012.6479012>.
- [10] NEUTS, Marcel F. «Probability distributions of phase type». En: *Liber Amicorum Professor Emeritus Dr. H. Florin* (1975), págs. 173-206.
- [11] NEUTS, Marcel F. *Matrix-geometric solutions in stochastic models: an algorithmic approach*. Courier Corporation, 1994. ISBN: 978-0-486-68342-3.
- [12] PAN, Feng; GAO, Shuang; CHEN, Chao; SONG, Chen, y ZENG, Fei. «Recent progress in resistive random access memories: Materials, switching mechanisms, and performance». En: *Materials Science and Engineering: R: Reports* 83 (2014), págs. 1-59. ISSN: 0927-796X. <https://doi.org/10.1016/j.mser.2014.06.002>.
- [13] RUIZ-CASTRO, Juan Eloy y DAWABSHA, Mohammed. «A discrete MMAP for analysing the behaviour of a multi-state complex dynamic system subject to multiple events». En: *Discrete Event Dynamic Systems* (dic. de 2018). ISSN: 1573-7594. <https://doi.org/10.1007/s10626-018-0274-0>.
- [14] RUIZ-CASTRO, Juan Eloy; DAWABSHA, Mohammed, y ALONSO, Francisco Javier. «Discrete-time Markovian arrival processes to model multi-state complex systems with loss of units and an indeterminate variable number of repairpersons». En: *Reliability Engineering & System Safety* 174 (2018), págs. 114-127. ISSN: 0951-8320. <https://doi.org/10.1016/j.res.2018.02.019>.

A. Apéndice

La representación de la distribución tipo fase lograda en la sección 4 es la siguiente:

1. El vector α toma los valores

(0,5416, 0,1237, 0,2346, 0, 0,0515, 0,0003, 0,0001, 0, 0,0026, 0, 0, 0,0121, 0, 0, 0,0215, 0,0001, 0, 0, 0,0119, 0).

2. La matriz T adopta la expresión que se ve en la página siguiente, en la figura 4.

$$\begin{pmatrix}
 -5.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5.9 & 0 & 0 & 0 \\
 0 & -8 & 0 & 0 & 0.18 & 4.17 & 0 & 0 & 0 & 0 & 0 & 0.87 & 0 & 0 & 0.02 & 2.77 & 0 & 0 & 0 & 0 \\
 0 & 0.85 & -18.91 & 0 & 3.67 & 2.85 & 0 & 0 & 0.02 & 0 & 0 & 2.64 & 0 & 0 & 0.5 & 0.87 & 0 & 7.43 & 0.08 & 0 \\
 0 & 0 & 0 & -5.9 & 0 & 0 & 5.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0.44 & 0 & 0 & -6.17 & 3.9 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0 & 0 & 0.05 & 1.47 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & -6.51 & 0 & 0 & 0 & 3.94 & 0 & 0 & 0 & 0.85 & 0 & 1.72 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & -5.9 & 0 & 1.87 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4.02 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5.9 \\
 0 & 0.77 & 0 & 0 & 1.39 & 0 & 0 & 0 & -6.15 & 0 & 0 & 0.28 & 0 & 0 & 3.39 & 0 & 0 & 0 & 0.32 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -7.07 & 2.38 & 0 & 0 & 4.69 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -6.08 & 0 & 0 & 0 & 0 & 0 & 0 & 6.08 & 0 & 0 \\
 0 & 0.06 & 0 & 0 & 0.18 & 0.87 & 0 & 0 & 0 & 0.02 & 0 & -7.58 & 0 & 0.02 & 0 & 6.42 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 5.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.06 & 6.69 & 0 & 0 & -6.75 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 2.78 & 0 & 0 & 0.17 & 1.94 & 0 & 0 & 0 & 0 & 0 & 1.47 & 0 & 0 & -6.9 & 0.53 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0.01 & 0 & 0 & 0 & 2.57 & 0 & 0 & 0 & 3.94 & 0 & -6.53 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 5.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5.9 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -6.77 & 0 & 0 \\
 0 & 0.33 & 0 & 0 & 3.01 & 0 & 0 & 0 & 0 & 0.07 & 0 & 0 & 0.18 & 0 & 0 & 2.41 & 0 & 0 & 0 & -6 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5.9
 \end{pmatrix}$$

Figura 4: Expresión de la matriz T .