#### Análise de Sobrevivência e Confiabilidade

Vicente G. Cancho garibay@icmc.usp.br

Departamento de Matemática Aplicada e Estatística
Universidade de São Paulo

## Caracterização dos tempos de sobrevivência

Seja T: o tempo até a ocorrência de um evento de interessa de uma unidade de experimental.

A distribuição de probabilidade de T pode ser especificada de três formas:

- função de sobrevivência (ou confiabilidade);
- função de densidade, e
- função risco.

A probabilidade do evento acontecer após o tempo t e é dada pela função de sobrevivência

$$S(t) = P(T > t)$$

A probabilidade do evento acontecer após o tempo t e é dada pela função de *sobrevivência* 

$$S(t) = P(T > t) = 1 - F(t).$$
 (1)

A probabilidade do evento acontecer após o tempo t e é dada pela função de *sobrevivência* 

$$S(t) = P(T > t) = 1 - F(t).$$
 (1)

onde F(t) é a função de distribuição acumulada da v.a T.

A probabilidade do evento acontecer após o tempo t e é dada pela função de sobrevivência

$$S(t) = P(T > t) = 1 - F(t).$$
 (1)

onde F(t) é a função de distribuição acumulada da v.a T.

#### **Propriedades**

• 
$$\lim_{t\to 0} S(t) = 1$$
;

A probabilidade do evento acontecer após o tempo t e é dada pela função de *sobrevivência* 

$$S(t) = P(T > t) = 1 - F(t).$$
 (1)

onde F(t) é a função de distribuição acumulada da v.a T.

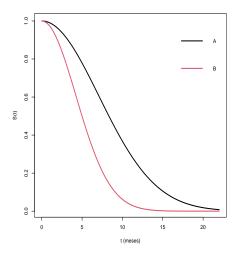
#### **Propriedades**

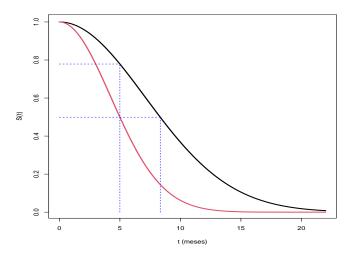
- $\lim_{t\to 0} S(t) = 1$ ;
- $\lim_{t\to\infty} S(t) = 0$ ;
- S(t) é uma função não crescente do tempo t.

#### Observação

- A primeira propriedade significa que no início do acompanhamento, todos os indivíduos não sofreram o evento de interesse.
- O segundo implica que se o tempo de acompanhamento for suficiente longo ( $t \to \infty$ ), todos os indivíduos sofrerão o evento em estudo.
- O segundo também significa que não há fração de cura; veja mais sobre este tópico em Ibrahim et al. (2001) e Rodrigues et al. (2009).
- A terceira propriedade é herdada diretamente de a relação com a função de distribuição cumulativa F(t)

- T: tempo até a falha de produtos:
- A Figura ao lado representa a função de confiabilidade de produtos de 2 marcas;
- O produto da marca A é superior que ao da marca B em relação da durabilidade;
- O produto da marca A é mais confiável que o produto da marca B.





- Note que produtos da marca A o tempo para que ao redor do 50% dos produtos ter falhado (tempo mediano) é aproximadamente 9 meses, enquanto para o produto da marca B é 5 meses.
- Outra informação que podemos retirar desta figura é o porcentual de produtos que estará em operação até um tempo determinado tempo de interesse.
- Assim, para os produtos da marca A, aproximadamente 80% de produtos ainda estarão em operação com 5 meses de uso, enquanto para a marca B é 50%.

## Função de densidade

A função de densidade da v.a. T é definido por

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{S(t) - S(t + \Delta t)}{\Delta t}$$

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

$$= -\frac{dS(t)}{dt} = -S'(t).$$

- A taxa de risco é uma forma útil de descrever a distribuição do "tempo até o evento" porque tem uma interpretação natural que se relaciona com o envelhecimento de uma população.
- Essa terminologia é muito popular na comunidade biomédica.
- Motivamos a definição de taxa de risco definindo primeiro taxa de mortalidade, que é uma versão discreta da taxa de risco.

