INTRODUCCIÓN AL ANÁLISIS DE SOBREVIVENCIA

CLAUDIA LORENA CERTUCHE VÉLEZ
CARLOS ANDRES MEDINA GAVIRIA
SANDRA PIEDAD PAME NORIEGA

UNIVERSIDAD DEL CAUCA
FACULTAD DE CIENCIAS NATURALES, EXACTAS
Y DE LA EDUCACIÓN
DEPARTAMENTO DE MATEMÁTICAS

POPAYÁN
2004

INTRODUCCIÓN AL ANÁLISIS DE SOBREVIVENCIA

Seminario de grado presentado como requisito parcial para optar al título de:

Licenciadas en Educación con Especialidad en Matemáticas a

CLAUDIA LORENA CERTUCHE VÉLEZ
SANDRA PIEDAD PAME NORIEGA

Y Matemático a

CARLOS ANDRÉS MEDINA GAVIRIA

Director:
Dr. LUIS EDUARDO MONTOYA

UNIVERSIDAD DEL CAUCA
FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA EDUCACIÓN
DEPARTAMENTO DE MATEMÁTICAS
POPAYÁN
2004

Dedicamos este trabajo a quienes nos
han apoyado incondicionalmente:
nuestros padres, familiares y amigos.

Nota de aceptación:

Director:

Dr. **Luis Eduardo Montoya**

Comité evaluador:

Mg. **Edwin Rengifo Canizales**

Dr. **Freddy Angel Amaya R.**

Fecha de sustentación: 13 de septiembre de 2004.

Agradecimientos

Los autores expresan sus más sinceros agradecimientos a:

Luis Eduardo Montoya, decano de la Facultad de Ciencias Naturales, Exactas y de la Educación y director de este trabajo, por sus valiosos aportes y su constante esfuerzo.

Edwin Rengifo, profesor del departamento de Matemáticas y miembro del comité evaluador, por su colaboración y sugerencias.

Freddy Amaya, profesor del departamento de Matemáticas y miembro del comité evaluador, por su colaboración y sugerencias.

Gerardo Loaiza, profesor del departamento de Matemáticas, por sus asesorías para la realización de este trabajo.

A nuestras familias por su apoyo, paciencia, amor y colaboración.

A la Universidad del Cauca, nuestro segundo hogar, que nos dió la oportunidad de crecer como ciudadanos y profesionales.

Contenido

1. Conceptos básicos **12**

1.1. Introducción 12

1.2. Función de sobrevivencia 13

1.3. Otras funciones de interés. 16

1.3.1. Función de riesgo 17

1.3.2. Esperanza de vida 24

1.3.3. Varianza 29

2. Modelos paramétricos **31**

2.1. Introducción 31

2.2. Modelo Gamma generalizado 32

2.2.1. Modelo Gamma 35

2.2.2. Modelo Exponencial 41

2.2.3. Modelo Weibull 46

2.3. Modelo Lognormal 55

2.4. Modelo Gompertz 63

2.5. Modelo Pareto 71

Conclusiones **78**

Bibliografía **79**

Anexos **80**

A. **81**

A.1. Aproximación gráfica para estimar parámetros 81

A.1.1. Modelo Weibull 81

A.2. Teoría básica para la construcción de tablas de mortalidad. 85

B. **89**

B.1. Algunas tablas de interés 89

B.1.1 Percentiles modelo Gamma con parámetros $\lambda = 0.2$ y $\beta = 3$ 89

B.1.2 Funciones de densidad de probabilidad, de sobrevivencia, de riesgo y la esperanza de vida para algunos modelos parámetricos. 90

B.1.3 Valores de la función de distribución acumulada normal estándar. . . 91

Índice de figuras

1.1. función de sobrevivencia 15

1.2. función de riesgo constante 18

1.3. función de riesgo creciente 18

1.4. función de riesgo decreciente 19

1.5. función de riesgo lognormal 19

2.1. funciones de sobrevivencia y de riesgo del modelo Gamma con parámetros $\lambda = 0.2$ y $\beta = 3$. 40

2.2. funciones de sobrevivencia y de riesgo del modelo Exponencial con parámetro $\lambda = 0.04$. 45

2.3. Estimación gráfica de los parámetros para el modelo Weibull. 53

2.4. funciones de sobrevivencia y de riesgo del modelo Weibull con parámetros $\lambda = 0.015$ y $\alpha = 1.75$. 54

2.5. función de sobrevivencia Lognormal con parámetros $\mu = 3.42$ y $\sigma^2 = 3.75$. . . 62

2.6. función de riesgo Lognormal con parámetros $\mu = 3.42$ y $\sigma^2 = 3.75$. 62

2.7. función de sobrevivencia y función de riesgo del modelo Gompertz con parámetros $m = 84$ y $\alpha = 0.1572$. 70

2.8. funciones de sobrevivencia y de riesgo del modelo Pareto con parámetros $\alpha = 4$ y $\theta = 5$ 76

A.1. Papel Weibull. 84

Resumen

En este documento se presenta el informe del seminario de grado "Introducción al análisis de sobrevivencia", realizado dentro del grupo de estudio y desarrollo investigativo en matemática aplicada, en la línea de estadistica y probabilidad del departamento de Matemáticas.

En la primera parte se presenta una teoría básica para el análisis de sobrevivencia; posteriormente algunos modelos de sobrevivencia parámetricos con sus respectivos ejemplos de aplicación.

Introducción

En todo tiempo, ciertas características de los seres vivos, de las máquinas y equipos, han estado sujetas a una multiplicidad de riesgos que atentan contra la sobrevivencia y el buen funcionamiento de los mismos. Debido a esto, **el análisis de sobrevivencia** surge como una metodología basada en técnicas actuariales, que utilizando elementos teóricos y prácticos en campos como la estadística, probabilidad y finanzas, aporta valiosa información para tomar decisiones estratégicas y proveer soluciones prácticas a problemas que involucran las posibles consecuencias de eventos futuros inciertos, como por ejemplo, el tiempo que se espera transcurra entre el nacimiento y el deceso de un ser vivo, o el tiempo de falla de una máquina, entre otros.

Uno de los objetivos del Análisis de Sobrevivencia es encontrar una función que describa la probabilidad de que **el elemento en estudio**, que puede ser, una persona, un animal, una máquina, entre otros, sobreviva un determinado tiempo y que dicha función tenga además, un buen grado de confiabilidad; esta función se denomina **función de sobrevivencia**. Sin embargo, no es fácil definir una única función para este propósito, porque hay que tener en cuenta que los riesgos varían en las distintas poblaciones, especies (de animales), calidades (de maquinarias), etc. Es por eso que se debe escoger la distribución de probabilidad adecuada para la variable de interés y que se ajuste muy bien a los parámetros del estudio, para que las inferencias realizadas sean confiables.

Se considera importante el estudio de algunas distribuciones de probabilidad, que representan tiempos de "vida"o de sobrevivencia de un elemento en estudio; y son útiles, ya que a partir de sus respectivas funciones de **densidad de probabilidad**, se pueden determinar funciones claves como: **la función de sobrevivencia**, **función de riesgo** y **la esperanza de vida**, para cada una de ellas. Además, las distribuciones de probabilidad de los tiempos de sobrevivencia son usualmente descritas o caracterizadas por estas cuatro funciones, que están estrechamente relacionadas (ver capítulo 1). Al conjunto de estas funciones para cada distribución de probabilidad se le llama **Modelo paramétrico de sobrevivencia**. Así, un problema básico del análisis de sobrevivencia, es estimar a partir de un conjunto de obsevaciones las funciones mencionadas y hacer inferencias acerca del modelo de sobrevivencia en la población.

Cabe resaltar las diversas aplicaciones de los modelos de sobrevivencia: en Procesos de control de calidad, como ocurre en el campo industrial donde se estudia el tiempo de falla de una unidad o alguna componente de la misma; en Economía, el tiempo trancurrido entre la oferta y la aceptación de un empleo ofrecido a una persona desempleada; en Demografía, el tiempo en el que un individuo llega a contraer matrimonio, o el tiempo que dura una relación de pareja; en el campo de la Medicina, los modelos de sobrevivencia se emplean para el estudio del tiempo de duración de enfermedades crónicas o los tratamientos de las mismas, el tiempo transcurrido hasta que ocurra la muerte de un paciente, el tiempo hasta la curación o aparición de una enfermedad, entre otros. En este último campo se han encontrado muchas más aplicaciones, entre ellas se destacan estudios en personas con enfermedades: cardiovasculares; de transmisión sexual; cáncer cervical, de mama, de laringe; leucemia aguda y transplantes de médula espinal, de riñón, etc. (ver capítulo 2).

Dichos modelos, aunque constituyen herramientas muy valiosas para el conocimiento e interpretación de los fenómenos, son también modelos transitorios, sujetos a verificación y perfeccionamiento, ya que cuando se tiene más información sobre el comportamiento del evento, esto puede generar un cambio en el modelo probabilístico.

En la mayor parte de la bibliografía referente a este tema se presenta la teoría en forma sintetizada, sin hacer claridad en muchos aspectos, entre ellos las condiciones bajo las cuales se cumplen los resultados teóricos descritos en el presente texto. Por tal motivo y dada la necesidad de modelar problemas relacionados con tiempos de "vida", nuestro aporte consiste en sistematizar, estudiar y presentar en detalle la interesante teoría en la que se fundamenta el análisis de sobrevivencia.

Capítulo 1

Conceptos básicos sobre análisis de sobrevivencia

1.1. Introducción

El análisis de Sobrevivencia es una metodología estadística que permite estudiar eventos, tales como muerte, incidencia de una enfermedad (o recaida debida a esta), desaparición de enfermedades (recuperación), desajuste de un mecanismo, o mal funcionamiento de este, o cualquier otra experiencia semejante; donde la variable respuesta sea el tiempo en el que pueden presentarse características como el deterioro, falla (en su funcionamiento), o sobrevivencia del elemento en estudio; que podría ser una persona, un animal, un sistema, una maquinaria, etc. Así, en el estudio de un determinado evento con las cararterísticas anteriores, la variable **Tiempo** representará el **Tiempo de Sobrevivencia**, sin importar cual sea el objeto en cuestión. Además, dicho tiempo puede ser medido en años, meses o días.

Aunque en un mismo análisis puede considerarse más de un evento, asumimos que sólo uno es considerado de interés, ya que el objetivo de este texto es dar la introducción al análisis de sobrevivencia, por eso estudiaremos una sola variable (análisis univariado). El caso en el que se presentan varios eventos, está emarcado en el análisis multivariado,que se dejará para posteriores estudios.

Puede presentarse el caso de que el elemento que se esté analizando abandone el estudio antes de que ocurra el evento de interés, registrándose sólo información parcial, llamada **Censura**. Hay generalmente tres razones relacionadas al elemento de estudio, para que ocurra la Censura:

1. Cuando no se experimenta el evento dentro del estudio.

2. Cuando se pierde el seguimiento durante el periodo de estudio.

3. Cuando se retira del estudio debido a factores inevitables.

Los casos antes mencionados son muy comunes en análisis de sobrevivencia y pueden presentarse en situaciones particulares. Por ejemplo:

- Para el primer caso, supongamos que el evento de interés es la recuperación del elemento en estudio. Si dicho elemento es un ser vivo, la censura puede presentarse cuando este no experimenta mejoría o muere; si se trata de una maquinaria, la censura puede presentarse si esta deja de funcionar debido a una descarga eléctrica o cualquier otra falla.

- Para el segundo caso, puede ser cuando el elemento en estudio desaparece y no se vuelve a tener contacto con el mismo , así que el seguimiento para este termina irremediablemente, sin saber si el evento de interés se presentó.

- El tercero, puede hacer referencia a factores como la muerte (si la muerte no es el evento de interés), o alguna otra razón que puede ser una reacción adversa a los medicamentos en el caso de los seres vivos; o si se trata de alguna maquinaria, cuando los repuestos no se pueden conseguir, entre otros.

Es importante resaltar que la censura debe ser incorporada en el Análisis de Sobrevivencia a pesar de que son datos parciales, ya que estos aportan valiosa información para el estudio.

En este capítulo, se estudiarán la función de sobrevivencia y otras funciones de interés, relacionadas con la misma, que dan un aporte fundamental y valiosa información al Análisis de sobrevivencia.

1.2. Función de sobrevivencia

Describiremos en esta sección, la función que según nuestro criterio, es de donde surge toda la teoría del Análisis de sobrevivencia y que nos aporta información sobre la probabilidad de "vida" de los elementos que estemos estudiando.

Definición 1.2.1. *Sea Ω el espacio muestral conformado por elementos de una población en estudio*[1]*; y $T : \Omega \to \mathbb{R}$ la variable aleatoria definida sobre el espacio de probabilidad $(\Omega, \mathcal{F}, P)$, que representa el tiempo de "vida" del elemento en estudio.*

[1] Dichos elementos pueden ser: seres humanos, animales, microorganismos, componentes de máquinas, sistemas o maquinarias, entre otros.

Definición 1.2.2. *Sea T la variable aleatoria continua que representa el tiempo de "vida"del elemento en estudio y F_T su correspondiente Función de Distribución de Probabilidad Acumulada. Consideremos la función S definida así:*

$$\begin{aligned} S : \mathbb{R}^+ \cup \{0\} &\to [0,1] \\ t &\to S(t) = P(T > t) = 1 - F_T(t) \end{aligned} \tag{1.1}$$

Así, la función S representa la probabilidad de que el elemento en estudio sobreviva a un tiempo t, llamado ***tiempo de sobrevivencia****.*

Esta función satisface las siguientes propiedades:

1. *S es decreciente*
   Como F_T es creciente, entonces $S = 1 - F_T$ es decreciente.

2. $S(0) = 1$
   Observe que $F_T(t) = 0$ si $t \leq 0$, entonces:
   $S(0) = P(T > 0) = 1 - F_T(0) = 1 - 0 = 1$

   Así que, al comienzo del estudio (que puede ser cuando nace el individuo o cuando ingresa al mismo), el 100 % de los elementos de la población están "vivos"(sobreviven).

3. Sea *w el tiempo mínimo al cual el elemento en estudio no puede sobrevivir (tiempo final del modelo).*Entonces $S(w) = 0$.

   - Consideremos a $w \in \mathbb{R}^+$:
     Como $F_T(w) = P(T \leq w)$ y como w es el tiempo mínimo al cual el elemento en estudio no puede sobrevivir, entonces:
     $P(T \leq w) = 1$, luego $S(w) = 1 - F_T(w) = 1 - 1 = 0$
   - Consideremos ahora el caso donde $w \to \infty$, entonces:

     $$\lim_{w\to\infty} S(w) = \lim_{w\to\infty} P(T > w) = \lim_{w\to\infty} [1 - F_T(w)] = 1 - \lim_{w\to\infty} F_T(w) = 1 - 1 = 0$$

     Así que, a medida que el tiempo de vida aumenta, la probabilidad de sobrevivencia del elemento en estudio tiende a cero.

Observación 1.2.1. *Como $F_T(t) = P(T \leq t)$ es la función de distribución de T, que representa la probabilidad de vivir a lo más t unidades de tiempo, entonces:*

$$F_T(t) = P(T \leq t) = (1 - S(t))\, I_{[0,+\infty)}(t) \ , \tag{1.2}$$

donde $I_{[0,+\infty)}(\cdot)$ es la función indicadora (característica) del conjunto $[0, +\infty)$ definida por:

$$I_{[0,+\infty)}(t) = \begin{cases} 1 & \text{si } t \in [0, +\infty) \\ \\ 0 & \text{en caso contrario.} \end{cases}$$

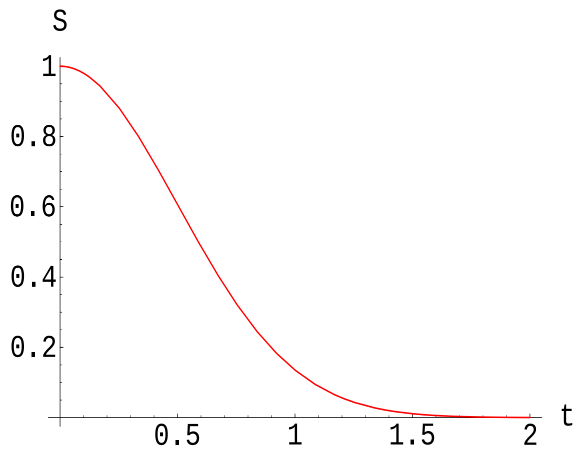

Figura 1.1: función de sobrevivencia

En la fig.(1.1) se pueden apreciar las características de la función de sobrevivencia anteriormente mencionadas.

- *Si T es continua y F_T es diferenciable, entonces podemos obtener la función de densidad de probabilidad de T, denotada f_T , de la siguiente forma:*

$$f_T(t) = \begin{cases} \dfrac{d}{dt}F_T(t) = -\dfrac{d}{dt}S(t) & \text{si } t > 0 \\ 0 & \text{si } t \leq 0. \end{cases} \tag{1.3}$$

  Por lo tanto, en este caso la función de sobrevivencia también se puede expresar como:

$$S(t) = \int_t^{\infty} f_T(x)dx \tag{1.4}$$

- *Si T es discreta, con valores $t_0, t_1, t_2, ...,$ con $t_j < t_{j+1}$ para $j = 0, 1, 2, ...$ y $t_0 = 0$, entonces:*

$$S(t) = \sum_{t_j > t} P(T = t_j) \tag{1.5}$$

Ejemplo 1.2.1. *Supongamos que la variable aleatoria T,la cual representa el tiempo de sobrevivencia de una persona, tiene un comportamiento uniforme en el intervalo $[0, \alpha]$. Construyamos la función de sobrevivencia para T y hallemos lo siguiente:*

a) *La probabilidad de que un recién nacido sobreviva más de a años.*

b) La probabilidad de que un recién nacido muera más allá de los a y antes de los b años.

c) la probabilidad de que una persona de edad a sobreviva a los b años.

