

Estatística e Modelos Probabilísticos - COE241

Aula de hoje

V.a. contínua CDF,

Função densidade,

V.a. Exponencial e suas variações,

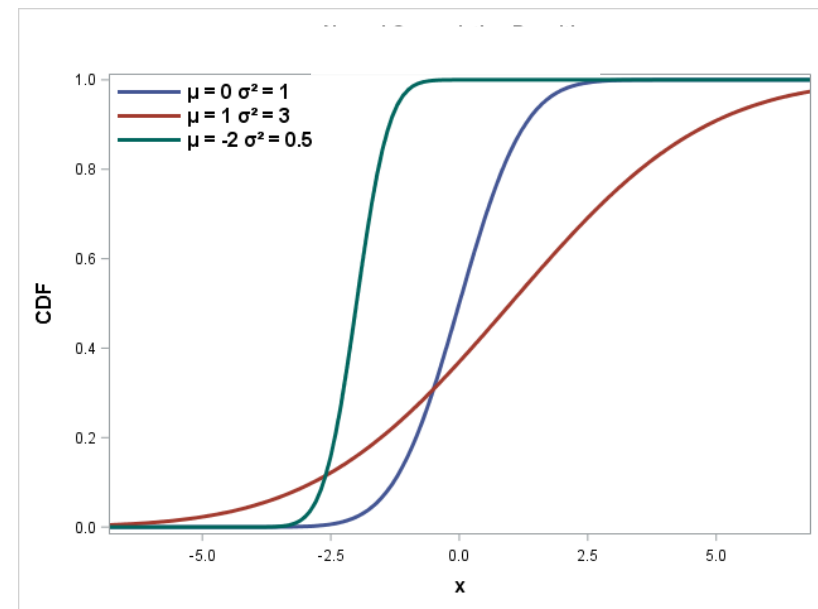
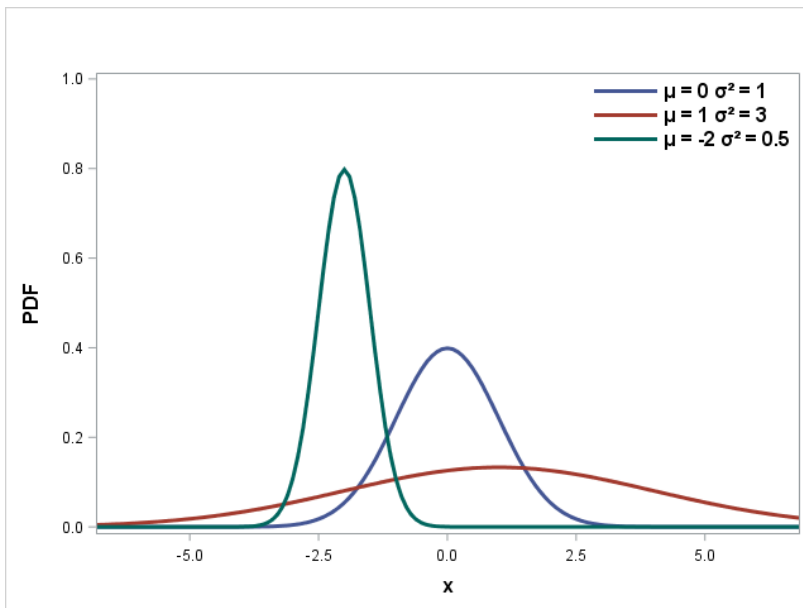
Normal,

Gamma,

Weibull

Variáveis Aleatórias Contínuas

- Duas funções são associadas a uma variável aleatória contínua:
- Função densidade de probabilidade
- Função distribuição de probabilidade



$$F_X(x) = P[X \leq x] = \int_{-\infty}^x f(X) dX$$

Variáveis Aleatórias Contínuas

- Distribution function:

$$F_X(x) = P(X \leq x), \quad -\infty < x < \infty$$

- If $F_X(x)$ is a continuous function of x , then X is a continuous random variable.

- (F1) $0 \leq F_X(x) \leq 1, \quad -\infty < x < \infty$
- (F2) $F_X(x)$: monotonically non-decreasing in x
- (F3) $\lim_{x \rightarrow -\infty} F_X(x) = 0$ and $\lim_{x \rightarrow \infty} F_X(x) = 1$
- (F4') $P(X = c) = P(c \leq X \leq c) = \int_c^c f_Y(y)dy = 0$

Variáveis Aleatórias Contínuas

- X : continuous rv, then, $f(x) = \frac{dF(x)}{dx}$ is the *pdf* of X .

- *CDF and pdf can be derived from each other*

$$P(X \leq x) = F(x) = \int_{-\infty}^x f(u) du, \quad -\infty < x < \infty$$

$$P(X \in (a, b]) = P(a < X \leq b) = \int_a^b f_X(u) du.$$

- *pdf* properties:

- **(f1)** $f(x) \geq 0$ for all x .

- **(f2)** $\int_{-\infty}^{\infty} f(x) dx = 1$.

Variáveis Aleatórias Contínuas: Exemplo

- Random variable X : time (years) to complete a project

$$f_X(x) = \begin{cases} kx(1-x), & 0 \leq x \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

- f_X clearly satisfies property **(f1)**.

