# Tema 2

# Modelos aleatorios de tiempos de vida

Numerosos modelos paramétricos son utilizados en el análisis de datos de tiempos de vida y en problemas relacionados con la modelización del envejecimiento y procesos de fallo. Entre los modelos univariantes, son unas pocas distribuciones las que toman un papel fundamental dado su demostrada utilidad en casos prácticos. Así se tiene la distribución exponencial, Weibull, log-normal, valores extremos, Gamma, logística y tipo fase. Por otro lado, generalmente, existen factores, externos o internos, que actúan sobre el tiempo de vida debiendo ser estudiado. Esto da lugar a los modelos de regresión que generalizan a los anteriores modelos teniendo en cuenta la información de estos factores de cada individuo.

## *2.1. El modelo de función de riesgo constante: La distribución exponencial*

Nos detenemos en esta sección a estudiar en profundidad la distribución exponencial dada su sencillez en el cálculo y las importantes propiedades que posee, que la convierten en un modelo utilizado en multitud de situaciones prácticas.

### 2.1.1. Definición

La distribución exponencial juega un papel fundamental en análisis de supervivencia, se trata de la distribución más usada en el análisis de datos de tiempo de vida.

Comprobamos que la distribución exponencial se obtiene como una forma límite de la distribución geométrica.

Supongamos un individuo sujeto a fallos tal que hay una probabilidad constante de que el fallo aparezca en un intervalo de tiempo corto, de amplitud $\delta$. El fallo causa el fin del estudio. El entorno que genera este fallo no tiene “memoria”, y en cada $\delta$–periodo, no importa lo que haya ocurrido antes, el dispositivo se enfrenta a dos posibilidades : ocurre un fallo con probabilidad p o no ocurre ninguno, con probabilidad $1-p$. Llamamos $\tau$ al tiempo de vida del individuo, entonces

$$P(\tau > t) = P\{\text{el fallo no apareció durante } K=[t/\delta]+1 \text{ periodos}\} = (1-p)^K.$$

Si K es grande y p pequeño, se puede tomar la siguiente aproximación

$$P\{\tau > t\} = (1-p)^K \approx \exp(-Kp) .$$

Si tomamos $\delta=1$, tenemos $P\{\tau > K\} = \exp(-Kp)$. La v.a., $\tau$ con función de supervivencia o función de fiabilidad $P\{\tau > t\} \approx \exp(-t\lambda)$, con $\lambda > 0$, se llama *distribución exponencial de parámetro $\lambda$*, escribimos exp($\lambda$). La función de distribución de esta v.a. viene dada por

$$F(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases},$$

y la correspondiente función de densidad

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}.$$

Es claro que la distribución exponencial tiene función de riesgo constante e igual a $\lambda$, que es a su vez el inverso de la media de la distribución y de la desviación típica.

### 2.1.2. Propiedad de no memoria

La siguiente propiedad caracteriza a la distribución exponencial, es la *propiedad de no memoria*, y se enuncia como sigue. Sea X una v.a. con distribución exponencial entonces

$$P\{X > x + t | X > x\} = P\{X > t\},$$

es decir, si interpretamos X como el tiempo de vida de un individuo, esta igualdad establece que la probabilidad de que el individuo sobreviva al tiempo x+t dado que ha sobrevivido al tiempo x, es igual a la probabilidad de que sobreviva al tiempo t al principio del estudio. La condición anterior puede también escribirse en términos de la función de supervivencia

$$S(x + t) = S(x)S(t).$$

Esta propiedad describe un proceso de vida sin envejecimiento, y caracteriza a la función exponencial. Esta última condición puede compararse con la relación equivalente obtenida para la distribución geométrica en el caso discreto.

***Teorema.*** Sea X una v.a. no negativa cuya función de supervivencia $S(x)$ satisface la condición $S(x + t) = S(x)S(t)$ (I), entonces $S(x) = \exp(-\lambda x)$, es decir X es una v.a. exponencial.

*Demostración*. Sea S(x) = 1−F(x), donde F(x) es la función de distribución de una v.a., X, tal que satisface (I), entonces, en particular, considerando x = t = 1/n, S(2/n)=S(1/n)S(1/n). Repitiendo este proceso m veces,

$$S(m/n) = \left(S(1/n)\right)^m,$$

y, tomando m=n,

$$S(1) = \left(S(1/n)\right)^n,$$

para todo n. De estas dos últimas expresiones se deduce que

$$S(m/n) = \left(S(1)\right)^{m/n},$$

para todo m/n número racional. De aquí, por la continuidad de S y por la propiedad de densidad de Q en R,

$$S(x) = \left(S(1)\right)^x,$$

para todo $x \in R$. Por otro lado, $S(1)=\left(S(1/2)\right)^2 \geq 0$, de modo que la expresión última representa una función exponencial, que podemos escribir

$$S(x) = e^{-\lambda x},$$

con $\lambda = -\ln S(1)$. Sólo queda comprobar que $\lambda > 0$, para ello observemos que $S(1) \le S(0) = 1$, de modo que si tomamos logaritmos, $\ln S(1) \le 0$, con lo que $\lambda > 0$. El caso $\lambda = 0$ es un caso trivial. □

La v.a. exponencial ha sido considerada como un buen modelo para el tiempo de fallo T, de algunos sistemas. En este caso F(t) se entiende como la probabilidad de que el sistema falle en el intervalo (0,t] y S(t) la probabilidad de que sobreviva.