Solución:

Como $S(t) = P(T > t)$, entonces: $S(t) = 1 - \dfrac{t}{\alpha}$ con $0 < t < \alpha$, así

$$S(t) = \begin{cases} 1, & \text{si } t = 0; \\ 0, & \text{si } t \geq \alpha. \end{cases}$$

Se puede ver que esta función cumple las propiedades de la función de sobrevivencia, esto es:

$S(0) = 1$, $S(t)$ es decreciente y si tomamos $w = \alpha$ entonces $S(w) = 0$.

a) $P(T > a) = S(a) = 1 - \dfrac{a}{\alpha} = \dfrac{\alpha - a}{\alpha}, \qquad \text{si } 0 \leq a \leq \alpha$

b)
$$\begin{aligned} P(a < T \leq b) &= P(T > a) - P(T > b) \\ &= S(a) - S(b) = 1 - \frac{a}{\alpha} - \left[1 - \frac{b}{\alpha}\right] \\ &= \frac{b}{\alpha} - \frac{a}{\alpha} = \frac{b-a}{\alpha}, \qquad \text{si } 0 \leq a < b \leq \alpha \end{aligned}$$

c)
$$\begin{aligned} P(b < T | T > a) &= \frac{P[(T > a) \bigcap (T > b)]}{P(T > a)} \\ &= \frac{P(T > b)}{P(T > a)} = \frac{S(b)}{S(a)} = \frac{\alpha - b}{\alpha - a}, \qquad \text{si } 0 \leq a < b \leq \alpha \quad \blacksquare \end{aligned}$$

1.3. Otras funciones de interés.

A continuación definiremos algunas funciones importantes relacionadas con la función de sobrevivencia y que son de interés en el análisis de sobrevivencia.

Sabemos que $S(t)$ expresa la probabilidad de que el elemento en estudio sobreviva al menos un tiempo t y además, cuando t aumenta, $S(t)$ decrece, entonces la probabilidad de "vida"del elemento de estudio decrece a medida que el tiempo aumenta; por tal motivo, si dicho elemento ha sobrevivido hasta un tiempo t, nos interesa ponderar el decrecimiento de su probabilidad de vida en el intervalo $[t, t+\Delta t]$, ya que en cada instante existe el riesgo de que el elemento falle y este riesgo debe ser inversamente proporcional al decrecimiento de dicha probabilidad en este intervalo.

Consideremos el siguiente cociente:

$$\frac{P(t \leq T < t+\Delta t | T \geq t)}{\Delta t}$$

que representa la ponderación, en el intervalo $[t, t+\Delta t]$, de la probabilidad de sobrevivir Δt unidades de tiempo, dado que el elemento sobrevive a t. Más aún, si existe el límite de este cociente cuando Δt se hace muy pequeño, obtenemos la razón de cambio en el instante t de la probabilidad de sobrevivencia del elemento en estudio.

Bajo estos argumentos definiremos a continuación la función de riesgo.

1.3.1. Función de riesgo

Definición 1.3.1. *Si T es una v.a. continua, la función de riesgo evaluada en el punto t, denotada $h(t)$, se define como la probabilidad instantánea de falla del elemento en estudio, en un intervalo pequeño de tiempo $[t, t+\Delta t]$, dado que éste ha sobrevivido al tiempo t. Esto es:*

$$h(t) = \lim_{\Delta t \to 0^+} \frac{P(t \leq T < t+\Delta t | T \geq t)}{\Delta t}, \tag{1.6}$$

siempre que el límite exista.

Nota: Otra forma de interpretar la función de riesgo es como el potencial de falla en el instante t por unidad de tiempo, dado que sobrevive al tiempo t.

El cociente entre la probabilidad condicional y el lapso de tiempo pequeño Δt, expresa una razón de cambio por unidad de tiempo, que depende de las unidades de medida de este, así que, no necesariamente es un valor entre 0 y 1.

En algunos textos, la función de riesgo es conocida también como **la proporción de falla condicional**[2] en epidemiología, **la fuerza de Mortalidad**[3] en demografía, **la Función**

[2] "Survival Analysis"; David G. Kleinbaum; Cap. 1

[3] "Cálculo Actuarial:Contingencias de vida individual"; Jaime Abel Huertas; cap.2

de intensidad[4] en procesos estocásticos, **la inversa de la Razón de Mill**[5] en economía.

En contraste con la función de sobrevivencia, el gráfico de la función de riesgo $h(.)$, no siempre empieza en 1 y no siempre decrece tendiendo a 0, más aún, su comportamiento no se puede generalizar; sin embargo, de la expresión (1.6), tanto la probabilidad en el numerador, como el Δt en el denominador son no negativas, así que esta función es no negativa.

La fig. (1.2), muestra una función de Riesgo constante para un posible estudio de personas saludables, donde el potencial instantáneo de riesgo de estas personas durante el periodo de estudio, será constante hasta que se enferme, sin importar el tiempo escogido.

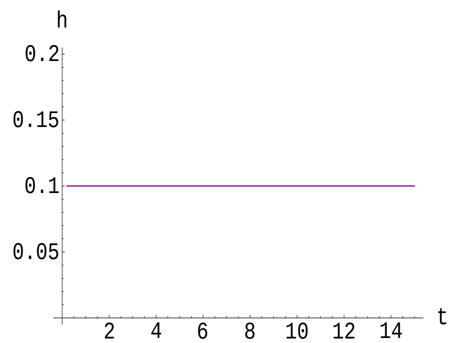

Figura 1.2: función de riesgo constante

La fig.(1.3) muestra una función de riesgo creciente y una función de este tipo se puede utilizar para modelar el riesgo que tienen pacientes con cierta enfermedad y que no respondan a un tratamiento, donde el evento de interés es la muerte. Como el tiempo de sobrevivencia crece para tal tipo de paciente y su pronóstico por consiguiente empeora, el potencial de riesgo de este también crece.

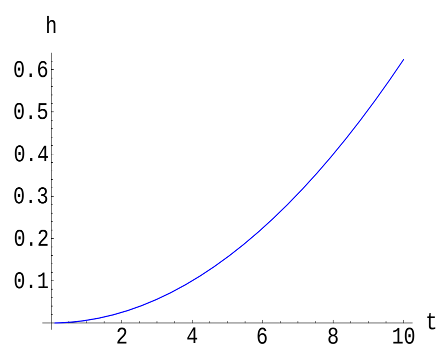

Figura 1.3: función de riesgo creciente

[4] $http://www.econ.au.dk/ec2bologna/papers/lb\_SCI.pdf$
[5] $http://www.economics.toronto.edu/siow/2801/canals\_stern.pdf$

La fig.(1.4) ilustra una función de riesgo decreciente; una función de este tipo puede utilizarse cuando el evento de interés es la muerte en personas que se están recuperando de una intervención quirúrgica, porque el potencial de riesgo para el paciente después de la cirugía usualmente decrece a medida que el tiempo aumenta.

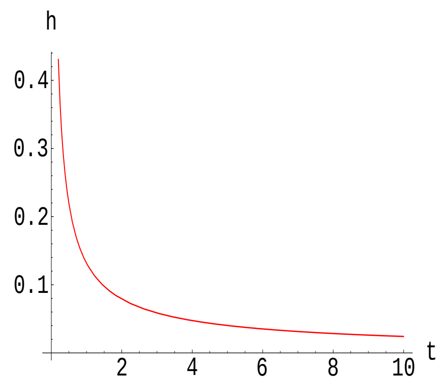

Figura 1.4: función de riesgo decreciente

La fig.(1.5) muestra una función que primero es creciente y después decreciente. Se puede modelar una función de este tipo en pacientes con enfermedades donde el potencial de riesgo para el enfermo crece rápidamente durante la enfermedad y decrece después de realizado el tratamiento, siempre que este tenga éxito.[6]

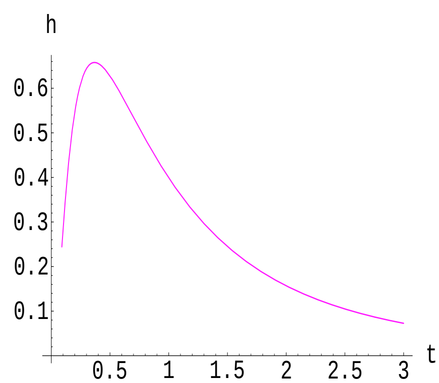

Figura 1.5: función de riesgo lognormal

Los modelos de sobrevivencia asociados a las funciones anteriores se presentarán y analizarán detalladamente en el Capítulo 2.

El siguiente resultado nos muestra la relación que existe entre la función de densidad de probabilidad y la función de sobrevivencia.

[6]Una revisión respecto al tema se puede encontrar en el libro: "Survival Analysis"por David G. Kleinbaum, Cap. 1

Teorema 1.3.1. *Si T es una variable aleatoria continua, en cada punto t, $t < w$, donde S es diferenciable se tiene que:*

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

Demostración: Sea t un punto donde S es diferenciable y $S(t) > 0$. Consideremos el siguiente límite y verifiquemos si existe:

$$\lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

Por propiedades de probabilidad esto es equivalente a:

$$\lim_{\Delta t \to 0} \frac{P(T \geq t \mid T \geq t) - P(T > t + \Delta t \mid T \geq t)}{\Delta t}$$

Por definición de probabilidad condicional se llega a:

$$\lim_{\Delta t \to 0} \frac{1 - \dfrac{P(T > t + \Delta t)}{P(T \geq t)}}{\Delta t}$$

Por definición (1.2.2) y el hecho de que T es una v.a. continua el límite anterior se puede expresar de la siguiente forma:

$$\lim_{\Delta t \to 0} \frac{1 - \dfrac{S(t + \Delta t)}{S(t)}}{\Delta t} = -\frac{1}{S(t)} \left[\lim_{\Delta t \to 0} \frac{S(t + \Delta t) - S(t)}{\Delta t} \right]$$

Como S es diferenciable en el punto t, entonces:

$$\lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{-S'(t)}{S(t)}$$

Luego, como el límite existe tenemos que:

$$h(t) = \frac{-S'(t)}{S(t)} \tag{1.7}$$

Y por (1.3), tenemos que:

$$h(t) = \frac{f_T(t)}{S(t)} \tag{1.8}$$

$\square$

Nota:
En algunos textos, la función de riesgo se presenta únicamente como:

$$h(t) = \frac{f_T(t)}{S(t)} \quad \text{con } S(t) \neq 0$$

Por otro lado, haciendo una similitud con el caso continuo, una forma natural para definir la función de riesgo en el caso discreto es:

Definición 1.3.2. *Si T es una v.a. discreta con valores $t_0, t_1, t_2, ...,$ donde $t_j < t_{j+1}$, $j = 0, 1, 2, ...$ y f_T su correspondiente función de densidad de probabilidad, la función de riesgo denotada por h se define como:*

$$h(t) = \frac{f_T(t)}{S(t)} \quad \textit{con } S(t) \neq 0$$

donde,

$$f_T(t) = \begin{cases} P(T = t) & \textit{si } t = t_j \textit{, con } j = 0, 1, 2, ... \\ \\ 0 & \textit{en otro caso.} \end{cases}$$

Así, $h(t_j)$ se puede expresar como:

$$h(t_j) = \frac{P(T = t_j)}{S(t_{j-1})} \tag{1.9}$$

Nótese que S se evalúa en el punto t_{j-1}, ya que $P(T > t_{j-1}) = P(T \geq t_j)$

Teorema 1.3.2. *Si T es una variable aleatoria discreta con valores $t_0, t_1, t_2, ...,$ donde $t_j < t_{j+1}$ para $j = 0, 1, 2, ...$ y $t_j < w$, entonces:*

$$h(t_j) = 1 - \frac{S(t_j)}{S(t_{j-1})}, \; j = 1, 2, ...$$

Demostración: Por (1.9) tenemos:

$$h(t_j) = \frac{P(T = t_j)}{S(t_{j-1})}$$

Por propidades de probabilidad:

$$P(T = t_j) = S(t_{j-1}) - S(t_j) \tag{1.10}$$

Luego,

$$h(t_j) = \frac{S(t_{j-1}) - S(t_j)}{S(t_{j-1})}$$

Por lo tanto,

$$h(t_j) = 1 - \frac{S(t_j)}{S(t_{j-1})} \qquad j = 1, 2, ... \tag{1.11}$$

□

Teorema 1.3.3. *Si T es una v.a. continua y b un número positivo, tal que S es diferenciable en $[0, b]$, entonces para cada $t \in [0, b]$, se tiene que:*

$$S(t) = \exp\left[-\int_0^t h(x)dx\right]$$

Demostración: Sea $t \in (0, b]$, $S(t) > 0$. Así que, por (1.7) tenemos:

$$h(t) = \frac{-S'(t)}{S(t)}$$

Además, en cada punto de diferenciabilidad de S la función primitiva de $h(\cdot)$ es $\ln[S(\cdot)]+C$ entonces:

$$h(t) = -\frac{d}{dt}[ln(S(t)) + C]$$

Integrando a ambos lados:

$$-\int h(t)\ dt = ln(S(t)) + C$$

Así que:

$$S(t) = \exp\left[-\int_0^t h(x)\ dx\right] \exp(-C)$$

Pero sabemos que:

$$S(0) = 1$$

Luego, $C = 0$ y por lo tanto:

$$S(t) = \exp\left[-\int_0^t h(x)\ dx\right] \tag{1.12}$$

$\square$

Teorema 1.3.4. *Si T es una variable aleatoria discreta con valores $t_0, t_1, t_2, ...,$ donde $t_j < t_{j+1}$ con $j = 0, 1, 2, ...$ y si para cada $t > 0$ consideramos el conjunto A_t definido así:*

$$A_t = \{t_j : t_j \leq t\,,\ P(T = t_j) > 0\}$$

Entonces, la función de sobrevivencia en t puede expresarse como:

$$S(t) = \prod_{t_j \in A_t} [1 - h(t_j)]$$

Demostración: Sabemos que:

$$S(t_j) = P(T > t_j) = P(T \geq t_{j+1})$$

Además, como a cada instante t se le puede asociar un eventoA_t, podemos hallar entonces, la función de sobrevivencia en t considerando dichos eventos, esto es:

$$S(t) = P\left[\bigcap_{t_j \in A_t} (T > t_j)\right]$$

Aplicando propiedades de probabilidad,

$$S(t) = P[T > t_0]P[T > t_1 \mid T > t_0]P[T > t_2 \mid T > t_1]...$$

Como $P[T > t_0] = S(0) = 1$, entonces:

$$S(t) = \prod_{t_j \in A_t} P(T > t_j \mid T > t_{j-1})$$

Por definición de probabiliad condicional,

$$S(t) = \prod_{t_j \in A_t} \frac{P(T > t_j)}{P(T > t_{j-1})}$$

Por definición de función de sobrevivencia,

$$S(t) = \prod_{t_j \in A_t} \frac{S(t_j)}{S(t_{j-1})}$$

Por teorema (1.3.2), sabemos que:

$$h(t_j) = 1 - \frac{S(t_j)}{S(t_{j-1})}$$

Entonces,

$$S(t) = \prod_{t_j \in A_t} [1 - h(t_j)] \tag{1.13}$$

$\square$

Nota: Cuando la v.a. T es continua, S diferenciable y h continua, la relación entre estas dos funciones puede verse en (1.7) y (1.12). En efecto, si se conoce la función de sobrevivencia se puede encontrar la correspondiente función de riesgo y viceversa.

Además, la función de riesgo es de interés por las siguientes razones:

- Suministra el porcentaje de falla condicional.
- Puede usarse para identificar la forma de un modelo de sobrevivencia específico.
- Los modelos de sobrevivencia pueden ser escritos en términos de la función de riesgo.

1.3.2. Esperanza de vida

Definición 1.3.3. *Para cada $x \geq 0$ fijo, tal que $S(x) > 0$, denotemos $T_x = \{y \in \Omega \mid T(y) \geq x\}$ el conjunto de todos los elementos de la población en estudio que sobreviven a un tiempo x. Sea $U_{T_x} = T - x$ la v.a. que representa el tiempo de "vida" que le queda al elemento en estudio, dado que ha sobrevivido a un tiempo x; y f_T la función de densidad de probabilidad de la v.a. T.* ***La Esperanza de vida después de vivir** x **años**, denotada*

$E[U_{T_x}]$ se define como:

$$E[U_{T_x}] = E[T-x|T>x] = \begin{cases} \dfrac{\displaystyle\int_x^{\infty}(t-x)f_T(t)dt}{S(x)} & \text{Si } T \text{ es una v.a. continua,} \\ \dfrac{\displaystyle\sum_{t_j>x}(t_j-x)P(T=t_j)}{S(x)} & \text{Si } T \text{ es una v.a. discreta.} \end{cases}$$

siempre que la integral o la sumatoria correspondiente sea convergente.

Observación 1.3.1. *Si $x=0$ la Esperanza de vida de cualquier individuo de la población es:*

- *Cuando T es una v.a. continua:*

  $$E[T] = \frac{\int_0^{\infty} tf_T(t)dt}{S(0)} = \int_0^{\infty} tf_T(t)dt$$

  Integrando por partes y usando (1.4), tenemos que:

  $$E[T] = \int_0^{\infty} S(t)dt$$

- *Cuando T es una v.a. discreta:*

  $$E[T] = \frac{\displaystyle\sum_{t_j>0} t_jP(T=t_j)}{S(0)} = \sum_{t_j>0} t_jP(T=t_j).$$

En algunos casos $E[T]$ también se denotará como μ_T.