#### Definição (Taxa de mortalidade)

A taxa de mortalidade no tempo t, onde t é geralmente considerado um número inteiro em termos de alguma unidade de tempo (por exemplo, anos, meses, dias, etc), é a proporção da população que falhou (morre) entre os tempos t e t+1 entre indivíduos vivos no tempo t, ou seja,

$$m(t) = P(t \le T < t + 1 | T > t)$$

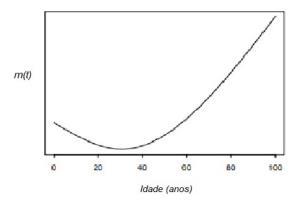


Figura: Padrão de mortalidade típico para humanos

- A taxa de risco h(t) é o limite de uma taxa de mortalidade se o intervalo de tempo for considerado pequeno (em vez de uma unidade).
- A taxa de risco é a taxa instantânea de falha (experimentando o evento) no tempo t, dado que um indivíduo está vivo no tempo t.

Especificamente a taxa de risco é definido pela seguinte equação

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t | T > t)}{\Delta t}.$$
 (2)

Portanto, se  $\Delta t$  é muito pequeno, temos

$$P(t \leq T < t + \Delta | T > t) \approx \Delta t h(t).$$

A definição da função de risco implica que]

$$h(t) = \frac{\lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t)}{\Delta t}}{P(T > t)} = \frac{f(t)}{S(t)}$$
$$= \frac{-S'(t)}{S(t)} = -\frac{d}{dt} \log(S(t))$$
(3)

Em (3), integrando os dois lados obtém-se

$$H(t) = \int_0^t h(u)du = -\log(S(t))$$

onde H(t) é referido como a função de risco cumulado.

Em (3), integrando os dois lados obtém-se

$$H(t) = \int_0^t h(u)du = -\log(S(t))$$

onde H(t) é referido como a função de risco cumulado.

#### Propriedades

•  $\lim_{t\to 0} H(t) = 0$ ;

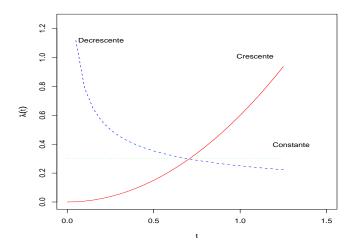
Em (3), integrando os dois lados obtém-se

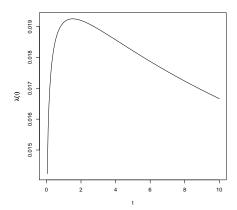
$$H(t) = \int_0^t h(u)du = -\log(S(t))$$

onde H(t) é referido como a função de risco cumulado.

#### Propriedades

- $\lim_{t\to 0} H(t) = 0$ ;
- $\lim_{t\to\infty} H(t) = \infty$ ;
- H(t) é uma função crescente do tempo t e ilimitada.





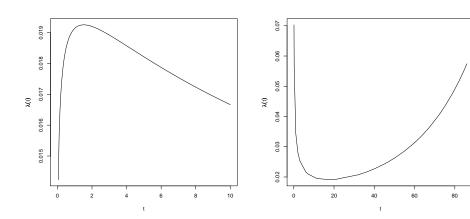


Figura: Função de taxa de falha unimodal e forma de banheira (ou U)



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

•

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

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} (\log S(t)),$$

•

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

•

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} (\log S(t)),$$

$$H(t) = -\log(S(t)),$$

•

•

•

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

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} (\log S(t)),$$

$$H(t) = -\log(S(t)),$$

$$S(t) = \exp(-H(t)) = \exp\left(-\int_0^t h(u)du\right) e$$

•

•

•

•

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

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} (\log S(t)),$$

$$H(t) = -\log(S(t)),$$

$$S(t) = \exp(-H(t)) = \exp\left(-\int_0^t h(u)du\right)$$
 e

$$f(t) = h(t)S(t) = h(t)\exp(-H(t)).$$

### Exemplo-1

Suponha que a taxa de falha instantânea de componentes elétricos, é expressa pela função linear: h(t) = 0.5t

- (a) Determine a função de confiabilidade, e a função de densidade.
- (b) Obtenha o tempo médio de falha.
- (c) Obtenha e interprete S(2).