### 2.1.3. El modelo de vida exponencial

Vamos a ver que, bajo ciertas consideraciones, el tiempo de vida de un individuo puede modelarse como una v.a. exponencial. Estas consideraciones las expresamos formalmente como *postulados.* Supongamos que el fallo ocurre sujeto a las siguientes hipótesis

(i) la ocurrencia del fallo en un intervalo de tiempo $[t_0, t_1]$, depende sólo de $t_1 - t_0$, y es independiente de $t_0$ ;
(ii) la probabilidad de que el sistema falle en un intervalo infinitesimal [t,t+h] es $\lambda h + o(h)$, con $\lambda > 0$ ;
(iii) la probabilidad de que el fallo ocurra en t=0 es 0.

Sea S(t)=P{T>t}, se puede demostrar que bajo las hipótesis anteriores, esta función es exponencial.

*Demostración.* De las hipótesis (i) y (ii) se deduce

$$S(t+h) = S(t)(1-\lambda h) + o(h),$$

para todo h>0, o dicho de otro modo

$$\frac{S(t+h)-S(t)}{h} = -\lambda S(t) + \frac{o(h)}{h};$$

y si tomamos límite cuando $h \to 0$, obtenemos la ecuación diferencial

$$\frac{d}{dt} S(t) = -\lambda S(t),$$

que tiene como solución,

$$S(t) = K e^{-\lambda x}.$$

Para determinar el valor de la constante K, hacemos uso de (iii) según la cual $S(0) = 1$, con lo que K=1. □

### 2.1.4. Propiedades de la distribución exponencial

La propiedad de no memoria puede expresarse en términos de la función de riesgo, en este caso significaría que el riesgo de fallo es constante, es decir $h(t) = \lambda$, para todo t. Esto implica que, para un individuo, la probabilidad de fallo en el siguiente intervalo infinitesimal de tiempo es independiente de su tiempo de vida desde el inicio del estudio. Físicamente, una función de riesgo constante sugiere que la población bajo consideración no se está desgastando o dicho de otro modo, no está envejeciendo.

***Propiedad 1. Función generatriz de momentos***

Viene dada por la siguiente expresión

$$M(s) = \frac{\lambda}{\lambda - s}\ ,\ \text{para } s<\lambda\ .$$

A partir de aquí, $E[X]=1/\lambda$ y $Var(X)=1/\lambda^2$, con lo que para cualquier distribución exponencial el coeficiente de variación es igual a 1.

***Propiedad 2. Generación aleatoria de valores de una variable exponencial***

Si U es una v.a. con distribución U(0,1), entonces $T=-\ln(U)$ es una v.a. con distribución exp(1).

Esta propiedad permite generar valores de una distribución exponencial a partir de la distribución uniforme: se generan valores aleatorios para U, entonces $T= -\ln(U)$ es una v.a. exponencial de parámetro 1 y $T/\lambda$ es una distribución exp($\lambda$).

***Propiedad 3. Suma de distribuciones exponenciales independientes idénticas***

Sean $T_1, T_2, ...T_n$, n v.a.i.i.d. exponenciales de parámetro $\lambda$, entonces la variable $S = \sum_{i=1}^{n} T_i$ , tiene distribución conocida, es una Erlang de índice n y parámetro $\lambda$.

***Propiedad 4. Propiedad de mínimo***

Sean $T_1, T_2, ...T_n$, v.a. independientes y con distribuciones exponenciales de parámetros $\lambda_1, \lambda_2, ...,\lambda_n$, respectivamente. Sea $T_{(1)}=min\{T_1, T_2, ...T_n, \}$. Entonces, $T_{(1)}$, sigue una distribución exp($\lambda_1+ \lambda_2+ ...+\lambda_n$).

En particular si todas las $T_i$ son idénticas y llamamos $\lambda$ al valor del parámetro, $T_{(1)}$, sigue una distribución exp($n\lambda$). Con lo cual, como consecuencia particular se tiene que el valor medio del mínimo es $1/n\lambda= E[T_i]/n$.