Teorema 1.3.5. *Sea T una v.a. continua. En todo $t<w$ donde S sea diferenciable, se tiene que :*

$$E[U_{T_x}] = \frac{\displaystyle\int_x^{\infty} S(t)dt}{S(x)}$$

Demostración: Por Definición (1.3.3), se tiene que:

$$E[U_{T_x}] = \frac{\displaystyle\int_x^\infty (t-x) f_T(t)dt}{S(x)}$$

$$E[U_{T_x}] = \frac{1}{S(x)} \left[\int_x^\infty t f_T(t)dt - x \int_x^\infty f_T(t)dt \right]$$

Teniendo en cuenta el resultado (1.4), se tiene que:

$$E[U_{T_x}] = \frac{1}{S(x)} \left[\int_x^\infty t f_T(t)dt - xS(x) \right]$$

Pero,

$$\int_x^\infty t f_T(t)dt = \int_0^\infty t f_T(t)dt - \int_0^x t f_T(t)dt$$

Integrando por partes y aplicando (1.3), tenemos que:

$$\int_x^\infty t f_T(t)dt = \int_0^\infty S(t)dt + xS(x) - \int_0^x S(t)dt$$

Entonces, retomando la integral inicial:

$$E[U_{T_x}] = \frac{\int_x^\infty (t-x) f_T(t)dt}{S(x)}$$

$$E[U_{T_x}] = \frac{\int_0^\infty S(t)dt + xS(x) - \int_0^x S(t)dt - xS(x)}{S(x)}$$

$$E[U_{T_x}] = \frac{\int_0^\infty S(t)dt - \int_0^x S(t)dt}{S(x)}$$

$$E[U_{T_x}] = \frac{\int_x^\infty S(t)dt}{S(x)}.$$

□

Teorema 1.3.6. *Si T una variable aleatoria continua, con función de distribución acumulada F_T, tal que $F_T(0) = 0$ y z es una función positiva, diferenciable y monótona, tal que $E[z(T)]$ existe, entonces:*

$$E[z(T)] = \int_0^\infty z(t) f_T(t)dt = z(0) + \int_0^\infty z'(t)[1 - F_T(t)]dt$$

Demostración: Por la observación (1.3.1) tenemos que:

$$E[T] = \int_0^{\infty} t f_T(t) dt$$

Consideremos la v.a. $Y = z(T)$. Como $E[z(T)]$ existe, entonces:

$$E[Y] = \int_0^{\infty} z(t) f_T(t) dt = \lim_{b \to \infty} \int_0^{b} z(t) f_T(t) dt$$

Hallemos la integral $I(b) = \displaystyle\int_0^{b} z(t) f_T(t) dt$

Integrando por partes a $I(b)$ y por ser z diferenciable, se tiene:

$$I(b) = F_T(b) z(b) - z(0) F_T(0) - \int_0^{b} z'(t) F_T(t) dt$$

Como $F_T(0) = 0$ y $z(0)$ es una constante, tenemos:

$$I(b) = F_T(b) z(b) - \int_0^{b} z'(t) F_T(t) dt$$

$$I(b) = z(b)[1 - S(t)] - \int_0^{b} z'(t)[1 - S(t)] dt$$

$$I(b) = z(b) - z(b)S(b) - \int_0^{b} z'(t) dt + \int_0^{b} z'(t) S(t) dt$$

$$I(b) = z(b) - z(b)S(b) - z(b) + z(0) + \int_0^{b} z'(t) S(t) dt$$

$$I(b) = z(0) - z(b)S(b) + \int_0^{b} z'(t) S(t) dt \tag{1.14}$$

Como z es monótona, consideremos los siguientes casos:

- Si z es creciente, entonces:

$$-z(b)S(b) + \int_0^{b} z'(t) S(t) dt \geq -z(b)S(b), \tag{1.15}$$

porque $\displaystyle\int_0^{b} z'(t) S(t) dt \geq 0$

Luego, de (1.14) y (1.15) tenemos:

$$E[Y] = \lim_{b\to\infty}\left[z(0) - z(b)S(b) + \int_0^b z'(t)S(t)dt\right] \geq \lim_{b\to\infty}[z(0) - z(b)S(b)] \qquad (1.16)$$

Como $E[Y]$ existe, entonces por (1.16)cuando $b \to \infty$ tenemos que $z(b)S(b)$ está acotado.

Además, como $\lim_{b\to\infty} S(b) = 0$. Se tiene que:

$$\lim_{b\to\infty}[z(b)S(b)] = 0$$

Luego, como $S(t) = 1 - F_T(t)$, tenemos que:

$$E[Y] = z(0) + \int_0^\infty z'(t)[1 - F_T(t)]dt$$

- Si z es decreciente:

  Como z es una función positiva, continua y decreciente, se tiene que z está acotada. Además,

  $$\lim_{b\to\infty} S(b) = 0$$

  Así que,

  $$\lim_{b\to\infty} z(b)S(b) = 0$$

  Por lo tanto, de (1.14) podemos concluir que:

  $$E[Y] = z(0) + \int_0^\infty z'(t)[1 - F_T(t)]dt \qquad \square$$

Definición 1.3.4. *Sea $q \in \mathbb{R}$, $0 < q < 1$. El q-ésimo percentil de la distribución de la v.a. Y, denotado por ϵ_q, se define como:*

$$\epsilon_q = inf\{\epsilon \in \mathbb{R} \mid F_Y(\epsilon) \geq q \quad \vee \quad S(\epsilon) \geq 1 - q\}$$

Observación 1.3.2.

a) *En el caso de que Y sea una v.a. continua, el q-ésimo percentil de la distribución de la v.a. Y se halla solucionando la ecuación:*

$$S(\epsilon_q) = 1 - q$$

b) *El 0.5 percentil de la distribución de la v.a. T es* ***el tiempo de vida mediano*** *de los elementos de la población en estudio.*

1.3.3. Varianza

Definición 1.3.5. *Sea T una v.a. con función de densidad de probabilidad f_T y esperanza de vida finita μ_T, entonces la varianza de T, $Var[T]$, está dada por:*

$$Var[T] = \begin{cases} \displaystyle\int_0^{\infty} (t-\mu_T)^2 f_T(t)dt & \textit{Si } T \textit{ es una v.a. continua} \\ \displaystyle\sum_{t_j>0} (t_j-\mu_T)^2 P(T=t_j) & \textit{Si } T \textit{ es una v.a. discreta.} \end{cases}$$

siempre que la integral o la sumatoria correspondiente sea convergente.

Teorema 1.3.7. *Si T es una v.a. continua con μ_T y μ_{T^2} finitas, entonces la varianza de T, $Var[T]$, está relacionada con la función de sobrevivencia por medio de la siguiente expresión:*

$$Var[T] = 2\int_0^{\infty} tS(t)dt - \left[\int_0^{\infty} S(t)dt\right]^2$$

Demostración: Por definición (1.3.5) tenemos:

$$Var[T] = \int_0^{\infty} (t-\mu_T)^2 f_T(t)dt$$

$$Var[T] = \int_0^{\infty} [t^2 - 2t\mu_T + (\mu_T)^2] f_T(t)dt$$

$$Var[T] = \int_0^{\infty} t^2 f_T(t)dt - 2\mu_T \int_0^{\infty} t f_T(t)dt + (\mu_T)^2 \int_0^{\infty} f_T(t)dt$$

Por la observación (1.3.1) y (1.4), se tiene que:

$$Var[T] = \int_0^{\infty} t^2 f_T(t)dt - (\mu_T)^2$$

Integrando por partes, haciendo uso de (1.3), se llega a:

$$Var[T] = 2\int_0^{\infty} tS(t)dt - \left[\int_0^{\infty} S(t)dt\right]^2 \qquad \square$$

Teorema 1.3.8. *Si T es una v.a. discreta con μ_T y μ_{T^2} finitas, entonces la varianza de T, $Var[T]$, está relacionada con la función de sobrevivencia por medio de la siguiente expresión:*

$$Var[T] = \sum_{t_j>0}(t_j)^2[S(t_{j-1}) - S(t_j)] - \left[\sum_{t_j>0}(t_j)[S(t_{j-1}) - S(t_j)]\right]^2$$

Demostración: Analogamente al teorema anterior llegamos a que:

$$Var[T] = \sum_{t_j>0}(t_j)^2 P(T = t_j) - 2\left[\sum_{t_j>0} t_j P(T = t_j)\right]^2 + \left[\sum_{t_j>0} t_j P(T = t_j)\right]^2 \left[\sum_{t_j>0} P(T = t_j)\right]$$

Como $\sum_{t_j>0} P(T = t_j) = 1$, entonces:

$$Var[T] = \sum_{t_j>0}(t_j)^2 P(T = t_j) - \left[\sum_{t_j>0} t_j P(T = t_j)\right]^2$$

y por (1.10) tenemos que:

$$Var[T] = \sum_{t_j>0}(t_j)^2[S(t_{j-1}) - S(t_j)] - \left[\sum_{t_j>0} t_j[S(t_{j-1}) - S(t_j)]\right]^2 \qquad \square$$

Observación 1.3.3. *Recordemos que, para toda v.a. Y siempre que existan μ_Y y μ_{Y^2}, se tiene que:*

$$Var[Y] = \mu_{Y^2} - (\mu_Y)^2$$

■

Hemos descrito hasta aquí la teoría básica para el análisis de sobrevivencia, la cual tendremos en cuenta para el estudio de algunos modelos paramétricos de sobrevivencia que se presentan en el siguiente capítulo.

Capítulo 2

Algunos Modelos Paramétricos

2.1. Introducción

A lo largo de la historia, los modelos paramétricos han tenido gran aplicación en campos como la industria, demografía, biomedicina, epidemiología, entre otras, empleándose para describir tiempos de falla (o de sobrevivencia) según el caso. Particularmente, dichos modelos son usados en las investigaciones que analizan datos de sobrevivencia, pues ellos ofrecen ideas acerca de la naturaleza de los diferentes parámetros y de las funciones relacionadas con el tiempo de sobrevivencia, tales como **la función de sobrevivencia**, **la función de riesgo** y **el tiempo de vida esperado**.

En esta sección se presentan algunos de estos modelos, entre los cuales destacamos:**Gamma generalizado, Gamma, Exponencial, Weibull, Log-normal, Gompertz y Pareto**.

Observación 2.1.1. *Consideremos la siguiente función:*

$$f(t) = \begin{cases} \dfrac{\alpha\lambda^{\beta}t^{\alpha\beta-1}\exp(-\lambda t^{\alpha})}{\Gamma(\beta)}, \quad \alpha > 0,\ \beta > 0 \ y \ \lambda > 0 & si\ t \in [0, +\infty) \\ 0 & en\ otro\ caso. \end{cases}$$

Probemos que esta función así definida es una función de densidad de probabilidad, esto es:

a) $f(t) \geq 0 \quad$ *para todo* $t \in \mathbb{R} \quad$ *y*

b) $\displaystyle\int_{-\infty}^{+\infty} f(t)dt = 1$

Solución:

a) *La primera parte es evidente, ya que α, β y λ son positivos.*

b)

$$\int_{-\infty}^{+\infty} f(t)dt = \int_{-\infty}^{0} 0dt + \int_{0}^{+\infty} \frac{\alpha\lambda^{\beta}t^{\alpha\beta-1}\exp(-\lambda t^{\alpha})}{\Gamma(\beta)}dt$$

$$\int_{-\infty}^{+\infty} f(t)dt = \int_{0}^{+\infty} \frac{\alpha\lambda(\lambda t^{\alpha})^{\beta-1}t^{\alpha-1}\exp(-\lambda t^{\alpha})}{\Gamma(\beta)}dt$$

Haciendo $u = \lambda t^{\alpha}$ tenemos que:

$$\int_{-\infty}^{+\infty} f(t)dt = \int_{0}^{+\infty} \frac{u^{\beta-1}\exp(-u)}{\Gamma(\beta)}du$$

Como $\displaystyle\int_{0}^{+\infty} u^{\beta-1}\exp(-u)du = \Gamma(\beta)$ se tiene que:

$$\int_{-\infty}^{+\infty} f(t)dt = \frac{\Gamma(\beta)}{\Gamma(\beta)} = 1$$

Por lo tanto, f así definida es una función de densidad de probabilidad.

2.2. Modelo Gamma generalizado

Sean α, β y λ números reales positivos. Decimos que la v.a. T tiene distribución Gamma generalizada, con parámetros α, β y λ, si su función de densidad de probabilidad, f_T, está dada por:

$$f_T(t) = \frac{\alpha\lambda^{\beta}t^{\alpha\beta-1}\exp(-\lambda t^{\alpha})}{\Gamma(\beta)} \; I_{[0,+\infty)}(t) \tag{2.1}$$

Proposición 2.2.1. *Las funciones de sobrevivencia y de riesgo, la esperanza de vida, la varianza y el q-ésimo percentil de la distribución Gamma generalizada son:*

a) *función de sobrevivencia:*

$$S(t) = 1 - \ell(\lambda t^{\alpha}, \beta), \quad donde \;\; \ell(\lambda t^{\alpha}, \beta) = \frac{\displaystyle\int_0^{\lambda t^{\alpha}} u^{\beta-1}\exp(-u)du}{\Gamma(\beta)}$$

se denomina la función Gamma incompleta.

b) *función de riesgo:*

$$h(t) = \frac{\alpha\lambda^{\beta}t^{\alpha\beta-1}\exp(-\lambda t^{\alpha})}{\Gamma(\beta)[1-\ell(\lambda t^{\alpha}, \beta)]}$$

c) *la esperanza de vida:*

$$E[T] = \frac{\Gamma(\beta + 1/\alpha)}{\lambda^{1/\alpha}\Gamma(\beta)},$$

d) *la varianza:*

$$Var[T] = \frac{1}{\lambda^{2/\alpha}\Gamma(\beta)}\left[\Gamma(\beta + 2/\alpha) - \frac{\Gamma^2(\beta + 1/\alpha)}{\Gamma(\beta)}\right]$$

e) *el q-ésimo percentil* ϵ_q *de la distribución Gamma generalizada es tal que:*

$$q = \int_0^{\lambda\epsilon_q^{\alpha}} \frac{u^{\beta-1}\exp(-u)du}{\Gamma(\beta)}$$

Demostraciones:

a) Por (1.4) se tiene:

$$S(t) = \int_t^{\infty} \frac{\alpha\lambda^{\beta}x^{\alpha\beta-1}\exp(-\lambda x^{\alpha})}{\Gamma(\beta)}dx$$

$$S(t) = \int_t^{\infty} \frac{\alpha\lambda(\lambda x^{\alpha})^{\beta-1}x^{\alpha-1}\exp(-\lambda x^{\alpha})dx}{\Gamma(\beta)}$$

Haciendo la sustitución: $u = \lambda x^{\alpha}$

$$S(t) = \int_{\lambda t^{\alpha}}^{\infty} \frac{u^{\beta-1}\exp(-u)du}{\Gamma(\beta)}$$

$$S(t) = \int_{0}^{\infty} \frac{u^{\beta-1}\exp(-u)du}{\Gamma(\beta)} - \int_{0}^{\lambda t^{\alpha}} \frac{u^{\beta-1}\exp(-u)du}{\Gamma(\beta)}$$

Se sabe que $\quad \Gamma(\beta) = \displaystyle\int_{0}^{\infty} u^{\beta-1}\exp(-u)du,$ entonces:

$$S(t) = 1 - \int_{0}^{\lambda t^{\alpha}} \frac{u^{\beta-1}\exp(-u)du}{\Gamma(\beta)}$$

y como $\quad \ell(\lambda t^{\alpha}, \beta) = \displaystyle\int_{0}^{\lambda t^{\alpha}} \frac{u^{\beta-1}\exp(-u)du}{\Gamma(\beta)}$ es la función Gamma incompleta, se tiene que:

$$S(t) = 1 - \ell(\lambda t^{\alpha}, \beta) \tag{2.2}$$

b) Por teorema(1.3.1), tenemos:

$$h(t) = \frac{\alpha\lambda^{\beta}t^{\alpha\beta-1}\exp(-\lambda t^{\alpha})}{\Gamma(\beta)[1 - \ell(\lambda t^{\alpha}, \beta)]} \tag{2.3}$$

c) Por la Observación (1.3.1):

$$E[T] = \int_{0}^{\infty} \frac{\alpha\lambda^{\beta}t^{\alpha\beta}\exp(-\lambda t^{\alpha})dt}{\Gamma(\beta)}$$

Haciendo la sustitución: $u = \lambda t^{\alpha}$

$$E[T] = \frac{1}{\lambda^{1/\alpha}\Gamma(\beta)} \int_{0}^{\infty} u^{(\beta+1/\alpha)-1}\exp(-u)du$$

$$E[T] = \frac{\Gamma(\beta + 1/\alpha)}{\lambda^{1/\alpha}\Gamma(\beta)} \tag{2.4}$$

d) Por observación (1.3.3) se tiene:

$$Var[T] = \frac{1}{\lambda^{1/\alpha}\Gamma(\beta)}\int_0^\infty \left(\frac{u}{\lambda}\right)^{1/\alpha} u^{(\beta+1/\alpha)-1}\exp(-u)du - \left[\frac{\Gamma(\beta+1/\alpha)}{\lambda^{1/\alpha}\Gamma(\beta)}\right]^2$$

$$Var[T] = \frac{1}{\lambda^{2/\alpha}\Gamma(\beta)}\int_0^\infty u^{(\beta+2/\alpha)-1}\exp(-u)du - \left[\frac{\Gamma(\beta+1/\alpha)}{\lambda^{1/\alpha}\Gamma(\beta)}\right]^2$$

Además, $\displaystyle\int_0^\infty u^{(\beta+2/\alpha)-1}\exp(-u)du = \Gamma(\beta+2/\alpha)$, entonces:

$$Var[T] = \frac{1}{\lambda^{2/\alpha}\Gamma(\beta)}\left[\Gamma(\beta+2/\alpha) - \frac{\Gamma^2(\beta+1/\alpha)}{\Gamma(\beta)}\right] \tag{2.5}$$

e) Por último, el q-ésimo percentil de esta distribución se puede hallar haciendo:

$$S(\epsilon_q) = 1 - q$$

Entonces,

$$S(\epsilon_q) = 1 - \int_0^{\lambda\epsilon_q^\alpha} \frac{u^{\beta-1}\exp(-u)du}{\Gamma(\beta)} = 1 - q$$

Por lo tanto, ϵ_q es tal que:

$$q = \int_0^{\lambda\epsilon_q^\alpha} \frac{u^{\beta-1}\exp(-u)du}{\Gamma(\beta)} \tag{2.6}$$

□

Nota: Despejar ϵ_q de la ecuación (2.6) no es tan inmediato, sin embargo, para casos particulares de la función gamma generalizada se realizará este procedimiento.

2.2.1. Modelo Gamma

Este modelo se aplica con frecuencia a los tiempos de vida de sistemas eléctricos y mecánicos; para modelar la abundancia de especies animales; y también para estudiar los períodos de incubación de enfermedades generadas por algún virus; entre otras.