Seja T o tempo de falha (em anos) de equipamentos. A função de risco acumulado de T

$$H(t) = \int_{t=0}^{t} (0,5u)du = 0,25t^2, \ t > 0$$

A função de sobrevivência (confiabilidade) de T resulta

$$S(t) = \exp\{-0.25t^2\} \ t > 0$$

A f.d.p de T resulta

$$f(t) = 0,5t \exp\{-0.25t^2\}$$
  $t > 0$ 

#### O tempo médio de falha

$$E(T) = \int_0^\infty tf(t)dt = \int_0^\infty 0.5t^2 \exp\{-0.25t^2\}dt$$
$$= 2\frac{\Gamma(3/2)}{2} = \sqrt{\pi}.$$
$$S(2) = \exp\{-0.25 \times 2^2\} = 0.36(36\%)$$

## Algumas medidas em análise de sobrevivência

 O tempo médio de vida é obtida pela área sob a função de sobrevivência. Isto é,

$$E(T) = \int_0^\infty t f(t) dt = \int_0^\infty S(t) dt$$

• A vida média residual (VMR) é obtida pela área sob a curva de sobrevivência à direita do tempo t e dividida por S(t). Ou seja,

$$VMR(t) = E[T-t|T>t] = \frac{\int_t^{\infty} (u-t)f(u)du}{S(t)} = \frac{\int_t^{\infty} S(u)du}{S(t)}.$$

## Algumas medidas em análise de sobrevivência

 O tempo médio de vida é obtida pela área sob a função de sobrevivência. Isto é,

$$E(T) = \int_0^\infty t f(t) dt = \int_0^\infty S(t) dt$$

• A vida média residual (VMR) é obtida pela área sob a curva de sobrevivência à direita do tempo t e dividida por S(t). Ou seja,

$$VMR(t) = E[T-t|T>t] = \frac{\int_t^{\infty} (u-t)f(u)du}{S(t)} = \frac{\int_t^{\infty} S(u)du}{S(t)}.$$

A VMR mede para indivíduos com idade *t* o tempo médio restante de vida.

• O q-ésimo quantil da distribuição T é o valor tal que  $t_q$  tal que

$$P(T \le t_q) = q, \, 0 < q < 1.$$

Para Exemplo 1, tem-se:

$$\begin{split} E[T] &= \int_0^\infty \exp\{-0, 25t^2\} dt = \int_0^\infty \exp\{-u\} \frac{2}{u^{1/2}} du \\ &= 2 \int_0^\infty u^{-\frac{1}{2} + 1 - 1} e^{-u} du = 2\Gamma(3/2) = \sqrt{\pi} = 1,772 \text{ anos} \end{split}$$

0 q-ésimo quantil

$$\exp\{-0, 25t_q^2\} = 1 - q, \implies t_q = -2\log(1 - q)^{1/2}$$

Daí o tempo mediano de falha  $t_{0,5}$  é,

$$t_{0,5} = 2(log(2))^2 = 1,665$$
 anos

### Modelos de Sobrevivência

Modelos mais utilizados na análise de sobrevivência e confiabilidade:

- Exponencial.
- Weibull.
- Gama.
- Log-Logística.
- Log-normal.
- Gama Generalizada.

#### Modelo Exponencial

Uma variável aleatória T tem distribuição exponencial com parâmetro  $\lambda>0$  se sua função de densidade de probabilidade (fdp) é dada por

$$f(t) = \lambda e^{-\lambda t}, \ t > 0.$$

A correspondente função de sobrevivência e função de risco são dados

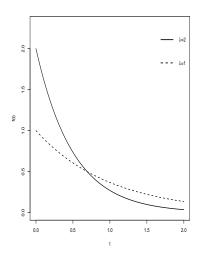
$$S(t)=e^{-\lambda t},\ t>0,$$

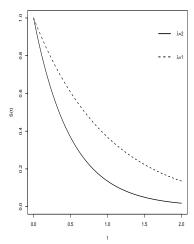
е

$$h(t) = \lambda, t > 0.$$

respectivamente.

Notação:  $X \sim \mathcal{E}(\lambda)$ .





O k-ésimo momento é dada por

$$E[T^k] = \lambda^{-k} \Gamma(k+1)$$

A média e variância são dados por

$$E(T) = 1/\lambda$$
,  $Var(T) = 1/\lambda^2$ .

• O *q*-ésimo quantil é dado por:

$$t_q = -\log(1-q)/\lambda, \ 0 < q < 1.$$

A mediana

$$t_{0.5} = \log(2)/\lambda$$
.



O tempo médio residual

$$VMR(t) = 1/\lambda$$
.