***Propiedad 5. Muestras ordenadas con reemplazamiento***

Sean $T_{(1)}\leq T_{(2)}\leq...\leq T_{(n)}$, los estadísticos ordenados de una muestra aleatoria $T_1, T_2, ...T_n$, de v.a.i.i.d. con distribución exp($\lambda$). Esta situación puede utilizarse para modelar los tiempos de vida de una muestra de n ítems cuyo tiempo de vida para todos siguen distribuciones exponenciales idénticas. Se inicia la observación en $T_{(0)}=0$, cuando transcurre el primer fallo, en $t_{(1)}$, el ítem es inmediatamente reemplazado por otro idéntico, y así sucesivamente. Llamamos $W_1=T_{(1)}$, $W_2=T_{(2)}-T_{(1)}$, ..., $W_n=T_{(n)}-T_{(n-1)}$, los tiempos entre fallos consecutivos. Es claro que $W_1$ tiene distribución exp($\lambda n$), por la propiedad 4. Como la distribución exponencial no tiene memoria y los ítems que fallan son reemplazados inmediatamente, después de cada fallo la situación es idéntica a la inicial, y por tanto, $W_i$ representa el mínimo de n vv.aa. exponenciales independientes, porque siempre es el tiempo en que ocurre el primero de n fallos, de modo que cada $W_i$ tiene distribución exponencial de parámetro $\lambda n$.

### *Propiedad 6. Muestras ordenadas sin reemplazamiento*

La misma situación que en 5 pero ahora el ítem que falla no es reemplazado. De modo que ahora $W_i$ tiene distribución exponencial de parámetro $(n-i+1)\lambda$, para cada $i=1,2,...,n$.

### *Propiedad 7. Distribución exponencial truncada*

Obtenemos una extensión de la distribución exponencial cuando se introduce un parámetro umbral, en este caso la función de densidad viene dada por

$$f(t) = \lambda \exp\{-\lambda(t-\mu)\}, \quad t \geq \mu \ ,$$

Dada una v.a. con función de densidad f puede obtenerse la distribución truncada en un intervalo, $[a,+\infty)$, como aquella distribución $f_{[a,\infty)}$, que verifica $f_{[a,\infty)}(t)=kf(t)$ y k es tal que $\int_a^\infty kf(t)dt = 1$. Por tanto la distribución exponencial con dos parámetros tal como la hemos definido no es más que la distribución exponencial truncada en $[\mu,+\infty)$.

Este tipo de distribuciones puede usarse para modelizar situaciones como la siguiente. Supongamos que se estudia el efecto de un tratamiento en individuos enfermos pasado un primer período de choque de dicho tratamiento. Si el enfermo recae es tratado como se hizo inicialmente. El tiempo de vida en estudio empieza a contar a partir de que ha transcurrido este periodo de tiempo.

### *Propiedad 8. Comparación de exponenciales*

Dada la sencillez que presenta en el cálculo esta distribución, es posible obtener expresiones que sería muy complicadas en otros casos.

Sean $T_1$ y $T_2$ dos vv.aa. independientes y con distribución exponencial de parámetros $\lambda_1$, $\lambda_2$, respectivamente. ¿Cuál es la probabilidad de que $T_2>T_1$?

$$P\{T_2 > T_1\} = \frac{\lambda_1}{\lambda_1 + \lambda_2} \ ,$$

Este resultado tiene aplicación en el estudio de tiempos de vida de enfermedades con distintas causas de fallo. Supongamos en este caso que hay dos causas que provocan la recaída, donde ambas causas son independientes y tienen tiempos de vida exponenciales como las anteriores. Entonces la probabilidad de que recaiga debido a la primera causa es $\lambda_1/(\lambda_1+\lambda_2)$. Se puede extender al caso de n causas distintas e independientes, en ese caso la probabilidad de recaída debido a la causa j es $\lambda_j/(\lambda_1+\lambda_2+...+\lambda_n)$.

## ***2.2. El modelo** Gamma*

Una variable aleatoria de tiempo de vida se dice que se distribuye según una Gamma de parámetros $\alpha$ y $\lambda$, $G(\alpha,\lambda)$, cuando su función de densidad viene dada por

$$f(t) = \frac{\lambda^\alpha t^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda t} \ ,$$

$\alpha$ se llama *parámetro de forma* y $\lambda$ es el *parámetro de escala*, y $\Gamma$ representa la *función gamma* que se define $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ , para todo $\alpha>0$, y verifica que $\Gamma(n)=(n-1)!$, para todo $n \in N$. Cuando $\alpha=1$, se tiene la distribución exponencial. La función de distribución recibe el nombre de función gamma incompleta y no se puede evaluar exactamente sino que está tabulada.

### 2.2.1. La distribución Erlang

Un caso particular de la distribución Gamma es la distribución de Erlang. Esta puede obtenerse a partir de la distribución exponencial como vemos a continuación. Supongamos un individuo sometido a fallos que llegan aleatoriamente en el tiempo de tal forma que los tiempos entre dos fallos consecutivos son vv.aa. independientes y exponenciales idénticas, de parámetro $\lambda$. Se supone que el individuo recae cuando sufre el golpe número k. Llamamos T al tiempo de vida del individuo, entonces

$$T = \tau_1+\tau_2+...+\tau_k,$$

siendo $\tau_i$ el tiempo transcurrido entre el fallo $i-1$ y el fallo i. T es la suma de k vv.aa. independientes, por lo tanto su función de densidad f, viene dada por la convolución de las funciones de densidad de la k variables, con lo que se obtiene que

$$f(t) = \frac{\lambda^n \, t^{n-1}}{(k-1)!} e^{-\lambda t} \quad ; \quad t > 0 \ ,$$

que es la función de densidad de una *distribución Erlang de parámetro $\lambda$ e índice k*, $E(k,\lambda)$.