Este modelo paramétrico es un caso particular del gamma generalizado con $\alpha = 1$. Sean β y λ números reales positivos. Decimos que la v.a. T tiene distribución Gamma, con parámetros β y λ, si su función de densidad de probabilidad, f_T, está dada por:

$$f_T(t) = \frac{\lambda^{\beta} t^{\beta-1} \exp(-\lambda t)}{\Gamma(\beta)} \; I_{[0,+\infty)}(t)$$

Corolario 2.2.1. *Las funciones de sobrevivencia y de riesgo, la esperanza de vida, la Varianza y el q-ésimo percentil de la distribución Gamma son:*

a) *función de sobrevivencia:*

$$S(t) = 1 - \ell(\lambda t, \beta), \quad \textit{donde } \ell(\lambda t, \beta) \textit{ es la función Gamma incompleta}$$

b) *función de riesgo:*

$$h(t) = \frac{\lambda^{\beta} t^{\beta-1} \exp(-\lambda t)}{\Gamma(\beta)[1 - \ell(\lambda t, \beta)]}$$

c) *la esperanza de vida:*

$$E[T] = \frac{\beta}{\lambda},$$

d) *la varianza:*

$$Var[T] = \frac{\beta}{\lambda^2}$$

e) *el q-ésimo percentil* ϵ_q *de la distribución Gamma es tal que:*

$$q = \int_0^{\lambda \epsilon_q} \frac{u^{\beta-1} \exp(-u) du}{\Gamma(\beta)}$$

Demostraciones:

Las siguientes demostraciones se siguen de la proposición (2.2.1) haciendo $\alpha = 1$.

a) Por (2.2) la función de sobrevivencia es:

$$S(t) = 1 - \int_0^{\lambda t} \frac{u^{\beta-1}\exp(-u)du}{\Gamma(\beta)}$$

Como $\quad \ell(\lambda t, \beta) = \displaystyle\int_0^{\lambda t} \frac{u^{\beta-1}\exp(-u)du}{\Gamma(\beta)} \quad$ es la función Gamma incompleta, se tiene que:

$$S(t) = 1 - \ell(\lambda t, \beta) \tag{2.7}$$

b) Por (2.3) la función de riesgo es:

$$h(t) = \frac{\lambda^\beta t^{\beta-1}\exp(-\lambda t)}{\Gamma(\beta)[1 - \ell(\lambda t, \beta)]} \tag{2.8}$$

c) Por (2.4) la esperanza de vida es:

$$E[T] = \frac{\Gamma(\beta+1)}{\lambda\Gamma(\beta)}$$

Como $\Gamma(\beta+1) = \beta\Gamma(\beta)$, entonces

$$E[T] = \frac{\beta}{\lambda} \tag{2.9}$$

d) Por (2.5) tenemos que la varianza es:

$$Var[T] = \frac{1}{\lambda^2\Gamma(\beta)}\left[\Gamma(\beta+2) - \frac{\Gamma^2(\beta+1)}{\Gamma(\beta)}\right]$$

Como $\Gamma(\beta+2) = (\beta+1)\Gamma(\beta+1)$ y a su vez $\Gamma(\beta+1) = \beta\Gamma(\beta)$ entonces:

$$Var[T] = \frac{\beta}{\lambda^2} \tag{2.10}$$

e) Por último, de (2.6) se tiene:

$$q = \int_0^{\lambda\epsilon_q} \frac{u^{\beta-1}\exp(-u)du}{\Gamma(\beta)} \tag{2.11}$$

□

Nota: Como se puede ver en la ecuación (2.11) despejar ϵ_q de esta expresión no es tan inmediato, sin embargo, existen paquetes computacionales tales como *Mathematica* 4 que permiten encontrar valores aproximados para integrales como la anterior, ingresando los parámetros λ y β.

Ejemplo 2.2.1. *Debido a la propagación de los ratones en el mundo, el control de sus poblaciones se ha convertido en un problema trascendental. Por tal motivo un laboratorio realiza una investigación sobre el efecto de un determinado veneno en dichos roedores. Para este propósito se tomó una muestra de* 20 *tiempos de falla (muerte) en horas a partir del momento de la aplicación del veneno (ver tabla* 1*). Se asume además que la v.a. tiempo de falla sigue una distribución gamma con parámetros desconocidos* λ *y* β.

A partir de las observaciones se estiman tales parámetros y con base en ellos, la función de sobrevivencia, la función de riesgo y la esperanza de vida, con el fin de determinar si el veneno produce un efecto rápido en los ratones.

TABLA 1: *tiempos de falla (muerte) de* 20 *ratones de laboratorio.*

Tiempo de falla (en horas)			
3	4	13	30
7	23	11	19
4	25	15	14
4	30	23	11
6	19	25	14

Solución *Sea* T *la v.a. que representa el tiempo de falla (muerte) en los ratones a partir del momento de la aplicación del veneno.*

Hallemos los estimadores $\hat{\lambda}$ *y* $\hat{\beta}$ *de los parámetros* λ *y* β *respectivamente por el método de los momentos. Consideremos una muestra aleatoria de* n*-observaciones* $T_1, T_2, .., T_n$ *de la variable* T *que se supone tiene una función de densidad de probabilidad Gamma con parámetros desconocidos* λ *y* β.

Así que debemos igualar dos pares de momentos poblacionales con dos pares de momentos muestrales.

Los dos primeros momentos poblacionales de la distribución gamma son:

$$\mu_1' = \mu = \frac{\beta}{\lambda} \qquad\qquad \mu_2' = \sigma^2 + \mu^2 = \frac{\beta}{\lambda^2} + \frac{\beta^2}{\lambda^2}$$

Ahora, se igualan estas cantidades a sus momentos muestrales correspondientes, después se despejan λ *y* β*:*

$$\mu_1' = \frac{\beta}{\lambda} = m_1' = \overline{T} \tag{2.12}$$

$$\mu_2' = \frac{\beta}{\lambda^2} + \frac{\beta^2}{\lambda^2} = m_2' = \frac{1}{n}\sum_{i=1}^{n} T_i^2 \tag{2.13}$$

De la ecuación (2.12), se tiene que:

$$\beta = \lambda m_1' \tag{2.14}$$

Así, al sustituir β en la ecuación (2.13) y despejar λ se obtiene:

$$\hat{\lambda} = \frac{m_1'}{m_2' - (m_1')^2}$$

Luego:

$$\hat{\lambda} = \frac{\overline{T}}{\frac{1}{n}\sum_{i=1}^{n} T_i^2 - \overline{T}^2} \tag{2.15}$$

Reemplazando $\hat{\lambda}$ en la ecuación (2.14) obtenemos:

$$\hat{\beta} = \frac{\overline{T}^2}{\frac{1}{n}\sum_{i=1}^{n} T_i^2 - \overline{T}^2} \tag{2.16}$$

Por lo tanto, los estimadores de momento para los parámetros λ y β son $\hat{\lambda}$ y $\hat{\beta}$, dados por (2.15) y (2.16) respectivamente.

De las ecuaciones (2.15), (2.16) y teniendo en cuenta los datos de la tabla 1, se halló la estimación de los parámetros λ y β, a saber:

$$\hat{\lambda} \approx 0.2 \quad y \quad \hat{\beta} \approx 3.0$$

Luego,

$$S(t) = 1 - \int_0^{0.2t} \frac{u^{2.0}\exp(-u)du}{\Gamma(3.0)} = 1 - \ell(0.2t\,;\,3.0)$$

$$h(t) = \frac{(0.2^{3.0})\,t^{2.0}\exp(-0.2t)}{\Gamma(3.0)[1-\ell(0.2t\,;\,3.0)]}; \quad E[T] = \frac{3.0}{0.2} \approx 15 \quad meses$$

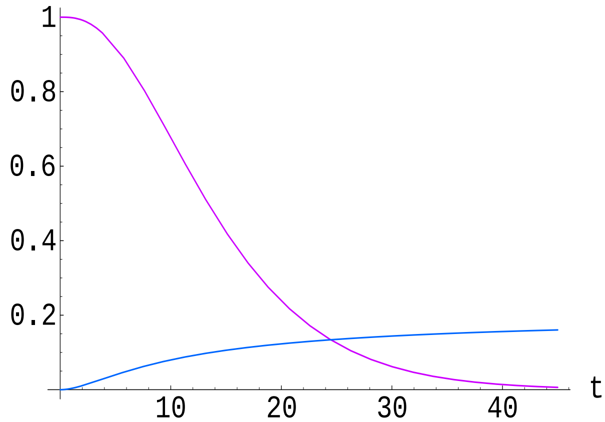

Figura 2.1: funciones de sobrevivencia y de riesgo del modelo Gamma con parámetros $\lambda = 0.2$ y $\beta = 3$.

- *Consideremos por ejemplo el valor $t = 18$ horas, entonces:*

$$S(18) \approx 0.30$$

Es decir, se espera que el $30\,\%$ de los ratones del laboratorio sobrevivan más de 18 horas después de la aplicación del veneno.

$$h(18) = \frac{(0.2^{3.0})\,18^{2.0}\exp[-0.2(18)]}{\Gamma(3.0)[1 - \ell(0.2(18)\,;\,3.0)]} \approx 0.12$$

Este resultado nos muestra que después de aplicado el veneno a los ratones del laboratorio, aproximadamente el $12\,\%$ de los que sobreviven después de las 18 horas, mueren por hora.

$$E[T] = \frac{3.0}{0.2} \approx 15 \quad \textit{horas}$$

Es decir, se espera que en promedio cualquier ratón al que se le aplique el veneno sobreviva 15 horas aproximadamente.

- *Determinemos si el veneno produce un efecto rápido en los ratones.*

  *Entonces, encontremos el tiempo alrededor del cual mueren el $90\,\%$ de los ratones, para ello utilizamos la Tabla B.1.1 del **anexo B**, de donde se obtiene que:*

$$\epsilon_{0.90} \approx 27 \quad \textit{horas}$$

Esto nos indica que aproximadamente al cabo de las 27 horas, se espera que el $90\,\%$ de los ratones del laboratorio a los cuales se les haya suministrado el veneno mueran. Podemos entonces observar que la mayoría de los ratones sometidos a este tratamiento tardan muchas horas en morir, y por lo tanto el veneno no es tan efectivo. ■

2.2.2. Modelo Exponencial

Esta distribución fué estudiada por primera vez en el siglo XIX por Clausius (1858), en la teoría cinética de los gases; más adelante se utilizó para estudios relacionados con salud, por Feigl y Zelen (1965); y por Sheps(1966). Actualmente se utiliza en la teoría de la confiabilidad (falla de componentes o dispositivos electrónicos) y para describir el instante de muerte en los seres humanos. Este modelo se considera un aporte fundamental al análisis de sobrevivencia moderno.

Este modelo paramétrico es un caso particular del modelo Gamma generalizado, con $\alpha = \beta = 1$. Sea λ un número real positivo. Decimos que la v.a. T tiene distribución Exponencial, con parámetro λ, si su función de densidad de probabilidad, f_T, está dada por:

$$f_T(t) = \lambda \exp(-\lambda t) \;\; I_{[0,+\infty)}(t), \qquad \text{con}\lambda > 0$$

Corolario 2.2.2. *Las funciones de sobrevivencia y de riesgo,la esperanza de vida,la varianza y el q-ésimo percentil de la distribucón Exponencial son:*

a) *función de sobrevivencia:*

$$S(t) = \exp(-\lambda t)$$

b) *función de riesgo:*

$$h(t) = \lambda$$

c) *la esperanza de vida:*

$$E[T] = \frac{1}{\lambda}$$

d) *la varianza:*

$$Var[T] = \frac{1}{\lambda^2}$$

e) *el q-ésimo percentil ϵ_q de la distribución exponencial es:*

$$\epsilon_q = -\frac{\ln(1-q)}{\lambda}$$

Demostraciones:

Las siguientes demostraciones se siguen de la proposición (2.2.1) haciendo $\alpha = \beta = 1$.

a) Por (2.2) la función de sobrevivencia es:

$$S(t) = 1 - \int_0^{\lambda t} \frac{u^0 \exp(-u) du}{\Gamma(1)}$$

Como $\Gamma(1) = 1$, entonces:

$$S(t) = 1 - \int_0^{\lambda t} \exp(-u) du$$

Resolviendo la integral, llegamos a que:

$$S(t) = \exp(-\lambda t) \tag{2.17}$$

b) Por (2.3) la función de riesgo es:

$$h(t) = \frac{\lambda \exp(-\lambda t)}{\Gamma(1)[1 - \ell(\lambda t, 1)]}$$

$$h(t) = \frac{\lambda \exp(-\lambda t)}{[1 - \ell(\lambda t, 1)]}$$

y resolviendo la integral $\ell(\lambda t, 1)$, se tiene:

$$h(t) = \frac{\lambda \exp(-\lambda t)}{\exp(-\lambda t)}$$

luego,

$$h(t) = \lambda \tag{2.18}$$

c) Por (2.4) la esperanza de vida es:

$$E[T] = \frac{\Gamma(2)}{\lambda\Gamma(1)}$$

Como $\Gamma(2) = 1$ y $\Gamma(1) = 1$, entonces:

$$E[T] = \frac{1}{\lambda} \tag{2.19}$$

d) De (2.5) la varianza es:

$$Var[T] = \frac{1}{\lambda^2\Gamma(1)}\left[\Gamma(3) - \frac{\Gamma^2(2)}{\Gamma(1)}\right]$$

Así que:

$$Var[T] = \frac{1}{\lambda^2} \tag{2.20}$$

e) Por último, de (2.6) se tiene:

$$q = \int_0^{\lambda\epsilon_q} \frac{\exp(-u)du}{\Gamma(1)}$$

Resolviendo la integral y sabiendo que $\Gamma(1) = 1$:

$$q = -\exp(-\lambda\epsilon_q) + 1 \qquad \text{luego,}$$

$$1 - q = \exp(-\lambda\epsilon_q)$$

Aplicando logaritmo natural a ambos lados de la igualdad:

$$-\lambda\epsilon_q = \ln(1-q)$$

$$\epsilon_q = -\frac{\ln(1-q)}{\lambda} \tag{2.21}$$

□

Ejemplo 2.2.2. *Durante el programa de mantenimiento mensual que realiza una industria vinícola, se necesita analizar el tiempo de sobrevivencia de una válvula mecánica de remontado (nueva y bien montada), usada en la fermentación del vino tinto, con el fin de implementar en la empresa un programa de mantenimiento preventivo. Para ello se toma una muestra aleatoria de* 10 *datos que representan los tiempos hasta que la válvula deja de funcionar a partir del momento de su instalación (ver tabla* 2*).*

Se estimará entonces, el parámetro λ *utilizando los datos de dicha tabla y con base en esta estimación la función de sobrevivencia, la función de riesgo y la esperanza de vida, que permitirá mantener a las válvulas en buenas condiciones de uso, renovándolas antes de que dejen de funcionar, evitando así detener la producción. Para tal efecto, asumimos que el tiempo de falla se distribuye exponencialmente con parámetro* λ.

TABLA 2: *tiempos de falla para* 10 *válvulas mecánicas de bombas de remontado.*

Válvulas mecánicas	*Tiempo de falla (en meses)*
1	*12.5*
2	*15*
3	*48*
4	*63.5*
5	*8*
6	*25*
7	*5*
8	*36*
9	*15*
10	*22*

Solución *Sea* T *la v.a. que representa el tiempo que transcurre desde el momento que se instala una válvula hasta que esta deja de funcionar.*

La estimación $\hat{\lambda}$ *de dicho parámetro se realizó por el método de los momentos y teniendo en cuenta que la distribución exponencial es un caso particular de la distribución Gamma cuando* $\beta = 1$, *entonces podemos retomar la ecuación (2.12), de la cual se tiene:*

$$\mu_1' = \frac{1}{\lambda} = m_1' = \overline{T}$$

De donde:

$$\hat{\lambda} = \frac{1}{\overline{T}} = \frac{1}{25} = 0.04$$

Así, las funciones de sobrevivencia y de riesgo y la esperanza de vida son respectivamente:

$$S(t) = \exp[-0.04t]; \qquad h(t) = 0.04; \qquad E[T] = 25 \quad \textit{meses}$$

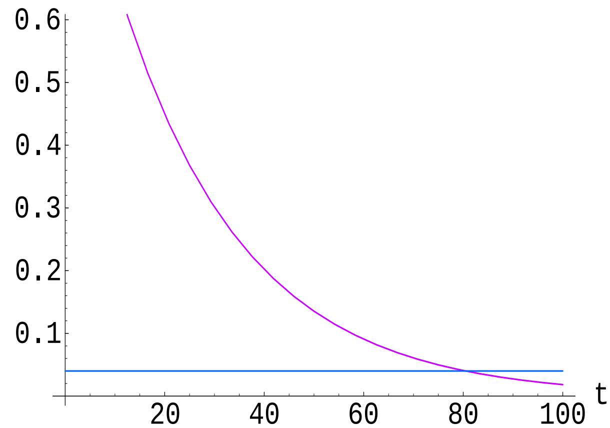

Figura 2.2: funciones de sobrevivencia y de riesgo del modelo Exponencial con parámetro $\lambda = 0.04$.

- *Consideremos por ejemplo el tiempo* $t = 26$ *meses. Analicemos las anteriores funciones para este tiempo:*

  $$S(26) = \exp[-0.04(26)] \approx 0.35$$

  Es decir, se espera entonces que el $35\,\%$ *de las válvulas continúen funcionando después de* 26 *meses a partir de su instalación.*

  $$h(26) \approx 0.04$$

  Este resultado nos muestra que aproximadamente el $4\,\%$ *de las válvulas que continúan funcionando después de los* 26 *meses, fallan por mes.*

  En este modelo, si se toma cualquier tiempo t *la función de riesgo es constante. En el ejemplo se tiene entonces que* $h(t) \approx 0.04$*, lo que nos indica que aproximadamente el* $4\,\%$ *de las válvulas que continúan funcionando después de cualquier tiempo* t*, fallan por mes.*

  $$E[T] \approx 25 \;\; \textit{meses}$$

  Es decir, se espera que en promedio una válvula nueva funcione correctamente alrededor de 25 *meses aproximadamente.*

- *Determinemos cuál es el tiempo aproximado en el que se deben empezar a hacer los ajustes necesarios para que el* $90\,\%$ *de las válvulas mecánicas no empiezen a fallar, para ello evaluemos el* 0.1*-percentil en la ecuación (2.21):*

  $$\epsilon_{0.1} = -\frac{\ln(1-0.1)}{0.04} \approx 2.63 \;\; \textit{meses}$$

Esto nos indica que a partir de los 3 meses aproximadamente, se deben empezar a hacer los ajustes necesarios para que el 90 % de las válvulas mecánicas de las bombas de remontado no empiezen a fallar. A partir de estas consideraciones se ve la necesidad de implementar planes de mantenimiento preventivo que permitan mantener las válvulas en buenas condiciones de uso, renovándolas antes de que se desgasten, evitando así que las bombas de remontado dejen de funcionar y por lo tanto, no se vea afectada la producción del vino en la industria y no ocasione pérdidas económicas a la misma.