A falta de memória

$$P(T > s + t | T > t) = P(T > t).$$

 Indica que a vida restante de um componente independe de sua idade atual.

- A distribuição exponencial é usada frequentemente para modelar componentes eletrônicos que geralmente não sofrem desgaste até muito tempo depois da vida esperada do produto no qual estão instalados.
- São exemplos os componentes de circuitos integrados de alta qualidade como diodos, transistores, resistores e capacitores.

- (a) Qual é a probabilidade de um destes ventiladores não falhar nas primeiras 8000 horas de funcionamento?
- (b) Determine e interprete o quantil 1%.
- (c) Determine o tempo de falha mediano.
- (c) Determine o quantil correspondente ao tempo médio.

- (a) Qual é a probabilidade de um destes ventiladores não falhar nas primeiras 8000 horas de funcionamento?
- (b) Determine e interprete o quantil 1%.
- (c) Determine o tempo de falha mediano.
- (c) Determine o quantil correspondente ao tempo médio.
  - Seja a v.a T representa o tempo até a falha de ventiladores de motores a diesel.

- (a) Qual é a probabilidade de um destes ventiladores não falhar nas primeiras 8000 horas de funcionamento?
- (b) Determine e interprete o quantil 1%.
- (c) Determine o tempo de falha mediano.
- (c) Determine o quantil correspondente ao tempo médio.
  - Seja a v.a T representa o tempo até a falha de ventiladores de motores a diesel.
  - Do enunciado tem-se  $T \sim \mathcal{E}(\lambda)$  com  $\lambda = 1/28700$ .

- (a) Qual é a probabilidade de um destes ventiladores não falhar nas primeiras 8000 horas de funcionamento?
- (b) Determine e interprete o quantil 1%.
- (c) Determine o tempo de falha mediano.
- (c) Determine o quantil correspondente ao tempo médio.
  - Seja a v.a T representa o tempo até a falha de ventiladores de motores a diesel.
  - Do enunciado tem-se  $T \sim \mathcal{E}(\lambda)$  com  $\lambda = 1/28700$ .
  - (a)  $Pr(T > 8000) = S(8000) = \exp\{\frac{8000}{28700}\} = 0,76.$



Se 8000 horas é o tempo de garantia dada pelo fabricante, significa que 24% é a fração esperada de ventiladores que falhariam na garantia.

• (b)  $t_{0,01} = -\log(1-0,01)/\lambda = 28700(-\log(0,99)) = 288 \text{horas},$ 

- (b)  $t_{0,01} = -\log(1-0,01)/\lambda = 28700(-\log(0,99)) = 288 \text{horas},$
- Isto significa, que é esperado que cerca de 1% dos ventiladores falhem nas primeiras 288 horas de uso

- (b)  $t_{0,01} = -\log(1-0,01)/\lambda = 28700(-\log(0,99)) = 288 \text{horas},$
- Isto significa, que é esperado que cerca de 1% dos ventiladores falhem nas primeiras 288 horas de uso
- (c)  $t_{0,5} = \log(2)/\lambda = 28700(\log(2)) = 19893$  horas

- (b)  $t_{0,01} = -\log(1-0,01)/\lambda = 28700(-\log(0,99)) = 288 \text{horas},$
- Isto significa, que é esperado que cerca de 1% dos ventiladores falhem nas primeiras 288 horas de uso
- (c)  $t_{0,5} = \log(2)/\lambda = 28700(\log(2)) = 19893$  horas
- (d) Para t<sub>q</sub> = 28700, qual é o valor de q? Dos dados q = 0,63. Ou seja o tempo médio de falha corresponde ao quantil 63%.

### Modelo Weibull

Uma v.a. T tem a distribuição Weibull parâmetros  $\lambda>0$  (escala) e  $\alpha>0$  (forma) se sua f.d.p dada por

$$f(t) = \alpha \lambda t^{\alpha - 1} e^{-\lambda t^{\alpha}}, \ t > 0$$
 (4)

Notação:  $T \sim W(\alpha, \lambda)$ .