### 2.2.2. Función de riesgo

Vamos a estudiar el carácter de la función de riesgo según los valores de los parámetros

$$h(t) = \frac{t^{\alpha-1} e^{-\lambda t}}{\int_t^{\infty} x^{\alpha-1} e^{-\lambda x} dx} \ ,$$

es una función decreciente en t si $\alpha<1$, es constante si $\alpha=1$ y es creciente en t si $\alpha>1$.

*Demostración.* En lugar de estudiar la monotonía de la función h(t), estudiamos la función 1/h(t), que podemos escribir, previo cambio de variable

$$\frac{1}{h(t)} = \int_0^{\infty} (1+u/t)^{\alpha-1} \, e^{-\lambda u} du \ .$$

La monotonía de esta función depende del comportamiento de

$$\left(1+\frac{u}{t}\right)^{\alpha-1}$$

como función de t. Ésta última es una función decreciente en t si y sólo si $\alpha>1$, es creciente si y sólo si $\alpha<1$ y es constante si y sólo si $\alpha=1$. □

Además h(t) se aproxima asintóticamente a $1/\lambda$ cuando $t\to\infty$, lo cual sugiere que la distribución gamma puede ser útil como un modelo de población cuando los individuos son sometidos a un programa de seguimiento regular. La razón de fallo puede crecer o decrecer algo inicialmente, pero después de algún tiempo la enfermedad tiende a estabilizarse y a partir de ahí el fallo es tan probable en un intervalo de tiempo como en otro de la misma amplitud.

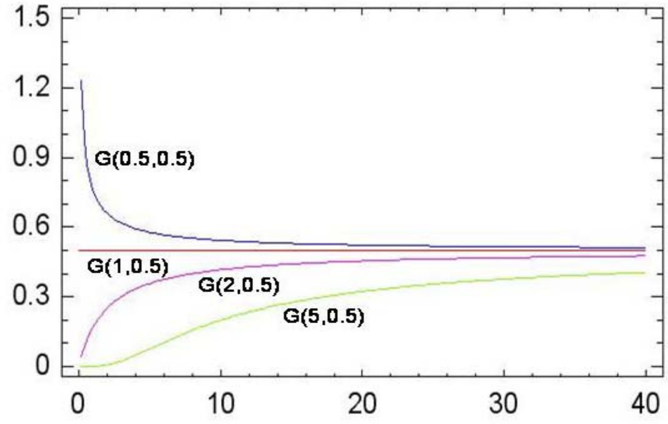


**Figura 5. Funciones de riesgo de distintas distribuciones Gamma**

## *2.3. La distribución de Weibull*

Un inconveniente de la distribución exponencial es que no sirve como modelo para el tiempo de vida en los que la razón de fallo no es una función constante, sino que la probabilidad condicional de fallo instantáneo varía con el tiempo. Cuando consideramos como generalización de una función constante las funciones potenciales, pasamos de la distribución exponencial a la distribución de Weibull.

### 2.3.1. Definición

La distribución de Weibull ha sido considerada en una gran variedad de situaciones prácticas. Mientras que la aplicabilidad de la distribución exponencial es limitada debido a la suposición de razón de fallo constante, la familia de distribuciones de Weibull incluyen razones de fallo crecientes y decrecientes. Como muchos fallos que encontramos en la práctica presentan una tendencia creciente, debido a envejecimiento o desgaste, esta distribución es útil para describir los patrones de este tipo de fallo.

En esta sección vamos a definir la distribución de Weibull a partir de su función razón de fallo, h(t). Sea T una v.a. tiempo de vida tal que la correspondiente razón de fallo viene dada por

$$h(t)= \lambda\beta(\lambda t)^{\beta-1},$$

con $\beta>0$, $\lambda>0$, $t>0$; $\beta$ es el parámetro de forma y $\lambda$ el parámetro de escala. Notamos a la distribución como $W(\lambda,\beta)$. Cuando $\beta=2$, es conocida como distribución de Rayleigh.

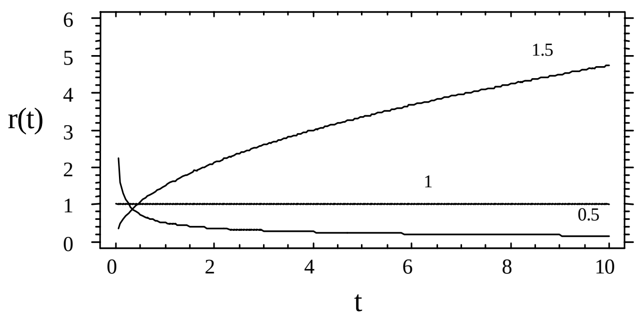


**Figura 6. Funciones de riesgo de distribuciones W(1, 1.5),W(1, 1) y W(1, 0.5)**

### 2.3.2. Propiedades de la distribución de Weibull

*Propiedad 1*. Si T1, T2, ..., Tn son n vv.aa. independientes e idénticamente distribuidas con distribución común Weibull, entonces min{T1, T2, ..., Tn} es también Weibull. La función de supervivencia es

$$S_{T_{(1)}}(t) = \left(S(t)\right)^n = \exp\left\{-n(\lambda\ t)^\beta\right\} \quad ,$$

y los parámetros pueden identificarse a partir de aquí fácilmente.