■

2.2.3. Modelo Weibull

La distribución Weibull fué introducida por el físico Sueco **Waloddi Weibull** en 1939 y ha sido utilizada extensamente en aplicaciones, tanto industriales como biomédicas. En el campo industrial es aplicable en problemas relativos a fallas por desgaste, "vida"de componentes y materiales usados, ya sea en la construcción de automóviles, aviones, barcos, etc., o para la construcción de las vías para el transporte terrestre. Por otro lado, en el campo biomédico es aplicada para desarrollar pruebas en productos farmacéuticos que permitan determinar la fecha de caducidad de medicamentos (Kapur y Lamberson, 1977; Chow y Liu, 1995; Lawless, 1982)[]. Además, por su flexibilidad ajusta bastante bien los datos que involucran tiempos de aparecimiento de ciertas enfermedades (tumores) en animales y humanos; considerándose así como uno de los modelos más apropiados para describir tiempos de vida.

El modelo Weibull también es un caso particular del modelo Gamma generalizado donde $\beta = 1$. Sean α y λ números reales positivos. Decimos que la v.a. T tiene distribución Weibull, con parámetros α y λ, si su función de densidad de probabilidad, f_T, está dada por:

$$f_T(t) = \lambda \alpha t^{\alpha-1} \exp(-\lambda t^{\alpha}) \;\; I_{[0,+\infty)}(t)$$

Corolario 2.2.3. *Las funciones de sobrevivencia y de riesgo, la esperanza de vida, la varianza y el q-ésimo percentil de la distribución Weibull son:*

a) *función de sobrevivencia:*

$$S(t) = \exp(-\lambda t^{\alpha})$$

b) *función de riesgo:*

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

c) *la esperanaza de vida:*

$$E[T] = \frac{\Gamma(1 + 1/\alpha)}{\lambda^{1/\alpha}}$$

d) *la varianza:*

$$Var[T] = \frac{\Gamma(1 + 2/\alpha) - [\Gamma(1 + 1/\alpha)]^2}{\lambda^{2/\alpha}}$$

e) *el q-ésimo percentil de la distribución Weibull es:*

$$\epsilon_q = \left[\frac{-\ln(1-q)}{\lambda}\right]^{1/\alpha}$$

Demostraciónes:

Las siguientes demostraciones se siguen de la proposición (2.2.1) haciendo $\beta = 1$.

a) Por (2.2) la función de sobrevivencia es:

$$S(t) = 1 - \int_0^{\lambda t^\alpha} \frac{u^0 \exp(-u)du}{\Gamma(1)}$$

Como $\Gamma(1) = 1$, entonces:

$$S(t) = 1 - \int_0^{\lambda t^\alpha} \exp(-u)du$$

Resolviendo la integral, llegamos a que:

$$S(t) = \exp(-\lambda t^\alpha) \tag{2.22}$$

b) Por (2.3) la función de riesgo es:

$$h(t) = \frac{\alpha\lambda t^{\alpha-1}\exp(-\lambda t^{\alpha})}{\Gamma(1)[1-\ell(\lambda t^{\alpha},1)]}$$

$$h(t) = \frac{\alpha\lambda t^{\alpha-1}\exp(-\lambda t^{\alpha})}{[1-\ell(\lambda t^{\alpha},1)]}$$

y resolviendo la integral $\ell(\lambda t^{\alpha},1)$, tenemos:

$$h(t) = \frac{\alpha\lambda t^{\alpha-1}\exp(-\lambda t^{\alpha})}{\exp(-\lambda t^{\alpha})}$$

Luego,

$$h(t) = \alpha\lambda t^{\alpha-1} \tag{2.23}$$

c) Por (2.4) la esperanza de vida es:

$$E[T] = \frac{\Gamma(1+1/\alpha)}{\lambda^{1/\alpha}\Gamma(1)}$$

Como $\Gamma(1) = 1$, entonces:

$$E[T] = \frac{\Gamma(1+1/\alpha)}{\lambda^{1/\alpha}} \tag{2.24}$$

d) Por (2.5) la varianza es:

$$Var[T] = \frac{1}{\lambda^{2/\alpha}\Gamma(1)}\left[\Gamma(1+2/\alpha) - \frac{\Gamma^2(1+1/\alpha)}{\Gamma(1)}\right]$$

Como $\Gamma(1) = 1$ entonces:

$$Var[T] = \frac{1}{\lambda^{2/\alpha}}\left[\Gamma(1+2/\alpha) - \Gamma^2(1+1/\alpha)\right] \tag{2.25}$$

e) Por último, de (2.6) se tiene:

$$q = \int_0^{\lambda\epsilon_q^{\alpha}} \frac{u^0\exp(-u)du}{\Gamma(1)}$$

Resolviendo la integral y sabiendo que $\Gamma(1) = 1$:

$$q = -\exp(-\lambda \epsilon_q^{\alpha}) + 1 \quad \text{luego,}$$

$$1 - q = \exp(-\lambda \epsilon_q^{\alpha})$$

Aplicando logaritmo natural a ambos lados de la igualdad:

$$-\lambda \epsilon_q^{\alpha} = \ln(1 - q)$$

$$\epsilon_q = \left[-\frac{\ln(1-q)}{\lambda} \right]^{1/\alpha} \tag{2.26}$$

$\square$

Observación 2.2.1. *En el medelo de sobrevivencia Weibull, la derivada de la función de riesgo está dada por:*

$$h'(t) = \alpha(\alpha - 1)\lambda t^{\alpha - 2} \tag{2.27}$$

A partir de (2.27) se tienen las siguientes características para la función de riesgo Weibull:

- *Para el caso $\alpha > 1$, la función de riesgo es creciente. A este modelo se le llama **Weibull creciente**.*
- *Para el caso $\alpha < 1$, la función de riesgo es decreciente. A este modelo se le llama **Weibull decreciente**.*
- *En el caso de $\alpha = 1$, la función de riesgo es constante y se obtiene el modelo de sobrevivencia **exponencial**, del que ya se habló en la sección anterior.*

Ejemplo 2.2.3. *Una empresa de alarmas ha querido hacer un estudio sobre la duración de ciertos sistemas mecánicos para detectores de presencia, que tienen la ventaja de ser muy económicos, pero de los cuales no se tiene la seguridad de que funcionen correctamente a largo plazo. Para ello se toma una muestra de los tiempos de falla de 15 sistemas seleccionados aleatoriamente y que han sido ordenados por su duración en meses (ver tabla 3). Dichos sistemas son sometidos a funcionamiento continuo hasta que se produce una falla, que puede ser debida a un daño de un componente por una descarga eléctrica,* *a un descuido por parte del operario, o el desgaste de sus partes.*

Teniendo en cuenta los siguientes datos, se desean estimar los parámetros α y λ, y con base en ellos, la función de sobrevivencia, la función de riesgo y la esperanza de vida, que permitirán analizar la calidad de los sistemas mecánicos y así, comercializarlos o no.

TABLA 3: *tiempos de falla de* 15 *sistemas mecánicos de detectores de presencia.*

Tiempos de falla (en meses)	*Porcentaje de sistemas que han fallado hasta el instante t*
2.2	0.066
3.4	0.132
4.4	0.198
5.3	0.264
6.5	0.330
7.4	0.396
8.3	0.462
9.3	0.528
10.2	0.594
11.4	0.660
12.5	0.726
14.0	0.792
16.0	0.858
19.0	0.924
26.0	0.990

Solución *Sea T la v.a. que representa el tiempo durante el cual un sistema mecánico de detector de presencia funciona correctamente antes de que se produzca una falla. Asumimos que esta v.a. sigue una distribución Weibull con parámetros desconocidos α y λ.*

Hallar los estimadores de los parámetros α y λ de la distribución Weibull con los métodos de los momentos y de máxima verosimilitud es complicado, puesto que se llega a ecuaciones implícitas en las que un parámetro está en términos de él mismo. Para resolver esta situación téngase en cuenta el siguiente análisis:

Se sabe que para el modelo Weibull $S(t) = 1 - F(t) = \exp[-\lambda t^{\alpha}]$ luego,

$$\frac{1}{1 - F(t)} = \exp \lambda t^{\alpha}$$

Evaluando $\ln$ *dos veces consecutivas, tenemos:*

$$\ln\ln\left[\frac{1}{1-F(t)}\right] = \ln\lambda + \alpha\ln t \tag{2.28}$$

Si a esta ecuación se le aplica que:

$$X(t) = \ln t \quad \text{(variable independiente)}$$

$$Y(t) = \ln\ln\left[\frac{1}{1-F(t)}\right] \quad \text{(variable dependiente)}$$

$$B = \ln\lambda \quad \text{(constante)} \tag{2.29}$$

$$A = \alpha \quad \text{(coeficiente director)} \tag{2.30}$$

Entonces se tiene:

$$Y(t) = AX(t) + B$$

Por comodidad se dá por entendida la dependencia de X *y* Y *de* t*, por eso de aquí en adelante no se anotará tal dependencia. Luego ésta ecuación se escribirá así:*

$$Y = AX + B$$

Como se puede observar la anterior ecuación representa una línea recta, entonces con base en el ***modelo de regresión lineal simple***[1] *encontremos para los datos presentados en la tabla 3 la recta que mejor ajuste dichas observaciones. Para llevar a cabo esto se necesita conocer los valores de* A *y* B*, los cuales a su vez permitirán determinar los estimadores de los parámetros* α *y* λ *respectivamente.*

Los estimadores de A *y* B *se encontrarán empleando el método de mínimos cuadrados, y en tal caso, basta con resolver las siguientes ecuaciones:*

$$\sum_{i=1}^{n} y_i = A\sum_{i=1}^{n} x_i + nB \tag{2.31}$$

$$\sum_{i=1}^{n} x_i y_i = A\sum_{i=1}^{n} x_i^2 + B\sum_{i=1}^{n} x_i \tag{2.32}$$

[1] Se puede ver en detalle este modelo en: Probabilidad y Estadística para ingeniería y ciencias; Mendenhall w, Sincich T; Cap.11

donde n es el número de datos observados.

Luego, para este ejemplo ajustaremos la línea de mínimos cuadrados de la ecuación (2.28), a los puntos de datos (x_i, y_i) $i = 1, 2, ..., 15$, donde el i-ésimo punto de datos es:

$$y_i = \ln\ln\left[\frac{1}{1 - F(t_i)}\right] \quad y \quad x_i = \ln t_i$$

Así, reemplazando los respectivos valores en las ecuaciones (2.31) y (2.32) se tiene que:

$$-5.86 = \hat{A}(32.27) + 15\hat{B}$$

$$-2.02 = \hat{A}(75.63) + \hat{B}(32.27)$$

Resolviendo el sistema de ecuaciones obtenemos que:

$$\hat{A} = 1.74 \quad y \quad \hat{B} = -4.13$$

De este modo, teniendo en cuenta las ecuaciones (2.29) y (2.30) los estimadores de los parámetros α y λ son respectivamente:

$$\hat{\alpha} = 1.74 \quad y \quad \hat{\lambda} = 0.016$$

Observación 2.2.2. *Una aproximación a los resultados anteriores se puede obtener utilizando el método gráfico, que consiste en la representación gráfica de los datos observados en un papel especial (ver* ***anexo A.1****). Así para este ejemplo se utilizó el papel Weibull y se obtuvo una línea recta (ver fig.2.3) para la cual se encontró gráficamente que los estimadores de α y λ son respectivamente:*

$$\hat{\alpha} \approx 1.75 \quad y \quad \hat{\lambda} = \left(\frac{1}{10.7}\right)^{1.74} \approx 0.016$$

El procedimiento para llegar a estos resultados se explica detalladamente en el método gráfico mencionado en el ***anexo A.1.1****.*

Por lo tanto, la función de sobrevivencia, la función de riesgo y la esperanza de vida, asociadas a estos parámetros son respectivamente:

$$S(t) = \exp(-0.016t^{1.74}), \quad h(t) = 0.028t^{0.74}, \quad E[T] = \frac{\Gamma(1.57)}{0.016^{1/1.74}} = 9.58 \;\; meses$$

Figura 2.3: Estimación gráfica de los parámetros para el modelo Weibull.

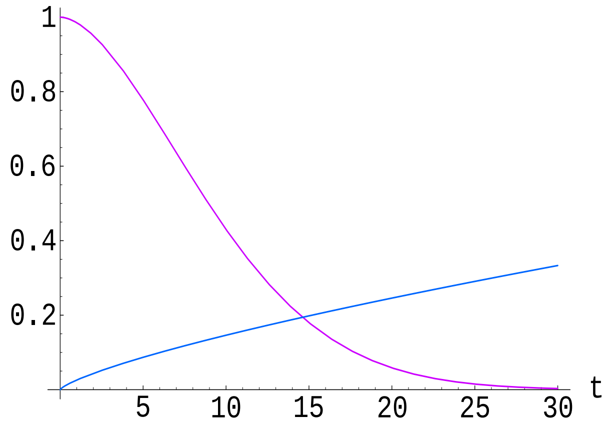

Figura 2.4: funciones de sobrevivencia y de riesgo del modelo Weibull con parámetros $\lambda = 0.015$ y $\alpha = 1.75$.

- *Consideremos por ejemplo el valor $t = 7$ meses, entonces:*

$$S(7) \approx 0.62$$

Es decir, se espera que el $62\,\%$ de los sistemas mecánicos detectores de presencia duren más de 7 meses a partir del momento en que son puestos en funcionamiento.

$$h(7) \approx 0.12$$

Este resultado nos muestra que aproximadamente el $12\,\%$ de los sistemas mecánicos detectores de presencia que continúan funcionando después de los 7 meses, fallan cada mes.

$$E[T] \approx 9.58 \quad \textit{meses}$$

Se espera que en promedio un sistema mecánico detector de presencia funcione correctamente alrededor de 10 meses aproximadamente.

- *Determinemos si los sistemas mecánicos detectores de presencia son efectivos para períodos de tiempo prolongados.*

Encontremos el tiempo alrededor del cual fallan el $90\,\%$ de tales sistemas, para ello evaluemos el 0.9 percentil en la ecuación (2.26) :

$$\epsilon_{0.9} = \left[-\frac{\ln(1-0.9)}{0.016} \right]^{1/1.74}$$

$$\epsilon_{0.9} \approx 17.4 \quad \textit{meses}$$

Esto nos indica que aproximadamente al cabo de los 17 *meses el* 90 % *de los sistemas mecánicos detectores de presencia fallan.*

Podemos entonces observar que la mayoría de dichos sistemas mecánicos no duran mucho tiempo y por lo tanto no serían efectivos a largo plazo. Luego si la empresa desea vender sistemas que además de ser económicos, funcionen por períodos de tiempo prolongados no sería recomendable sacarlos al mercado.

■

2.3. Modelo Lognormal

Este modelo tiene un papel importante en evaluaciones de riesgo, especialmente para modelar tiempos de falla y de sobrevivencia, como por ejemplo, el tiempo transcurrido desde una intervención quirúrgica de un paciente hasta su muerte. Es aplicable también en muchos procesos químicos cuando se analiza el tiempo de duración de un componente metálico en una máquina después de que ha sido sometido a altas temperaturas. En procesos físicos, para modelar el tiempo de vida de un aislamiento eléctrico o el tiempo de duración de una máquina que ha sido reparada. En procesos biológicos y ecológicos; para estudiar el tiempo transcurrido hasta que una bahía es totalmente contaminada; también puede ser aplicado en procesos toxicológicos; por ejemplo, cuando se analiza el tiempo de sobrevivencia de personas que consumen un alimento vencido; entre otras.

Consideremos la v.a.continua T. Se dice que T tiene distribución lognormal, si la v.a. $Y = \ln(T)$ tiene distribución normal.

La distribución lognormal tiene dos parámetros μ y σ^2, la media y la varianza de Y respectivamente. Su función de densidad de probabilidad está dada por:

$$f_T(t) = \frac{\exp\left[-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^2\right]}{t\sqrt{2\pi}\sigma} \; I_{(0,+\infty)}(t)$$

Proposición 2.3.1. *Las funciones de sobrevivencia y de riesgo, la esperanza de vida, la varianza y el q-ésimo percentil de la distribución Lognormal son:*

a) *función de sobrevivencia:*

$$S(t) = 1 - \Phi[z(t)]$$

donde $\Phi[z(\cdot)]$ *es la función de distribución acumulada de la v.a.* $z(T) = \dfrac{\ln(T)-\mu}{\sigma}$*,la cual tiene distribución normal estándar.*

b) *función de riesgo:*

$$h(t) = \frac{\exp\left[-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^2\right]}{t\sqrt{2\pi}\sigma(1-\Phi[z(t)])}$$

c) *la esperanza de vida:*

$$E[T] = \exp\left[\mu + \frac{\sigma^2}{2}\right]$$

d) *la varianza:*

$$Var[T] = \exp[2\mu + \sigma^2][\exp(\sigma^2) - 1]$$

e) *el q-ésimo percentil de la distribución lognormal es tal que:*

$$q = \Phi[z(\epsilon_q)]$$

Demostraciónes:

a) Por (1.4), se tiene:

$$S(t) = \int_t^{\infty} f_T(x)dx \qquad \text{Luego,}$$

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

sea $z(x) = \dfrac{\ln(x)-\mu}{\sigma}$, así, $z(t) = \dfrac{\ln(t)-\mu}{\sigma}$ entonces:

$$S(t) = \frac{1}{\sqrt{2\pi}}\int_{z(t)}^{\infty}\exp\left[-\frac{1}{2}z^2\right]dz$$

Como en la ecuación anterior el integrando es la función de densidad de una v.a. con distribución normal estándar, entonces:

$$S(t) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z(t)} \exp\left[-\frac{1}{2}z^2\right] dz$$

$$S(t) = 1 - \Phi[z(t)] \tag{2.33}$$

b) Por teorema (1.3.1) tenemos que:

$$h(t) = \frac{\exp\left[-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^2\right]}{t\sqrt{2\pi}\sigma(1-\Phi[z(t)])} \tag{2.34}$$

c) Por la observación (1.3.1), se tiene:

$$E[T] = \frac{1}{\sqrt{2\pi}\sigma} \int_0^{\infty} \exp\left[-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^2\right] dt$$

sea $z = \dfrac{\ln(t)-\mu}{\sigma}$, así, $t = \exp[\sigma z + \mu]$ entonces:

$$E[T] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}z^2\right] \exp[\sigma z + \mu] dz$$

$$E[T] = \frac{1}{\sqrt{2\pi}} \exp[\mu] \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(z^2 - 2\sigma z + \sigma^2) + \frac{1}{2}\sigma^2\right] dz$$

$$E[T] = \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp[-\frac{1}{2}(z-\sigma)^2] dz\right] \exp\left[\mu + \frac{\sigma^2}{2}\right]$$

Pero, $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\infty} \exp[-\frac{1}{2}(z-\sigma)^2] dz = 1$ luego,

$$E[T] = \exp\left[\mu + \frac{\sigma^2}{2}\right] \tag{2.35}$$

d) Por la observación (1.3.3) se tiene:

$$Var[T] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}z^2\right] \exp[2\sigma z + 2\mu] dz - \left[\exp(\mu + \sigma^2/2)\right]^2$$

$$Var[T] = \frac{1}{\sqrt{2\pi}} \exp[2\mu] \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(z^2 - 4\sigma z + 4\sigma^2) + 2\sigma^2\right] dz$$

$$- \left[\exp(\mu + \sigma^2/2)\right]^2$$

$$Var[T] = \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(z - 2\sigma)^2\right) dz\right] \exp\left[2\mu + 2\sigma^2\right]$$

$$- \left[\exp(\mu + \sigma^2/2)\right]^2, \quad \text{pero}$$

$$1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(z - 2\sigma)^2\right] dz, \quad \text{luego}$$

$$Var[T] = \exp\left[2\mu + 2\sigma^2\right] - \left[\exp(\mu + \sigma^2/2)\right]^2$$

$$Var[T] = \exp\left[2\mu + 2\sigma^2\right] - \exp(2\mu + \sigma^2)$$

$$Var[T] = \exp\left[2\mu + \sigma^2\right] \left[\exp(\sigma^2) - 1\right] \tag{2.36}$$

e) Por la observación (1.3.2 a.) tenemos que:

$$S(\epsilon_q) = 1 - q, \quad \text{entonces}$$

$$1 - \Phi[z(\epsilon_q)] = 1 - q \quad \text{con } z(\epsilon_q) = \frac{\ln(\epsilon_q) - \mu}{\sigma}, \text{ así:}$$

$$q = \Phi[z(\epsilon_q)] \tag{2.37}$$

□

Nota: Como se puede ver en la ecuación (2.37) despejar ϵ_q de esta expresión no es tan inmediato, sin embargo, para esto se puede hacer uso de la tabla B.1.3 del **anexo B**, que muestra los valores de la función de distribución acumulada normal estándar, la cual facilita estos cálculos.