A função de sobrevivência e função de riscos são dada por,

$$S(t) = e^{-\lambda t^{\alpha}}, \ t > 0 \tag{5}$$

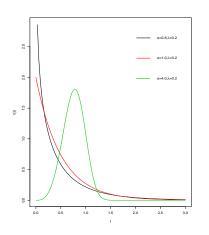
е

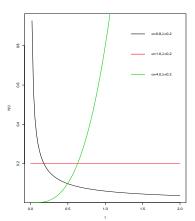
$$h(t) = \alpha \lambda t^{\alpha - 1}, \ t > 0 \tag{6}$$

A forma da função de risco tem as seguintes formas:

- Se  $\alpha=1$  a função de risco é constante (distribuição exponencial)
- Se  $\alpha < 1$  a função de risco é decrescente,
- Se  $\alpha > 1$  a função de risco é crescente.

## Gráfico da f.d.p e a função de risco da distribuição Weibull





# Momentos e a função quantil do modelo Weibull

• O k-ésimo momento é dada por

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

Média e variância

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

onde

$$\Gamma(s) = \int_0^\infty u^{s-1} e^{-u} du$$

é a função gama.



## Modelo Weibull

ullet O p-ésimo quantil, denotado por  $t_p$ 

$$t_p = \lambda^{-1/\alpha} [-\log(1-p)]^{1/\alpha}, \ 0$$

• A mediana,  $md = t_{0,5}$ 

$$t_{0,5} = \lambda^{-1/\alpha} [\log(2)]^{1/\alpha}.$$

### Modelo Weibull

O tempo médio residual

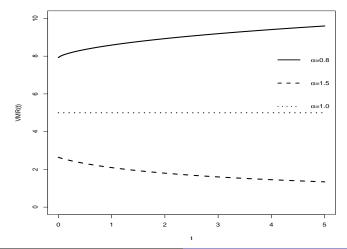
$$VMR(t) = rac{\lambda^{-1/lpha}\Gamma(1/lpha,\lambda t^lpha)}{lpha e^{-\lambda t^lpha}}$$

onde

$$\Gamma(s,t) = \int_t^\infty u^{s-1} e^{-u} du$$

é a função gama incompleta superior.

# Gráfico do VMR para o modelo Weibull para $\alpha=0.8,1,1.5$ e $\lambda=0,2$



### Modelo Gamma

Uma v.a. T tem a distribuição Gamma com parâmetros  $\lambda > 0$  (escala) e  $\alpha > 0$  (forma) se sua f.d.p dada por

$$f(t) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} t^{\alpha - 1} e^{-\lambda t}, \ t > 0$$
 (7)

A função de sobrevivência e função de risco são dada por,

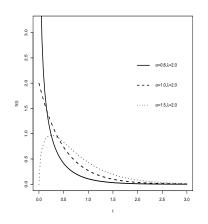
$$S(t) = \frac{\Gamma(\alpha, \lambda t)}{\Gamma(\alpha)}, \ t > 0$$
 (8)

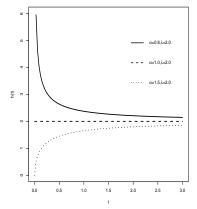
е

$$h(t) = \frac{\lambda^{\alpha} t^{\alpha - 1} e^{-\lambda t}}{\Gamma(\alpha, \lambda t)}, \ t > 0$$
 (9)

Notação:  $T \sim G(\alpha, \lambda)$ 

# Gráfico da f.d.p e a função de risco da distribuição Gamma





### Modelo Gamma

O k-esimo momento

$$E(T^k) = \frac{\Gamma(\alpha + k)}{\lambda^k \Gamma(\alpha)}, \ k = 1, 2, \dots,$$

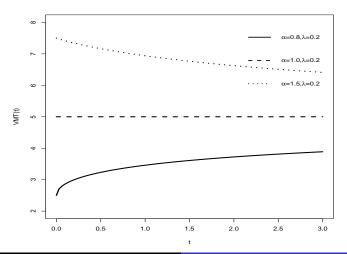
Média e Variância

$$E(T) = \frac{\alpha}{\lambda}, \ \ \mathsf{Var}(T) = \frac{\alpha}{\lambda^2}.$$

Tempo médio de vida residual

$$VMR(t) = \frac{\Gamma(\alpha + 1, \lambda t)}{\lambda \Gamma(\alpha, \lambda t)} - t, \ t > 0$$
 (10)