*Propiedad 2.* Si X tiene distribución exp(λ), entonces $T=X^p$ tiene distribución Weibull, W( $\lambda^p$ ,1/p).
De donde se deduce que la distribución exponencial es un caso particular de la distribución Weibull para β=1.

*Propiedad 3.* La función de riesgo es creciente para β>1 y h(t)→∞ cuando t→∞. Es decreciente para β<1 y tiende asintóticamente a 0 cuando t→∞, ver figura 6.

*Propiedad 4.* Si (λt)β<<1, una buena aproximación de F(t) es (λt)β.

*Propiedad 5.* Si β→∞, se tiene un tiempo de vida constante pues F(t)→0.

*Propiedad 6.* El i-ésimo momento de la distribución es igual a

$$E\left[T^i\right] = \Gamma\left(1+i/\beta\right)\lambda^{-i},$$

siendo $\Gamma(\cdot)$ la función Gamma.

## *2.4. Distribuciones de valores extremos*

### 2.4.1. Definición

Esta distribución describe adecuadamente ciertos fenómenos extremos tales como temperaturas mínimas, precipitaciones caídas en épocas de sequía, fuerza eléctrica de materiales, y ciertos tipos de datos de vida, por ejemplo mortalidad humana a edades avanzadas. Es, por otra parte, importante por su estrecha relación con la distribución Weibull, y en función de esto, los datos Weibull suelen ser analizados convenientemente en términos de la distribución de valores extremos más simple. Por todas estas razones, esta distribución juega un papel fundamental en análisis de supervivencia y surge de manera natural relacionada con estructuras en serie.

Sea $T_1, T_2,..., T_n$ variables aleatorias independientes e idénticamente distribuidas, (no necesariamente representando longitudes de vida) todas con distribución F, continua. Definimos las variables mínimo y máximo

$$T_{(1)} = \mathit{min}\ \{T_1, T_2,..., T_n\} = U_n$$
$$T_{(n)} = \mathit{max}\ \{T_1, T_2,..., T_n\} = V_n$$

es decir, los valores extremos.

Las funciones de distribución de $U_n$ y $V_n$ pueden expresarse fácilmente en términos de F, de la siguiente forma

$$F_{U_n}(t) = 1 - (1 - F(t))^n$$

y

$$F_{V_n}(t) = F(t)^n .$$

A pesar de la sencillez de estas expresiones, estas fórmulas resultan difíciles de manipular en la práctica, si por ejemplo, F representa una distribución normal, nos vemos obligados a trabajar con potencias de la función de distribución correspondiente, lo que resulta incómodo.

Sin embargo, en muchas aplicaciones prácticas, n es muy grande, por lo que podemos explorar técnicas asintóticas que, bajo condiciones generales para F pueden conducir a sencillas representaciones de las distribuciones de $U_n$ y $V_n$. En particular, podemos considerar la siguiente transformación $Y_n = n\,F(U_n)$ y obtener la distribución correspondiente, es decir, para $y > 0$

$$P\left(Y_n \le y\right) = P\left(F\left(U_n\right) \le \frac{y}{n}\right) = P\left(U_n \le F^{-1}\left(\frac{y}{n}\right)\right) =$$

$$= F_{U_n}\left[F^{-1}\left(\frac{y}{n}\right)\right] = 1 - \left[1 - F\left(F^{-1}\left(\frac{y}{n}\right)\right)\right]^n = 1 - \left(1 - \frac{y}{n}\right)^n .$$

Si tomamos ahora límites cuando $n \to \infty$,

$$P(Y_n \le y) = 1 - \left(1 - \frac{y}{n}\right)^n \to 1 - e^{-y} ,$$

para $y > 0$. Esto significa que $Y_n$ converge en distribución (en ley) a una v.a. exponencial de parámetro 1, y por lo tanto que

$$U_n \xrightarrow{L} F^{-1}\left(\frac{Y}{n}\right),$$

con $Y \to \exp(1)$.

De manera similar, podemos definir $W_n = n(1 - F(V_n))$. Usando un argumento análogo al anterior, podemos comprobar, para $w > 0$

$$P\left(W_n \le w\right) = 1 - \left(1 - \frac{w}{n}\right)^n ,$$

con lo que, si tomamos límites en infinito,

$$W_n \xrightarrow{L} W ,$$

siendo $W \to \exp(1)$, por lo tanto, si deshacemos los cambios

$$V_n \xrightarrow{L} F^{-1}\left(1 - \frac{W}{n}\right).$$

Según todo el razonamiento anterior, sería lógico esperar que la distribución límite de $U_n$ y $V_n$ dependerá de F. Sin embargo, resulta que hay únicamente tres posibles tipos de distribución límite para el extremo mínimo $U_n$ y sólo tres posibles tipos de distribuciones límite para el extremo máximo $V_n$. Si bien todas son equivalentes sin más que establecer la transformación adecuada.