Ejemplo 2.3.1. *En un hospital se lleva a cabo un estudio con personas que padecen Leucemia (cáncer en la sangre) y que han sido sometidos a fuertes procesos de quimioterapia, produciéndose efectos nocivos que requerían transplantes de médula ósea. Se registró el tiempo (en meses) de* 20 *pacientes desde el momento en el que se realizó el transplante de la médula, hasta el momento en el que cada uno de ellos falleció (ver tabla* 4*). Se desea determinar si el hospital donde se tomaron los datos ofrece buenas espectativas de vida a los pacientes que padecen dicha enfermedad y que son intervenidas quirúrgicamente.*

Asumiendo que el tiempo de falla (muerte) se distribuye lognormalmente con parámetros desconocidos μ y σ^2; se desean estimar estos parámetros a partir de las observaciones y con base en ellos, la función de sobrevivencia, la función de riesgo y la esperanza de vida, con el fin de estudiar la efectividad de los transplantes en el hospital.

TABLA 4: *tiempos de falla (muerte) de pacientes que han sido sometidos a un transplante de médula ósea.*

Tiempo de falla (en meses)			
2	8	20	250
3	9	22	400
4	10	24	600
5	12	30	750
7	15	50	930

Solución *Sea T la v.a. que representa el tiempo transcurrido desde el momento en el que el paciente es sometido a un transplante de médula ósea hasta que ocurre su muerte .*

Hallemos la estimación $\hat{\mu}$ y $\widehat{\sigma^2}$ de los parámetros μ y σ^2 respectivamente. Para esto consideremos una muestra aleatoria $T_1, T_2, ..., T_n$ distribuida lognormalmente con parámetros desconocidos μ, σ^2. Hallemos los estimadores de máxima verosimilitud para dichos parámetros:

Sea $\theta = (\mu, \sigma^2)$, entonces

$$L(\theta; t_1, ..., t_n) = \prod_{i=1}^{n} \frac{\exp\left[-\frac{1}{2}\left(\frac{\ln t_i - \mu}{\sigma}\right)^2\right]}{t_i\sqrt{2\pi}\sigma}, \quad \text{así}$$

$$L(\theta; t_1, ..., t_n) = \frac{\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\ln t_i - \mu}{\sigma}\right)^2\right]}{(2\pi)^{n/2}\sigma^n\prod_{i=1}^{n} t_i}$$

Aplicando ln a ambos lados:

$$\ln[L(\theta; t_1, ..., t_n)] = \ln\left[\frac{1}{(2\pi)^{n/2}\sigma^n}\right] - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\ln t_i - \mu}{\sigma}\right)^2 - \ln\left[\prod_{i=1}^{n} t_i\right]$$

$$\ln[L(\theta; t_1, ..., t_n)] = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln t_i - \mu)^2 - \sum_{i=1}^{n}\ln t_i$$

Al derivar con respecto a μ y σ^2, tenemos:

$$\frac{d\ln[L(\theta; t_1, ..., t_n)]}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(\ln t_i - \mu)$$

$$\frac{d\ln[L(\theta; t_1, ..., t_n)]}{d\mu} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(\ln t_i - \mu)^2$$

Al igualar las derivadas a cero y resolver simultaneamente las ecuaciones se obtienen los estimadores:

$$\hat{\mu} = \frac{\sum_{i=1}^{n}\ln T_i}{n} \tag{2.38}$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n}(\ln T_i - \hat{\mu})^2}{n} \tag{2.39}$$

Observación 2.3.1. *Haciendo algunos cálculos matemáticos se tiene que:*

$$E[\hat{\mu}] = E\left[\frac{1}{n}\sum_{i=1}^{n}\ln T_i\right] = \frac{n}{n-1}\mu \quad y$$

$$E[\widehat{\sigma^2}] = E\left[\frac{1}{n}\sum_{i=1}^{n}(\ln T_i - \hat{\mu})^2\right] = \frac{n}{n-1}\sigma^2 \quad \textit{luego:}$$

$$\mu = E\left[\frac{1}{n-1}\sum_{i=1}^{n}\ln T_i\right] \quad y$$

$$\sigma^2 = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(\ln T_i - \hat{\mu})^2\right],$$

así que los estimadores insesgados para μ y σ^2 son respectivamente:

$$\hat{\mu}' = \frac{1}{n-1}\sum_{i=1}^{n}\ln T_i \quad y \tag{2.40}$$

$$\widehat{\sigma^2}' = \frac{1}{n-1}\sum_{i=1}^{n}(\ln T_i - \hat{\mu}')^2 \tag{2.41}$$

Luego, de las ecuaciones (2.40), (2.41) y las observaciones de la tabla 4 tenemos que:

$$\hat{\mu}' \approx 3.49 \quad y \quad \widehat{\sigma^2}' \approx 3.71$$

Por lo tanto,

$$S(t) = 1 - \Phi\left[\frac{\ln(t) - 3.49}{1.92}\right]$$

$$h(t) = \frac{\exp\left[-\frac{1}{2}\left(\frac{\ln(t) - 3.49}{1.92}\right)^2\right]}{1.92\sqrt{2\pi}t\left[1 - \Phi\left(\frac{\ln(t) - 3.49}{1.92}\right)\right]}$$

$$E[T] = \exp\left[3.49 + \frac{3.71}{2}\right] \approx 209.55 \quad \textit{meses.}$$

- *Consideremos por ejemplo el valor $t = 120$ meses (10 años), entonces:*

$$S(120) \approx 0.25$$

Es decir, se espera que el 25 % de las personas sometidas a un transplante de médula ósea sobrevivan más de 10 años.

$$h(120) \approx 0.005$$

Esto nos indica que aproximadamente el 0.5 % de los individuos que se someten a un transplante de médula ósea y que sobreviven después de los 10 años, mueren por mes.

$$E[T] \approx 209.55 \quad meses$$

Así que, se espera en promedio que cada persona que se someta a un transplante de médula ósea en dicho hospital, sobreviva alrededor de 209 meses (17 años) aproximadamente.

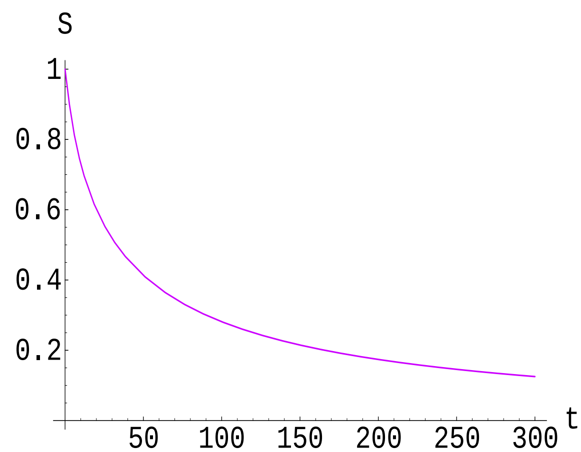

Figura 2.5: función de sobrevivencia Lognormal con parámetros $\mu = 3.42$ y $\sigma^2 = 3.75$.

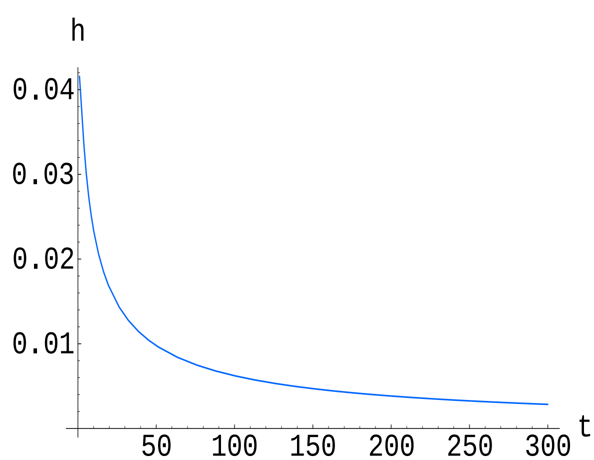

Figura 2.6: función de riesgo Lognormal con parámetros $\mu = 3.42$ y $\sigma^2 = 3.75$.

- *Determinemos el tiempo alrededor del cual se espera que sobrevivan el* 10 % *de las personas que son sometias a un transplante de médula ósea.*

  Como no se puede despejar ϵ_q *de la ecuación (2.37), entonces utilizamos la tabla B.1.3 del* ***anexo B****, de donde se obtiene que:*

$$1.285 = \frac{\ln \epsilon_{0.90} - 3.49}{\sqrt{3.71}}, \quad \textit{luego:}$$

$$\epsilon_{0.90} \approx 389.58 \quad \textit{meses} \quad \textit{(32 años aprox.)}$$

  Por lo tanto, al cabo de los 32 *años aproximadamente el* 90 % *de la población falla. Se espera entonces que este hospital sea un abanderado en las operaciones de médula ósea en el país, ya que ofrece buenas espectativas de vida a los pacientes que padecen leucemia y que son sometidos a esta intervención quirúrgica.*

■

2.4. Modelo Gompertz

Uno de los modelos más importantes utilizado por los actuarios para describir la mortalidad en los seres humanos es el Gompertz, debido a que la mayoría de las tablas de mortalidad (modelos de sobrevivencia presentados en forma tabular) se pueden ajustar a dicho modelo.

La distribución Gompertz consta de dos parámetros positivos α y θ; α indica el incremento en la mortalidad con la edad (llamado parámetro de envejecimiento) y θ el nivel inferior de mortalidad en la tabla.

Sean α y θ números reales positivos. Decimos que la v.a. T tiene distribución Gompertz, con parámetros α y θ, si su función de densidad de probabilidad, f_T, está dada por:

$$f_T(t) = \theta \exp(\alpha t) \exp\left[\frac{\theta}{\alpha}[1 - \exp(\alpha t)]\right] \; I_{[0,+\infty)}(t) \tag{2.42}$$

Se sabe, del capítulo 1, que $h(t)$ mide la susceptibilidad a la muerte en edad t, es por eso que su recíproco $1/h(t)$ se ha usado para medir la resistencia a la muerte. Benjamín Gompertz (1825) suponía que a través del tiempo el hombre va perdiendo la fuerza de vida (o resistencia a la muerte) en partes proporcionales, de esta manera encontró que para todo $t < w$ donde $1/h(t)$ sea diferenciable se tiene que:

$$\frac{d}{dt}\left(\frac{1}{h(t)}\right) = c_1 \frac{1}{h(t)} \qquad \text{esto es,}$$

$$c_1 = \frac{d}{dt} \ln \left[\frac{1}{h(t)}\right] \tag{2.43}$$

Resolviendo (2.43) y despejando $h(t)$ se tiene:

$$h(t) = \exp(-c_1 t - c_2)$$

Sea $\alpha = -c_1$ y $\theta = \exp(-c_2)$, entonces:

$$h(t) = \theta e^{\alpha t}, \qquad \text{donde } \alpha > 0 \text{ y } \theta > 0$$

Observación 2.4.1. *Se dice que la moda m de la curva de mortalidad, es la edad para la cual se espera el mayor número de muertes. Además, según los actuarios, la fuerza de mortalidad (función de riesgo) evaluada en la moda coincide [2] con el parámetro de envejecimiento α, esto es:*

$$h(m) = \theta \exp(\alpha m) = \alpha$$

luego, $\theta = \alpha \exp(-\alpha m)$ y así:

$$h(t) = \alpha \exp[(t - m)\alpha] \tag{2.44}$$

Por teorema (1.3.3) sabemos que:

$$S(t) = \exp\left[-\int_0^t h(x)dx\right] \qquad \textit{luego:}$$

$$S(t) = \exp\left[-\int_0^t \alpha e^{(x-m)\alpha} dx\right]$$

$$S(t) = \exp\left[-\alpha e^{-m\alpha} \int_0^t e^{\alpha x} dx\right]$$

$$S(t) = \exp\left[e^{-m\alpha}(1 - e^{\alpha t})\right] \tag{2.45}$$

[2]Una revisión sobre este tema en *http://www.acst.mq.edu.au/research_papers/acstdem/rp98_002.pdf*

Por (1.3) tenemos:

$$f(t) = -\frac{d}{dt}S(t), \qquad \textit{así:}$$

$$f(t) = -\frac{d}{dt}\exp\left[e^{-m\alpha}(1-e^{\alpha t})\right]$$

$$f(t) = \alpha e^{(-m\alpha+\alpha t)}\exp\left[e^{-m\alpha}(1-e^{\alpha t})\right] \;\; I_{(0,+\infty)}(t) \tag{2.46}$$

Nota: Esta función de densidad de probabilidad es equivalente a la que se presenta en (2.42) y de ahora en adelante se trabajará con ésta por facilidad de cálculos con tablas de vida.

Corolario 2.4.1. *Para la f.d.p. (2.46), las funciones de sobrevivencia y de riesgo, la esperanza de vida, la varianza y el q-ésimo percentil son:*

a) *función de sobrevivencia:*

$$S(t) = \exp\left[e^{-m\alpha}(1-e^{\alpha t})\right]$$

b) *función de riesgo:*

$$h(t) = \alpha\exp[(t-m)\alpha]$$

c) *la esperanza de vida:*

$$E[T] = \frac{\exp(e^{-m\alpha})}{\alpha}\int_{e^{-m\alpha}}^{\infty}\frac{e^{-z}}{z}dz$$

d) *la varianza:*

$$Var[T] = \exp(e^{-m\alpha})\int_{e^{-m\alpha}}^{\infty}\frac{e^{-z}}{z}dz\left[\frac{2m}{\alpha} - \frac{\exp(e^{-m\alpha})}{\alpha^2}\int_{e^{-m\alpha}}^{\infty}\frac{e^{-z}}{z}dz\right] +$$

$$\frac{2\exp(e^{-m\alpha})}{\alpha^2}\int_{e^{-m\alpha}}^{\infty}\frac{e^{-z}\ln z}{z}dz$$

e) *el q-ésimo percentil de la distribución Gompertz es:*

$$\epsilon_q = m + \frac{\ln\left[e^{-m\alpha} - \ln(1-q)\right]}{\alpha} \qquad \textit{siempre que:} \quad 1 - \exp(e^{-m\alpha}) < q < 1$$

Demostraciones:

a) y **b)** se siguen de la observación (2.4.1).

c) Por la observación (1.3.1), se tiene:

$$E[T] = \int_0^{\infty} \exp\left[e^{-m\alpha}[1 - e^{\alpha t}]\right] dt$$

sea $z = e^{-m\alpha + \alpha t}$, así $t = \dfrac{\ln z + m\alpha}{\alpha}$ entonces:

$$E[T] = \frac{\exp(e^{-m\alpha})}{\alpha} \int_{e^{-m\alpha}}^{\infty} \frac{e^{-z}}{z} dz \tag{2.47}$$

Nota: Una forma de aproximar $\displaystyle\int_a^{\infty} \frac{e^{-z}}{z} dz$, utilizando integración numérica[3] es:

$$\int_a^{\infty} \frac{e^{-z}}{z} dz \approx -\gamma - \pi - \ln a + \sum_{n=1}^{\infty} \frac{(-1)^{n+1} a^n}{n(n!)}$$

donde $a > 0$ y $\gamma = 0{,}5772157$ es la llamada constante de Euler.
Por lo tanto, (2.47) se puede aproximar así:

$$E[T] \approx \exp[e^{m\alpha}] \left[-\frac{\gamma}{\alpha} - \frac{\Pi}{\alpha} + m + \frac{1}{\alpha} \sum_{n=1}^{\infty} \frac{(-1)^{n+1} (e^{-m\alpha})^n}{n(n!)} \right] \tag{2.48}$$

d) Por el teorema(1.3.7) siempre que μ_T y μ_{T^2} existan, se tiene:

$$Var[T] = 2\int_0^{\infty} tS(t)dt - \left[\int_0^{\infty} S(t)dt\right]^2 \qquad \text{Luego,}$$

$$Var[T] = 2\int_0^{\infty} t \exp\left[e^{-m\alpha}(1 - e^{\alpha t})\right] dt - \left[\int_0^{\infty} \exp\left[e^{-m\alpha}(1 - e^{\alpha t})\right] dt\right]^2$$

[3] Una revisión de esta aproximación en http://www.mathworld.wolfam.com/exponentialintegral.html

sea $z = e^{-m\alpha+\alpha t}$, entonces:

$$Var[T] = 2\int_{e^{-m\alpha}}^{\infty} \left(\frac{\ln z + m\alpha}{\alpha^2 z}\right) \exp(e^{-m\alpha} - z)\ dz - \left[\frac{\exp(e^{-m\alpha})}{\alpha}\int_{e^{-m\alpha}}^{\infty} \frac{e^{-z}}{z} dz\right]^2$$

$$Var[T] = \frac{2\exp(e^{-m\alpha})}{\alpha}\left[\frac{1}{\alpha}\int_{e^{-m\alpha}}^{\infty} \frac{e^{-z}\ln z}{z} dz + m\int_{e^{-m\alpha}}^{\infty} \frac{e^{-z}}{z} dz\right]$$

$$- \frac{\exp(2e^{-m\alpha})}{\alpha^2}\left[\int_{e^{-m\alpha}}^{\infty} \frac{e^{-z}}{z} dz\right]^2$$

$$Var[T] = \exp(e^{-m\alpha})\int_{e^{-m\alpha}}^{\infty} \frac{e^{-z}}{z} dz \left[\frac{2m}{\alpha} - \frac{\exp(e^{-m\alpha})}{\alpha^2}\int_{e^{-m\alpha}}^{\infty} \frac{e^{-z}}{z} dz\right]$$

$$+ \frac{2\exp(e^{-m\alpha})}{\alpha^2}\int_{e^{-m\alpha}}^{\infty} \frac{e^{-z}\ln z}{z} dz \qquad (2.49)$$

e) Por la observación (1.3.2 a) tenemos que:

$S(\epsilon_q) = 1 - q$ luego,

$1 - q = \exp\left[e^{-m\alpha}(1 - e^{\alpha\epsilon_q})\right]$

Aplicando ln a ambos lados, siempre que $0 \leq q < 1$ tenemos:

$\ln(1 - q) = e^{-m\alpha}(1 - e^{\alpha\epsilon_q})$

$$\epsilon_q = m + \frac{\ln\left[e^{-m\alpha} - \ln(1 - q)\right]}{\alpha}\ , \quad \text{si } 1 - \exp(e^{-m\alpha}) < q < 1 \qquad (2.50)$$

□

Ejemplo 2.4.1. *Una aseguradora muy reconocida en el país está adelantando una investigación sobre la mortalidad de sus asegurados en un lapso de r años, para garantizar que el pago de beneficios debido a la ocurrencia de siniestros no afecte económicamente a la empresa y continúe siendo competitiva a largo plazo.*

Sea T la variable aleatoria que representa la edad de un asegurado en el momento de su fallecimiento; se asume que esta variable sigue una distribución Gompertz con moda m y parámetro de envejecimiento α, ambos desconocidos.