# Gráfico do VMR para o modelo Gama para lpha=0.8,1,1.5 e $\lambda=0,2$



## O modelo log-logistica

Uma v.a T tem distribuição log-logistica com parâmetros  $\alpha >$  e  $\lambda >$  0 se sua fdp é dada por

$$f(t) = \frac{\alpha \lambda t^{\alpha - 1}}{(1 + \lambda t^{\alpha})^2}.$$

A correspondente função de sobrevivência é dada por

$$S(t) = (1 + \lambda t^{\alpha})^{-1},$$

Notação:  $T \sim LL(\alpha, \lambda)$ 

## O modelo log-logistica

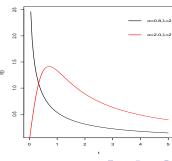
A função de risco da distribuição log-logistica é dada por

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

com  $\alpha > 0$  e  $\lambda > 0$ .

Formas da função de risco

- $\alpha \leq 1$  : é decrescente;
- $\alpha > 1$ : é unimodal.



## O modelo log-logistica

O p-ésimo quantil, denotado por  $t_p$ 

$$t_p = \lambda^{-1/\alpha} \left[ \frac{p}{1-p} \right]^{1/\alpha}, \ 0$$

A mediana

$$t_{0,5}=\lambda^{-1/\alpha}.$$

## Modelo log-Normal

Uma v.a T tem distribuição log-normal se sua f.d.p é dada por

$$f(t) = \frac{1}{\sqrt{2\pi}t\sigma} \exp\left\{-\frac{1}{2}\left(\frac{\log t - \mu}{\sigma}\right)^2\right\}, \ t > 0,$$

onde  $\mu \in R$  é média de log T, assim como  $\sigma > 0$  é o desvio padrão, ou seja, log  $T \sim N(\mu, \sigma^2)$ .

#### A média e variância

$$E(T) = \exp\{\mu + \sigma^2/2\}, \ e \ Var(T) = \exp\{2\mu + \sigma^2\}(\exp\{\sigma^2\} - 1).$$

#### O q-ésimo quantil

$$t_q = \exp\{\mu + \sigma z_q\}$$

onde  $z_q$  é q-ésimo quantil da distribuição normal padrão.

## Modelo log-Normal

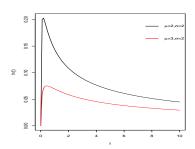
A função de sobrevivência não tem uma forma analática explicita e é representado por

$$S(t) = \Phi\left(-rac{\log t - \mu}{\sigma}
ight)$$

onde  $\Phi(.)$  é a função de distribuição acumulada da distribuição normal padrão.

A correspondente função de risco é dado por:

$$h(t) = \frac{\exp\left\{-\frac{1}{2}\left(\frac{\log t - \mu}{\sigma}\right)^{2}\right\}}{\sqrt{2\pi}t\sigma\Phi\left(-\frac{\log t - \mu}{\sigma}\right)}.$$



## Modelo Gama Generalizado

Uma v.a. T tem a distribuição Gamma com parâmetros  $\lambda > 0$ ,  $\alpha > 0$  e p > 0 se sua f.d.p dada por

$$f(t) = \frac{p\lambda}{\Gamma(\frac{\alpha}{p})} (\lambda t)^{\alpha - 1} e^{-(\lambda t)^{p}}, \ t > 0$$
 (11)

A função de sobrevivência e função de risco são dada por,

$$S(t) = \frac{\Gamma(\frac{\alpha}{p}, (\lambda t)^p)}{\Gamma(\frac{\alpha}{p})}, \ t > 0$$
 (12)

е

$$h(t) = \frac{(\lambda t)^{\alpha - 1} e^{-(\lambda t)^{p}}}{\Gamma(\frac{\alpha}{p}, (\lambda t)^{p})}, \ t > 0$$
 (13)

## Modelo Gama Generalizado

O k-ésimo momento

$$E(T^k) = \frac{\Gamma(\frac{\alpha+k}{p})}{\Gamma(\frac{\alpha}{p})}, \ k = 1, 2, \dots$$

- $GG(\alpha, \lambda, p = 1) = G(\alpha, \lambda)$ .
- $GG(\alpha, \lambda, p = \alpha) = W(\alpha, \lambda^{\alpha}).$

### Exercícios

- (a) Para os modelos log-logística, log-normal e gama generalizado determine o valor médio residual e mostre o comportamento dessas característica ao longo do tempo.
- (b) Para modelo gama generalizado, determine a média e mediana.