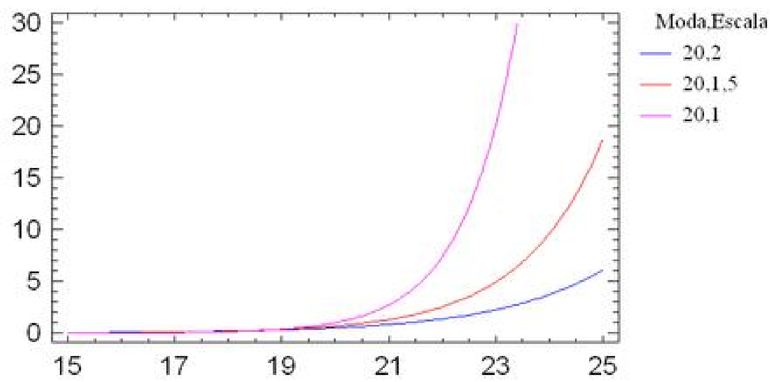


Figura 10. Razones de fallo de diferentes distribuciones de valores extremos.

### 2.4.2. La distribución de Gumbel

Si la función de densidad f, de la variable T, se aproxima a 0 exponencialmente cuando $t \to \infty$, la distribución límite del mínimo, $U_n = T_{(1)}$, es de la forma

$$F_{T_{(1)}}(t) = 1 - \exp\left(-e^{(t-\vartheta)/\alpha}\right), \quad -\infty < t < \infty$$

con $\alpha > 0$, $\vartheta$ constantes. El parámetro $\alpha$ es la moda de la distribución y $\vartheta$ el parámetro de escala. Esta distribución fue denominada por Gumbel la distribución asintótica tipo I del mínimo (extremo menor).

La distribución estándar corresponde a $\alpha = 1$ y $\vartheta = 0$, en este caso la razón de fallo tiene la forma

$$h_{T_{(1)}}(t) = e^{t}, \ -\infty < t < \infty.$$

Como $T_{(1)}$ puede tomar valores negativos, no resulta, esta distribución, adecuada como modelo de tiempo de vida. En este caso consideramos la distribución truncada en 0, es decir

$$F^{0}_{T_{(1)}}(t) = 1 - \exp\left[e^{-\frac{\vartheta}{\alpha}}\left(1 - e^{\frac{\vartheta}{\alpha}}\right)\right], \text{ para } t > 0.$$

Si la densidad f se aproxima a 0 exponencialmente cuando $t \to \infty$, la distribución límite del máximo, $V_n = T_{(n)}$, es de la forma

$$F_{T_{(n)}}(t) = \exp\left(-e^{-(t-\vartheta)/\alpha}\right), \quad -\infty < t < \infty$$

con $\alpha > 0$, $\vartheta$ constantes. Esta distribución fue denominada por Gumbel la distribución asintótica tipo I del máximo (extremo mayor).

### 2.4.3. La distribución Weibull

Weibull (1951) comprobó que la resistencia final de un determinado material podía ser descrita mediante una distribución de valores extremos (Gumbel) con $\vartheta = 108$ Kg/cm$^2$, y $\alpha$ = 9.27 Kg/cm$^2$. La distribución de Weibull puede considerarse como la distribución asintótica tipo III del mínimo, con la siguiente parametrización

$$F_{T_{(1)}}(t) = 1 - e^{-((t-\vartheta)/\eta)^{\beta}}, \quad t \geq 0;$$

con $\vartheta > 0$, $\eta > 0$, y $\beta > 0$.

Además, si tenemos en cuenta la definición de distribución Weibull que dimos en la sección 2.3, esto es

$$T \to W(\lambda, \beta) \Rightarrow S(t) = e^{-(\lambda t)^{\beta}}, \; t > 0.$$

Podemos considerar el siguiente cambio de variable Y = ln T, con lo cual

$$\overline{F}_Y(y) = e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}}, \quad -\infty < y < \infty,$$

con $\mu = -\ln \lambda$ y $\sigma = 1/\beta$. Es decir, la distribución de Weibull y la distribución de valores extremos guardan una relación paralela a la que guardan la distribución lognormal y normal y también la distribución log–logística y logística.