*Si consideramos la **tabla** 5 se puede observar que para esta muestra de asegurados el primer tiempo registrado de fallecimiento fué a los 20 años, así que $t_0 = 20$ años (edad inicial) y que la edad máxima estimada que puede alcanzar un individuo tomado de dicha muestra es de 99 años, por tanto, $w = 100$ años. Se sabe además que el número esperado de sobrevivientes al inicio del estudio es $l_0 = 100000$ personas.*

Con base en las observaciones de la tabla 5 deben hallarse los estimadores de los parámetros y con estos, la función de sobrevivencia, la función de riesgo y la esperanza de vida; que nos permitirá hacer un estudio acerca de algunos aspectos de la mortalidad en la población asegurada y analizar si la mayoría de ellos continúa con vida en un lapso amplio de tiempo.

Solución:

*En el **anexo A.2** se presenta la teoría básica para la costrucción de tablas de mortalidad. Un ejemplo de esto es la tabla 5, en la cual se consignan los siguientes datos: la edad t hasta el fallecimiento (generalmente en años enteros); el número esperado de sobrevivientes hasta una edad t (l_t); la probabilidad de morir antes de 1 año después de tener una edad t (q_t); y el número esperado de muertes entre t y $t+1$ años (d_t).*

***TABLA 5**: datos de mortalidad en una región colombiana.*

Edad t(en años)	l_t	q_t	d_t
20	100000	0.003450	345
21	99655	0.003462	345
28	97222	0.003600	350
29	96872	0.003623	351
35	94745	0.003779	358
36	94387	0.003803	359
44	91412	0.004463	408
45	91004	0.004692	427
54	85776	0.009362	803
55	84973	0.009827	835
63	77557	0.015266	1184
64	76373	0.016917	1292
72	62794	0.034749	2182
73	60612	0.037847	2294
84	25783	0.158050	4075
85	21708	0.186982	4059
91	3739	0.368280	1377
92	2362	0.412362	974
97	53	0.735849	39
98	14	0.857143	12

Fuente: Cálculo Acuarial, contingencias de vida individual; Abel Huertas; pg.216

Se sabe por (A.4) que:

$$S(t) = \frac{l_t}{l_0},$$

donde l_t representa el número esperado de sobrevivientes a la edad t, así para $t = 84$ años tenemos que:

$$S(84) \approx 0.25783$$

Además por (A.5), la probabilidad de morir antes de 1 año dada una edad t, denotada por ${}_1q_t = q_t$ es:

$$q_t = \frac{S(t) - S(t+1)}{S(t)}, \quad \text{así:}$$

$$q(84) \approx 0.158050$$

y esto también se puede ver como la probabilidad de morir entre 84 y 85 años.

De igual modo por (A.9), el número esperado de muertes entre los 84 y 85 años, denotado por ${}_1d_{84} = d_{84}$, es:

$$d(84) \approx 4075$$

Se espera entonces que de los 10000 individuos iniciales mueran 4075 con 84 años de edad.

Según la observación (2.4.1), la moda m de la distribución Gompertz será la edad correspondiente al mayor valor esperado de muertes en la tabla, esto es, $m = 84$ años; luego $k = h(84)$. Usando este resultado y de la ecuación (A.14) :

$$h(t+r) = \frac{q_t}{1 - r(q_t)},$$

Se reemplaza $t = 83$ y $r = 1$ para obtener:

$$k = h(84) \approx 0.1572$$

Por lo tanto,

$$S(t) = \exp[e^{-84(0.1572)}(1 - e^{0.1572t})]$$

$$h(t) = 0.1572\; e^{-84(0.1572)} e^{0.1572t}$$

Además, gracias a la aproximación dada en (2.48) se tiene que:

$$E[T] \approx 80.66 \;\; \textit{años}$$

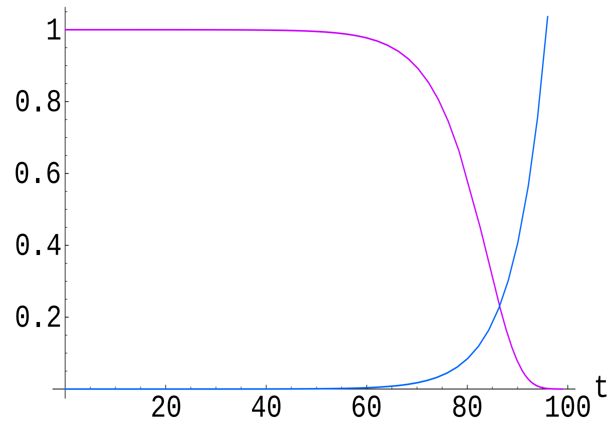

Figura 2.7: función de sobrevivencia y función de riesgo del modelo Gompertz con parámetros $m = 84$ y $\alpha = 0.1572$.

- *Consideremos por ejemplo el valor $t = 87$ años, entonces:*

  $$S(87) \approx 0.2013$$

  Se espera entonces que aproximadamente el 20 % de las personas de dicha población sobrevivan más allá de 87 años.

  $$h(87) \approx 0.2519$$

  Esto nos indica que aproximadamente el 25 % de la población que sobrevive después de los 87 años, muere por año.

  $$E[T] \approx 80.66 \quad \text{años}$$

  Se espera entonces que en promedio, cada persona que vive en la población sobreviva hasta un tiempo de 81 años aproximadamente.

- *Hallemos ahora el 0.10 percentil de este modelo para determinar cuántos años se espera que sobreviva el 90 % de la población de asegurados:*

  De la ecuación (2.50) se tiene que:

  $$\epsilon_{0.10} = 84 + \frac{\ln\left[e^{-84(0.1572)} - \ln(0.9)\right]}{0.1572} \approx 69.68 \quad \textit{años}$$

Por lo tanto, se espera que el 90 % de los asegurados sobrevivan hasta una edad de 70 años aproximadamente, la cual es una edad de sobrevivencia aceptable para la aseguradora, porque esperaría tener que cancelar siniestros solamente al 10 % de la población trancurrido ese tiempo y esto puede resultar favorable para que la empresa sea competitiva a largo plazo.

■

2.5. Modelo Pareto

El modelo Pareto, descubierto por Vilfredo Pareto es muy utilizado en el análisis de datos en sistemas de manufactura que presentan defectos, o en algún componente estructural de un sistema particular.

Pareto consideró el análisis de datos como una parte importante dentro de un programa de mejoramiento de la calidad de cualquier sistema, porque esto le permite a los ingenieros y diseñadores enfocar su atención en los defectos más críticos de un producto o proceso, para que una vez identificados estos defectos, se desarrollen e implementen las acciones correctivas necesarias.

Sean α y θ números reales positivos. Decimos que la v.a. T tiene distribución Pareto, con parámetros α y θ, si su función de densidad de probabilidad, f_T, está dada por:

$$f_T(t) = \frac{\alpha\theta^\alpha}{t^{\alpha+1}} \; I_{[\theta,+\infty)}(t) \quad \text{con } \alpha > 1 \text{ y } \theta > 0$$

Proposición 2.5.1. *Las funciones de sobrevivencia y de riesgo, la esperanza de vida, la varianza y el q-ésimo percentíl de la distribución Pareto son:*

a) *función de sobrevivencia:*

$$S(t) = \left[\frac{\theta}{t}\right]^\alpha$$

b) *función de riesgo:*

$$h(t) = \frac{\alpha}{t}$$

c) *la esperanza de vida:*

$$E[T] = \frac{\alpha\theta}{\alpha - 1}$$

d) *la varianza:*

$$Var[T] = \frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)} \qquad con\ \alpha > 2$$

e) *el q-ésimo percentil de la distribución Pareto es:*

$$\epsilon_q = \frac{\theta}{(1-q)^{1/\alpha}}$$

Demostración:

a) Por (1.4), se tiene:

$$S(t) = \int_t^{\infty} f_T(x)dx \qquad \text{Luego,}$$

$$S(t) = \int_t^{\infty} \frac{\alpha\theta^\alpha}{x^{\alpha+1}}dx$$

$$S(t) = \alpha\theta^\alpha \int_t^{\infty} \frac{1}{x^{\alpha+1}}dx$$

$$S(t) = -\frac{\alpha\theta^\alpha}{\alpha}\left[\lim_{b\to\infty}\frac{1}{b^\alpha} - \frac{1}{t^\alpha}\right]$$

$$S(t) = \left[\frac{\theta}{t}\right]^\alpha \tag{2.51}$$

b) Por teorema (1.3.1) tenemos que:

$$h(t) = \frac{t^\alpha\theta^\alpha\alpha}{t^\alpha\theta^\alpha t}$$

$$h(t) = \frac{\alpha}{t} \tag{2.52}$$

c) Por la observación (1.3.1), se tiene:

$$E[T] = \int_0^{\infty} t f_T(t) dt \qquad \text{luego, como } t \geq \theta \text{, tenemos:}$$

$$E[T] = \alpha\theta^{\alpha} \int_{\theta}^{\infty} \frac{1}{t^{\alpha}} dt$$

$$E[T] = \frac{\alpha\theta^{\alpha}}{1-\alpha} \left[\lim_{b\to\infty} \frac{1}{b^{\alpha-1}} - \frac{1}{\theta^{\alpha-1}} \right]$$

$$E[T] = \frac{\alpha\theta}{\alpha-1} \tag{2.53}$$

d) Por la observación (1.3.3) se tiene:

$$Var[T] = \int_{\theta}^{\infty} t^2 \frac{\alpha\theta^{\alpha}}{t^{\alpha+1}} dt - \left[\frac{\alpha\theta}{\alpha-1} \right]^2$$

$$Var[T] = \alpha\theta^{\alpha} \int_{\theta}^{\infty} \frac{1}{t^{\alpha-1}} dt - \left[\frac{\alpha\theta}{\alpha-1} \right]^2$$

$$Var[T] = \frac{\alpha\theta^{\alpha}}{2-\alpha} \left[\lim_{b\to\infty} \frac{1}{b^{\alpha-2}} - \frac{1}{\theta^{\alpha-2}} \right] - \left[\frac{\alpha\theta}{\alpha-1} \right]^2 \qquad \text{con } \alpha > 2$$

$$Var[T] = \frac{\alpha\theta^2}{\alpha-2} - \left[\frac{\alpha\theta}{\alpha-1} \right]^2$$

$$Var[T] = \frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)} \qquad \text{con } \alpha > 2 \tag{2.54}$$

e) Por la observación (1.3.2 a.) tenemos que:

$$S(\epsilon_q) = 1 - q \quad \text{luego,}$$

$$1 - q = \frac{\theta^{\alpha}}{\epsilon_q^{\alpha}}$$

$$\epsilon_q = \frac{\theta}{(1-q)^{1/\alpha}} \tag{2.55}$$

□

Ejemplo 2.5.1. *Puesto que hoy día las enfermedades del corazón se presentan con mayor frecuencia, y en muchos casos es necesario el implante permanente de un marcapasos que regule el ritmo del latido cardíaco, la unidad de cardiología de cierto hospital ha querido hacer un estudio sobre la vida de la batería de estos aparatos electrónicos de una marca A; su finalidad es determinar alrededor de que tiempo fallan solo el 10 % de dichos marcapasos para tomar medidas preventivas en futuras intervenciones cardiovasculares.*

Para este estudio se toma entonces una muestra de los tiempos de falla en años (tiempo que transcurre hasta que la batería del marcapasos se descarga) de 12 baterías de tales marcapasos (ver tabla 6) y se asume que el tiempo de falla sigue una distribución Pareto con parámetros desconocidos α y θ.

Así, a partir de las observaciones se estiman tales parámetros y con base en ellas la función de sobrevivencia, la función de riesgo y la esperanza de vida.

TABLA 6: *tiempos de falla para una muestra de 12 baterías de marcapasos*

Tiempo de falla (en años)		
5.0	5.1	8.0
11.5	5.2	5.2
5.6	5.1	7.7
5.0	11.4	5.2

Solución *Sea T la v.a que representa la vida de la batería de un marcapasos.*

Los estimadores $\hat{\alpha}$ y $\hat{\theta}$ de los parámetros α y θ respectivamente, se obtienen por medio del método de los momentos como se muestra a continuación.

Los dos primeros momentos poblacionales de la distribución Pareto son:

$$\mu_1' = \mu = \frac{\alpha\theta}{\alpha - 1} \qquad \mu_2' = \sigma^2 + \mu^2 = \frac{\alpha\theta^2}{(\alpha-1)(\alpha-2)} + \left[\frac{\alpha\theta}{\alpha-1}\right]^2$$

Ahora, se igualan estas cantidades a sus momentos muestrales correspondientes, después se despejan α y θ:

$$\mu_1' = \frac{\alpha\theta}{\alpha - 1} = m_1' = \overline{T} \tag{2.56}$$

$$\mu_2' = \frac{\alpha\theta^2}{(\alpha-1)(\alpha-2)} + \left[\frac{\alpha\theta}{\alpha-1}\right]^2 = m_2' = \frac{1}{n}\sum_{i=1}^{n} t_i^2 \tag{2.57}$$

De la ecuación (2.56), se tiene que:

$$\theta = \frac{(\alpha - 1)m_1'}{\alpha} \tag{2.58}$$

Así, al sustituir θ en la ecuación (2.57) se obtiene:

$$m_2' = \frac{\alpha \left[\frac{m_1'(\alpha - 1)}{\alpha}\right]^2}{(\alpha - 1)^2(\alpha - 2)} + \frac{\alpha^2 \left[\frac{m_1'(\alpha - 1)}{\alpha}\right]^2}{(\alpha - 1)^2}$$

De donde,

$$(m_1')^2(\alpha - 1)^2 = (\alpha^2 - 2\alpha)m_2'$$

Completando cuadrados tenemos:

$$(m_1')^2(\alpha - 1)^2 = [(\alpha - 1)^2 - 1]m_2'$$

$$m_2' = (\alpha - 1)^2[m_2' - (m_1')^2]$$

Al despejar α de la ecuación anterior se tiene:

$$\hat{\alpha} = 1 + \sqrt{\frac{m_2'}{m_2' - (m_1')^2}} \tag{2.59}$$

Y reemplazando este valor en (2.58):

$$\hat{\theta} = \frac{m_1'\sqrt{\frac{m_2'}{m_2' - (m_1')^2}}}{1 + \sqrt{\frac{m_2'}{m_2' - (m_1')^2}}} \tag{2.60}$$

Donde,

$$m_1' = \overline{T} \quad y \quad m_2' = \frac{1}{n}\sum_{i=1}^{n} T_i^2$$

Por lo tanto, los estimadores de momento para los parámetros α *y* θ *son* $\hat{\alpha}$ *y* $\hat{\theta}$ *de las ecuaciones (2.59)y (2.60) respectivamente.*

Así que a partir de las ecuaciones (2.59)y (2.60) y de los datos observados, una estimación de los parámetros es:

$$\hat{\alpha} \approx 4.0 \quad y \quad \hat{\theta} \approx 5.0$$

Luego,

$$S(t) = \left[\frac{5}{t}\right]^4; \quad h(t) = \frac{4}{t}; \quad con \;\; t \geq 5 \quad E[T] = \frac{20}{3} \approx 6.66 \;\; años.$$

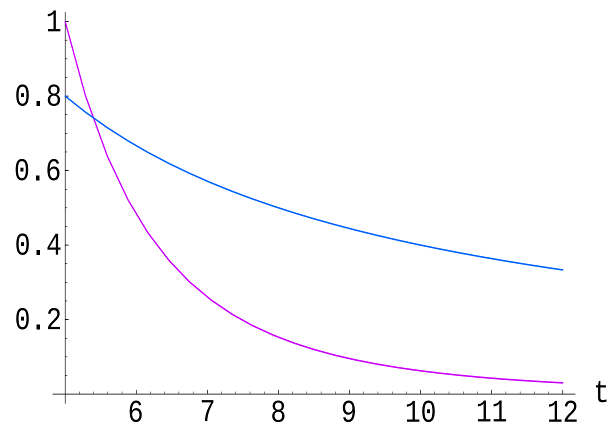

Figura 2.8: funciones de sobrevivencia y de riesgo del modelo Pareto con parámetros $\alpha = 4$ y $\theta = 5$

- *Consideremos por ejemplo el valor* $t = 8$ *años, entonces:*

  $$S(8) \approx 0.15$$

  Es decir, se espera que el $15\,\%$ *de las baterías de los marcapasos duren más de 8 años.*

  $$h(8) \approx 0.5$$

  Este resultado nos muestra que aproximadamente el $50\,\%$ *de las baterías de cualquier marcapaso que continuan funcionando después de los 8 años, fallan por año.*

  $$E[T] \approx 6.66 \;\; años.$$

  Es decir, se espera que en promedio la vida de la batería de un marcapaso dure alrededor de 7 años aproximadamente.