## *2.5. Modelos asociados a la distribución normal*

### 2.5.1. La distribución Normal

Una v.a. T se dice que tiene distribución N($\mu$, $\sigma$) si la función de densidad es

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu)^2/2\sigma^2}.$$

Si llamamos $\Phi(t)$ a la función de distribución de la N(0,1), entonces se puede comprobar que la función de distribución de la N($\mu$, $\sigma$) es

$$F(t) = P\{T \leq t\} = \Phi\left(\frac{t-\mu}{\sigma}\right)$$

y, por tanto, la función de riesgo viene dada por

$$h(t) = \frac{1}{\sigma} \cdot \frac{\phi((t-\mu)/\sigma)}{1 - \Phi((t-\mu)/\sigma)}.$$

Si $r_{\Phi}(t)$ representa la función de riesgo de la normal estándar, se tiene la siguiente igualdad

$$r(t) = \frac{1}{\sigma} \cdot r_{\Phi}\left(\frac{t-\mu}{\sigma}\right).$$

El siguiente gráfico muestra la función de riesgo de la N(0,1), es una función creciente para todo t y se aproxima asintóticamente a h(t)=t.

.

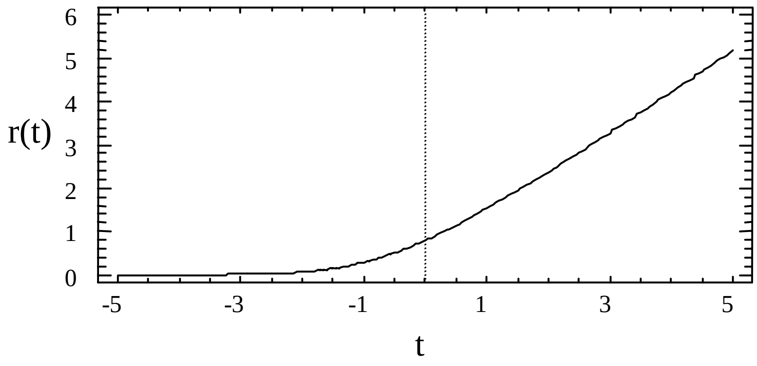


**Figura 7. Razón de fallo de la distribución N(0,1).**

Cuando una v.a. tiene distribución normal pero con una cota superior y/o inferior para los valores de la variable, la distribución resultante se llama *distribución normal truncada*. Cuando sólo hay una cota inferior (superior), la distribución se dice truncada a la izquierda (derecha), si existen las dos cotas se dice doblemente truncada.

La distribución normal truncada en 0 es a menudo usada como distribución de tiempo de vida.

### 2.5.2. La distribución lognormal

Una v.a. T tiene distribución lognormal si X=lnT tiene distribución normal. Nótese que X toma cualquier valor real mientras que T siempre es mayor que 0. La función de densidad de una distribución lognormal de parámetros $\mu$ y $\sigma$ es

$$f(t) = \frac{1}{\sqrt{2\pi}\, t\sigma} e^{-(\ln t - \mu)^2 / 2\sigma^2} ,$$

para t>0. La figura 8 muestra la razón de fallo de la distribución lognormal para distintos valores de los parámetros.

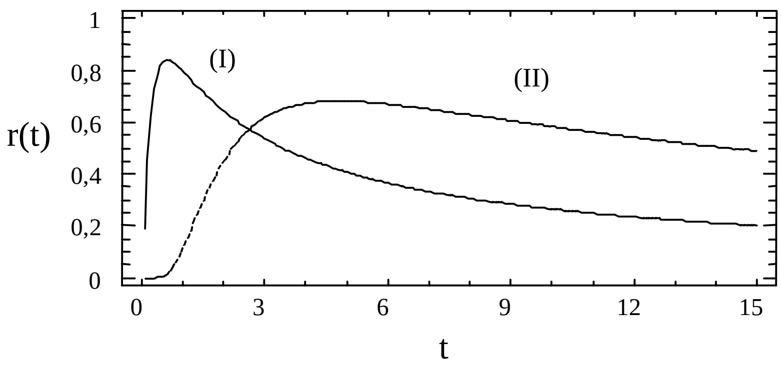


**Figura 8. Razones de fallo de distribuciones lognormales con: (I) $\mu$=0 y $\sigma$=1; y (II) $\mu$=1 y $\sigma$=0.5.**

## *2.6. Modelos asociados a la distribución logística*

Veamos cómo se llega a esta distribución mediante un ejemplo. Supongamos que se estudia la concentración de una droga para que sea mortal. Se administra la droga a un paciente en diferentes dosis, d. Sea p la probabilidad de que el sujeto sometido a tratamiento muera, y supongamos que esta probabilidad depende de la dosis que recibe. La relación más simple que podemos establecer es una ecuación lineal:

$$p = \alpha + \beta d,$$

pero este no es un buen modelo, ya que mientras que p es una probabilidad y sólo toma valores en (0,1), d puede tomar cualquier valor real. Para resolver este problema, reemplazamos p por una función, g(p), que lleva el intervalo (0,1) en $(-\infty, +\infty)$, y tiene la siguiente forma

$$g(p) = \ln\frac{p}{1-p} \ .$$

Así pues, el modelo logístico viene dado por

$$\ln\frac{p}{1-p} = \alpha + \beta d \ ,$$

y la distribución logística (estándar) se define a partir de su función de distribución de la siguiente forma

$$F(t) = \frac{1}{1+e^{-t}} \ .$$

En general, diremos que T sigue distribución logística de parámetros $-\infty < \mu < \infty$ y $\sigma > 0$, si la función de supervivencia viene dada por

$$S(t) = \frac{1}{1+e^{\left(\frac{t-\mu}{\sigma}\right)}} \ , \quad -\infty < t < \infty$$

Sea T una v.a. con distribución logística estándar y consideremos la siguiente transformación $Y = e^T$. Y se dice entonces que tiene distribución log–logística. Es fácil obtener las características de esta distribución, por ejemplo, a partir de la última expresión, la función de supervivencia es

$$S_Y(y) = \frac{1}{1+(\lambda t)^p} \ , \ t > 0,$$

donde $\lambda = e^{-\mu}$ y $p = 1/\sigma$.