- *Determinemos el tiempo alrededor del cual fallan el* 10 % *de los marcapasos pertenecientes a la firma A.*

  Encontremos entonces el 0.1 *percentil para esta distribución, con la ayuda de la ecuación (2.55). Esto es:*

$$\epsilon_{0.1} = \frac{5}{(1-0.1)^{1/4}} \approx 5.2 \quad \textit{años}$$

  Lo que nos indica que al cabo de 5 *años aproximadamente el* 10 % *de las baterías de estos marcapasos han fallado. Así que, cuando se usen marcapasos de la firma A en implantaciones cardiovasculares, la unidad de cardiología deberá tener en cuenta que a partir de este tiempo los pacientes necesitarán estar en controles más restringidos con el fín de evitar complicaciones mayores y dar una buena garantía al procedimiento que fueron sometidos.* ■

Conclusiones

1. Se estudiaron en detalle los conceptos básicos del análisis de sobrevivencia, tales como las funciones de sobrevivencia y de riesgo, la esperanza de vida, la varianza y el q-ésimo percentil de algunas distribuciones.

2. Se describieron algunos modelos de sobrevivencia parámetricos que se usan en investigaciones con tiempos de vida.

3. Se aplicaron los conceptos de análisis de sobrevivencia a ejemplos concretos para cada modelo estudiado.

4. Se presentaron algunos resultados teóricos cuyas demostraciones no aparecen en la literatura consultada.

Bibliografía

[HUE] HUERTAS, Jaime Abel. *Cálculo Actuarial: Contingencias de vida individual*. Unibiblos. Primera Edición. 2001.

[HOU] HOUGAARD, Philip. *Analysis of Multivariate Survival Data*. Springer.

[KLE] KLEINBAUM, David G. *Survival Analysis*. Springer.

[KMO] KLEIN, John P. MOESCHBERGER, Melvin L. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer.

[MOH] MONTGAMERY, Douglas C. HINES, William W. *Probabilidad y Estadistica para la Ingeniería*. Compañia editorial Continental S.A de C.V México. Segunda Edición.

[MWS] MENDENHALL, William. WACKERLY, Dennis D. SCHEAFFER, Richard L. *Estadistica Matemática con Aplicaciones*. Grupo editorial Iberoamèricano. Segunda Edición.

[MSI] MENDENHALL, William. SINCICH, Terry. *Probabilidad y Estadistica para Ingeniería y Ciencias*. editorial Prentice Hall. Cuarta Edición.1997

[MIL] MILLER, Rupert G. *Survival Analysis*. Wiley Classics Library.

[PLE] PREDOMINGO, Alejandro. LETON, Emilio. *Fundamentos Teóricos del Análisis de Supervivencia*. Glaxowellcome.

[WAM] WALPOLE, Ronald. MYERS, Raymond. *Probabilidad y Estadistica*. McGraw-Hill. Cuarta Edición.

Páginas Web:

http://www.econ.au.dk/ec2bologna/papers/lb_SCI.pdf

http://www.economics.toronto.edu/siow/2801/canals_stern.pdf

http://www.acst.mq.edu.au/research_papers/acstdem/rp98_002.pdf

http://www.mathworld.wolfam.com/exponentialintegral.html

http://www.weibull.com/AccelTestWeb/calculating_the_parameter_of_weibull_distribution.htm

ANEXOS

Anexo A

A.1. Aproximación gráfica para estimar parámetros

Este método también conocido como el "método gráfico", es una manera geométrica de estimar los parámetros de una determinada distribución, haciendo uso de un papel apropiado para el modelo que se estudia. Generalmente se representa en el eje X los tiempos de falla y en el eje Y las probabilidades acumuladas deacuerdo con el modelo en consideración.

Si se conoce la distribución de los datos en estudio, la representación gráfica de estos en el papel adecuado es una línea recta, o se aproxima a una gráfica de este tipo cuando la distribución de los datos es aproximadamente la f.d.p. del modelo representado en el papel.

No obstante, aunque el método gráfico es rápido, fácil de usar y de beneficio adicional para algunos modelos como el Lognormal y Exponencial, también presenta la siguiente desventaja, a saber:

- No se pueden establecer propiedades estadísticas sobre los estimadores de los parámetros.

A.1.1. Modelo Weibull

- Estimación del parámetro α:

  Se sabe que $S(t) = 1 - F(t) = \exp[-\lambda t^{\alpha}]$ luego,

$$\frac{1}{1 - F(t)} = \exp \lambda t^{\alpha}$$

Evaluando ln dos veces consecutivas, tenemos:

$$\ln\ln\left[\frac{1}{1-F(t)}\right] = \ln\lambda + \alpha\ln t$$

Haciendo las siguientes sustituciones:

$$\begin{aligned} X &= \ln t \quad \text{(variable independiente)} \\ Y &= \ln\ln\left[\frac{1}{1-F(t)}\right] \quad \text{(variable dependiente)} \\ B &= \ln\lambda \quad \text{(constante)} \\ A &= \alpha \quad \text{(coeficiente director)} \end{aligned} \tag{A.1}$$

Se tiene que:

$$Y = AX + B \quad \text{(ecuación de una recta)}$$

Analicemos el corte de la recta $Y = AX + B$ con el eje X. Para ello hacemos $Y = 0$ y las sustituciones respectivas, esto es:

$$0 = \alpha\ln(t_0) + \ln(\lambda)$$

Haciendo los cálculos algebraicos necesarios, se tiene:

$$t_0 = \frac{1}{\lambda^{1/\alpha}} \tag{A.2}$$

Ahora hallemos $F(t_0)$ de la ecuación (A.1), así:

$$0 = \ln\ln\left[\frac{1}{1-F(t_0)}\right]$$

Y haciendo los cálculos algebraicos respectivos se tiene:

$$F(t_0) \approx 0.632$$

Luego, las coordenadas del punto de intersección entre las rectas $Y = AX + B$ e $Y = 0.632$ son:

$$t_0 = \frac{1}{\lambda^{1/\alpha}} \ \text{ y } \ F(t_0) = 0.632$$

Por lo tanto, para cualquier valor de α en t_0 se espera que el 63.2 % de la población falle.

De esta forma, el estimador $\hat{\alpha}$ del parámetro α, se encuentra haciendo pasar una recta denotada Y' por el punto $(x_0, 0.632)$ (donde x_0 es una potencia de 10, tal que el dato observado más pequeño es mayor o igual a x_0) y que sea paralela a la recta $Y = AX + B$ obtenida con los datos observados. Luego, el valor $\hat{\alpha}$ se puede leer directamente en una escala tabulada de 0 a 7 marcada en dicho papel y viene dado por la intersección de la recta Y' con la línea que contiene la escala (ver fig. A.1).

- Estimación del parámetro λ:

  Conociendo el valor del estimador $\hat{\alpha}$, de la ecuación (A.2) se despeja λ y se encuentra de esta manera el estimador $\hat{\lambda}$:

$$\hat{\lambda} = \left(\frac{1}{t_0}\right)^{\alpha} \tag{A.3}$$

Figura A.1: Papel Weibull.

A.2. Teoría básica para la construcción de tablas de mortalidad.

Una tabla de mortalidad contiene los elementos básicos que permiten calcular las probabilidades de muerte y sobrevivencia en una población homogénea. Estas tablas son muy utilizadas día a día por los actuarios en sus tareas cotidianas tanto para el cálculo de primas, jubilaciones, bases técnicas de seguros, planes y fondos de pensiones, como en determinados trabajos financieros.

Definición A.2.1. *Una tabla de mortalidad es una serie cronológica (registro estadístico) que expresa la reducción progresiva de un grupo inicial de individuos de la misma edad, por efecto de los fallecimientos.*

Nota: La tabla de mortalidad también es un modelo de sobrevivencia presentado en formato tabular, donde en cada columna se consignan los siguientes datos: la edad t hasta el fallecimiento (generalmente en años enteros); el número esperado de sobrevivientes hasta una edad t (l_t); la probabilidad de morir antes de k años ($k \in \mathbb{Z}^+$) después de tener una edad t (${}_kq_t$); y el número esperado de muertes entre t y $t+k$ años (${}_kd_t$).

Notación: En una tabla de mortalidad de una cierta población, la expresión l_t denotará el número esperado de sobrevivientes a una edad t, y l_0 es el tamaño inicial de la población (raiz de la tabla), que generalmente está fijado en miles de unidades.

Observación A.2.1. *Consideremos la v.a. L_t definida así:*

L_t: Número de sobrevivientes a la edad t de un grupo inicial l_0.

$$L_t = \sum_{j=1}^{l_0} I_{j,t}$$

donde $I_{j,t}$ es la variable indicadora de la sobrevivencia del individuo j en el instante t. Así:

$$I_{j,t} = \begin{cases} 1 & \text{si el } j\text{-ésimo individuo sobrevive a la edad } t \\ \\ 0 & \text{en otro caso.} \end{cases}$$

Luego, $I_{j,t}$ es una v.a. de tipo Benoulli y por tanto, L_t tiene una distribución binomial con parámetros $n = l_0$ (tamaño inicial de la población) y probabilidad de éxito (que la persona sobreviva a la edad t) $p = S(t)$. Entonces:

$$E[L_t] = l_t = np = l_0 S(t) \tag{A.4}$$

$$Var[L_t] = np(1-p) = l_0 S(t)[1 - S(t)]$$

Definición A.2.2. *Sea T la v.a. que expresa el tiempo futuro de vida de una persona, entonces:*

- *La probabilidad de morir antes de k años, dado que ha sobrevivido a una edad t, denotada por ${}_kq_t$, está dada por:*

$${}_kq_t = P[T \leq t + k \mid T > t] = \frac{P[t < T \leq t + k]}{P[T > t]} = \frac{P[T > t] - P[T \geq t + k]}{P[T > t]}$$

$${}_kq_t = \frac{S(t) - S(t+k)}{S(t)} = 1 - \frac{S(t+k)}{S(t)} \tag{A.5}$$

- *La probabilidad de sobrevivir k años, dado que ha sobrevivido a una edad t, denotada ${}_kp_t$, está dada por:*

$${}_kp_t = P[T > t + k \mid T > t] = \frac{P[T > t + k]}{P[T > t]}$$

$${}_kp_t = \frac{S(t+k)}{S(t)} \tag{A.6}$$

Por lo tanto,

$${}_kq_t = 1 - {}_kp_t \tag{A.7}$$

Observación A.2.2. *Consideremos la v.a. ${}_kD_t$ que expresa el número de muertes entre las edades de t y $t + k$, la cual tiene distribución binomial con parámetros $n = l_0$ y $P = F[t + k] - F[t]$.*

Como $P = F[t + k] - F[t] = S(t) - S(t + k)$, entonces el número esperado de muertes entre t y $t + k$, denotado por ${}_kd_t$, es:

$$_kd_t = E[_kD_t] = l_0[S(t) - S(t+k)] \tag{A.8}$$

y por (A.4) tenemos que:

$$_kd_t = l_t - l_{t+k} \tag{A.9}$$

Aplicando (A.4), (A.6) y (A.9) se tiene:

$$_kq_t = \frac{S(t) - S(t+k)}{S(t)} = \frac{l_t - l_{t+k}}{l_t} = \frac{_kd_t}{l_t} \tag{A.10}$$

Observación A.2.3. *Supongamos que el comportamiento de l_t entre t y $t+1$ es lineal. Sea además $0 \leq k \leq 1$, entonces:*

Haciendo una interpolación y aplicando (A.9) se tiene:

$$l_{t+k} = l_t + k(l_{t+1} - l_t) = l_t - k(d_t) \tag{A.11}$$

Por (A.7) se llega a:

$$k(q_t) = 1 - {_kp_t} = {_kq_t} \tag{A.12}$$

Utilizando nuevamente (A.9) tenemos:

$$_kq_t = \frac{S(t) - S(t+k)}{S(t)} = k\left[\frac{S(t) - S(t+1)}{S(t)}\right] = \frac{kS(t) - kS(t+1)}{S(t)}$$

Así,

$$S(t) - S(t+k) = kS(t) - kS(t+1)$$

Esto es:

$$S(t+k) = S(t) - kS(t) + kS(t+1)$$

Luego:

$$\frac{d}{dk}S(t+k) = -S(t) + S(t+1) \tag{A.13}$$

Utilizando (1.7) y (A.13) tenemos:

$$h(t+k) = \frac{-\frac{d}{dk}S(t+k)}{S(t+k)} = \frac{S(t) - S(t+1)}{S(t+k)}$$

Así, dividiendo al numerador y denominador por $S(t)$ con $t < w - k$ y aplicando (A.5) y (A.7) se tiene:

$$h(t+k) = \frac{q_t}{1 - k(q_t)} \tag{A.14}$$

Anexo B

B.1. Algunas tablas de interés

TABLA B.1.1: Percentiles para el modelo Gamma con parámetros $\lambda = 0.2$ y $\beta = 3$

ϵ_q	q	ϵ_q	q	ϵ_q	q	ϵ_q	q
0	0.000000	26	0.891213	52	0.998007	78	0.999977
1	0.001148	27	0.905242	53	0.998311	79	0.999981
2	0.007926	28	0.917612	54	0.998570	80	0.999984
3	0.023115	29	0.928489	55	0.998789	81	0.999986
4	0.047423	30	0.938031	56	0.998976	82	0.999989
5	0.080301	31	0.946382	57	0.999134	83	0.999990
6	0.120513	32	0.953676	58	0.999268	84	0.999992
7	0.166502	33	0.960032	59	0.999381	85	0.999993
8	0.216642	34	0.965562	60	0.999478	86	0.999994
9	0.269379	35	0.970364	61	0.999559	87	0.999995
10	0.323324	36	0.974526	62	0.999628	88	0.999996
11	0.377286	37	0.978129	63	0.999686	89	0.999997
12	0.430291	38	0.981243	64	0.999736	90	0.999997
13	0.481570	39	0.983930	65	0.999777	91	0.999998
14	0.530546	40	0.986246	66	0.999812	92	0.999998
15	0.576810	41	0.988239	67	0.999842	93	0.999998
16	0.620096	42	0.989953	68	0.999867	94	0.999999
17	0.660260	43	0.991424	69	0.999888	95	0.999999
18	0.697253	44	0.992686	70	0.999906	96	0.999999
19	0.731103	45	0.993768	71	0.999921	97	0.999999
20	0.761897	46	0.994693	72	0.999934	98	0.999999
21	0.789762	47	0.995485	73	0.999944	99	0.999999
22	0.814858	48	0.996161	74	0.999953	100	1.000000
23	0.837361	49	0.996738	75	0.999961		
24	0.857461	50	0.997231	76	0.999967		
25	0.875348	51	0.997650	77	0.999972		

TABLA B.1.2: Funciones de densidad de probabilidad, de sobrevivencia, de riesgo y la esperanza de vida para algunos modelos parámetricos.

Modelo paramétrico	$f_T(t)$	$S(t)$	$h(t)$	$E[T]$
Gamma generalizado $\alpha > 0,\ \beta > 0,\ \lambda > 0$ $t \geq 0$	$\frac{\alpha\lambda^{\beta}t^{\alpha\beta-1}e^{-\lambda t^{\alpha}}}{\Gamma(\beta)}$	$1-\ell(\lambda t^{\alpha},\beta)$	$\frac{f_T(t)}{S(t)}$	$\frac{\Gamma(\beta+1/\alpha)}{\lambda^{1/\alpha}\Gamma(\beta)}$
Gamma $\beta > 0,\ \lambda > 0,\ t \geq 0$	$\frac{\lambda^{\beta}t^{\beta-1}\exp(-\lambda t)}{\Gamma(\beta)}$	$1-\ell(\lambda t,\beta)$	$\frac{f_T(t)}{S(t)}$	$\frac{\beta}{\lambda}$
Exponencial $\lambda > 0,\ t \geq 0$	$\lambda\exp(-\lambda t)$	$\exp(-\lambda t)$	λ	$\frac{1}{\lambda}$
Weibull $\alpha > 0,\ \lambda > 0,\ t \geq 0$	$\lambda\alpha t^{\alpha-1}\exp(-\lambda t^{\alpha})$	$\exp(-\lambda t^{\alpha})$	$\lambda\alpha t^{\alpha-1}$	$\frac{\Gamma(1+1/\alpha)}{\lambda^{1/\alpha}}$
Lognormal $\mu \geq 0,\ \sigma > 0,\ t > 0$	$\frac{\exp\left[-\frac{1}{2}[z(t)]^2\right]}{t\sqrt{2\pi}\sigma}$	$1-\Phi[z(t)]$	$\frac{f_T(t)}{S(t)}$	$\exp\left[\mu+\frac{\sigma^2}{2}\right]$
Gompertz $m \geq 0,\ \alpha > 0,\ t \geq 0$	$\frac{\alpha e^{(1-e^{\alpha t})e^{-m\alpha}}}{e^{(m\alpha-\alpha t)}}$	$e^{-\int_0^t h(x)dx}$	$\alpha e^{(t-m)\alpha}$	$\int_0^{\infty} S(t)dt$
Pareto $\alpha > 0,\ \theta > 0,\ t \geq \theta$	$\frac{\alpha\theta^{\alpha}}{t^{\alpha+1}}$	$\left[\frac{\theta}{t}\right]^{\alpha}$	$\frac{\alpha}{t}$	$\frac{\alpha\theta}{\alpha-1},\ \alpha > 1$

$*\ell(\lambda t^{\alpha},\beta)$ es la función Gamma incompleta presentada en la proposición(2.2.1) pg.33.

$*\Phi[z(\cdot)]$ es la función de distribución acumulada normal estándar de la v.a.

$$z(T) = \frac{\ln(T)-\mu}{\sigma}.$$

TABLA B.1.3: Valores de la función de distribución acumulada normal estándar.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left[-\frac{1}{2}u^2\right] du = P(Z \leq z)$$

Fuente: P.Meyer, Probabilidad y aplicaciones estadísticas. pgs.348 y 349.

TABLA B.1.3 (continuación)