La correspondiente función de riesgo es

$$h_Y(y) = \frac{\lambda p (\lambda t)^{p-1}}{1+(\lambda t)^p} \ ,$$

y, entre las propiedades de esta función, destaca que ofrece distintas formas en función del parámetro p, lo cual es interesante desde el punto de vista de la modelización. Es decir, si $p < 1$, es una función decreciente desde infinito. Si $p = 1$, es decreciente desde el valor de $\lambda$. Y, por último, si $p > 1$, presenta una forma parecida a la razón de fallo de la distribución lognormal, salvo en la cola derecha, es decir, es una función que parte de

0 y crece hasta alcanzar un máximo en el punto $y = \frac{(p-1)^{1/p}}{\lambda}$, y a partir de ese momento decrece a 0, con lo que esta distribución puede considerarse una candidata a tener en cuenta como modelo en el caso en el que la distribución lognormal no se ajuste bien cuando t crece a infinito.

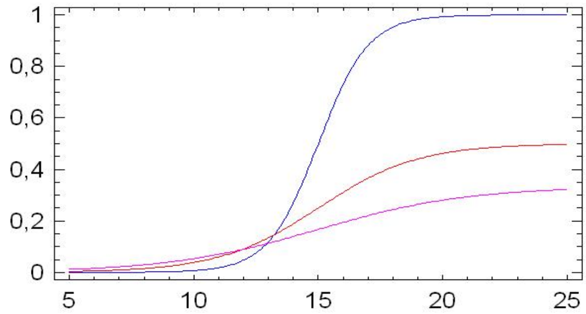


Figura 9. Razones de fallo correspondiente a distribuciones logísticas $\mu$=15 y $\sigma$=1,2,3 (desde arriba).

## *2.7. La distribución tipo fase*

La distribución tipo fase ha sido utilizada en distintos campos como la supervivencia (estudio del cáncer, sida, etc), fiabilidad (estudio de sistemas reparables, etc) y en teoría de colas. Para introducir esta distribución veamos el siguiente ejemplo.

Supongamos una enfermedad que puede atravesar distintas fases, comunicadas todas ellas entre sí. Comenzando el estudio en alguna de estas fases con una cierta probabilidad, tras un tiempo de vida exponencial cambia de fase a otra distinta con una cierta probabilidad. Desde esta nueva fase, tras un tiempo de vida exponencial (no igual necesariamente que el anterior) cambia de fase a otra con una cierta probabilidad y así sucesivamente. Supongamos que en un instante de tiempo la enfermedad alcanza una fase en la que se para el tiempo de vida (curación, muerte, recaída determinada, etc.). Se define una distribución tipo fase como el tiempo hasta alcanzar esta última fase.

Si el tiempo en la fase i se distribuye como una exponencial de parámetro $\lambda_i$, entonces este parámetro es el riesgo de salida de esta fase i. El riesgo de cambio desde esta fase i a la fase j viene dado por $\lambda_{ij} = p_{ij}\, \lambda_i$ , siendo $p_{ij}$ la probabilidad de saltar al estado j desde el estado i cuando cambia de fase. Entonces definimos la matriz A como aquella matriz que tiene como entrada (i, j) el elemento $\lambda_{ij}$ si $i \neq j$ y $-\lambda_i$ si i=j para i, j, todas las fases transitorias en estudio. Llamamos $A^0$ a un vector columna cuya entrada k viene dada por el riesgo de pasar desde la fase k a la fase absorbente final. Sea el vector $\alpha$ el que tiene como elemento i la probabilidad de que inicialmente se esté en la fase transitoria i, para las fases transitorias. Entonces la función de densidad de una distribución tipo fase de parámetros $\alpha$, A , denotamos por PH($\alpha$, A) viene dada por

$$f(t) = \alpha \exp\{At\} A^0 \quad , t > 0 .$$

La función de supervivencia es

$$S(t) = \alpha \exp\{At\} e \quad , t > 0 \, ,$$

siendo e un vector columna de unos de dimensión apropiada.

El tiempo medio de vida de una distribución tipo fase es $E[T] = -\alpha A^{-1} e$. Finalmente, la función de riesgo es igual a

$$h(t) = \frac{\alpha \exp\{At\} A^0}{\alpha \exp\{At\} e} \quad , t > 0 \, .$$