- To be a *pdf*, it must also satisfy **(f2)**,

$$\int_0^1 kx(1-x)dx = 1 \xrightarrow{\text{gives}} k \left(\frac{x^2}{2} - \frac{x^3}{3} \right) \Big|_0^1 = 1 \quad \text{or, } k = 6$$

- Prob. of completing project in less than 4 months,

$$P(X < 4/12) = F_X(1/3) = \int_0^{1/3} f_X(x)dx = \frac{7}{27}, \text{ or } 26\%$$

Variáveis Aleatórias: Exponencial

- São muito usadas para avaliar a confiabilidade e desempenho de sistemas
- Possui propriedade “memoryless”
- Está relacionada com a distribuição de Poisson
- É usada para modelar:
 - Tempo entre a chegada de pacotes em uma rede
 - Tempo de serviço
 - Tempo para reparo e tempo até que ocorra uma falha de um componente

Variáveis Aleatórias: Exponencial

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{if } 0 \leq x < \infty \\ 0, & \text{otherwise} \end{cases}$$

where the base of natural logarithm, $e = 2.7182818284$

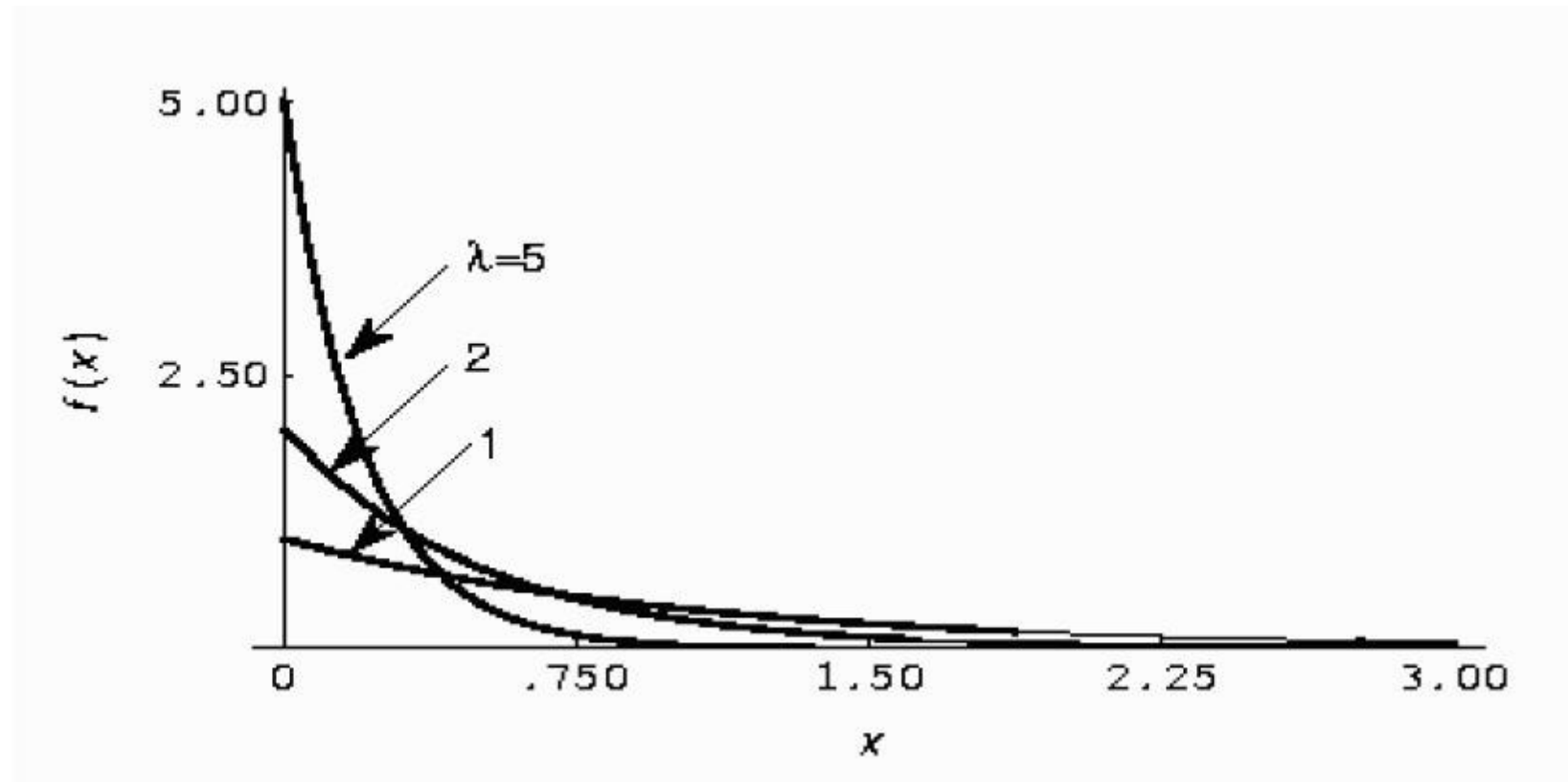
$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

■ Also $P(X > t) = \int_t^{\infty} f(x)dx = e^{-\lambda t}$

and,

$$\begin{aligned} P(a < X \leq b) &= \int_a^b f(x)dx = F(b) - F(a) \\ &= e^{-\lambda a} - e^{-\lambda b} \end{aligned}$$

Variáveis Aleatórias: função densidade da Exponencial



Exponencial: propriedade memoryless



X – v.a. Exponencial, $X > t$

Y – v.a. que representa o tempo que falta para ocorrer o evento de interesse

$$Y = X - t \quad X = Y + t$$

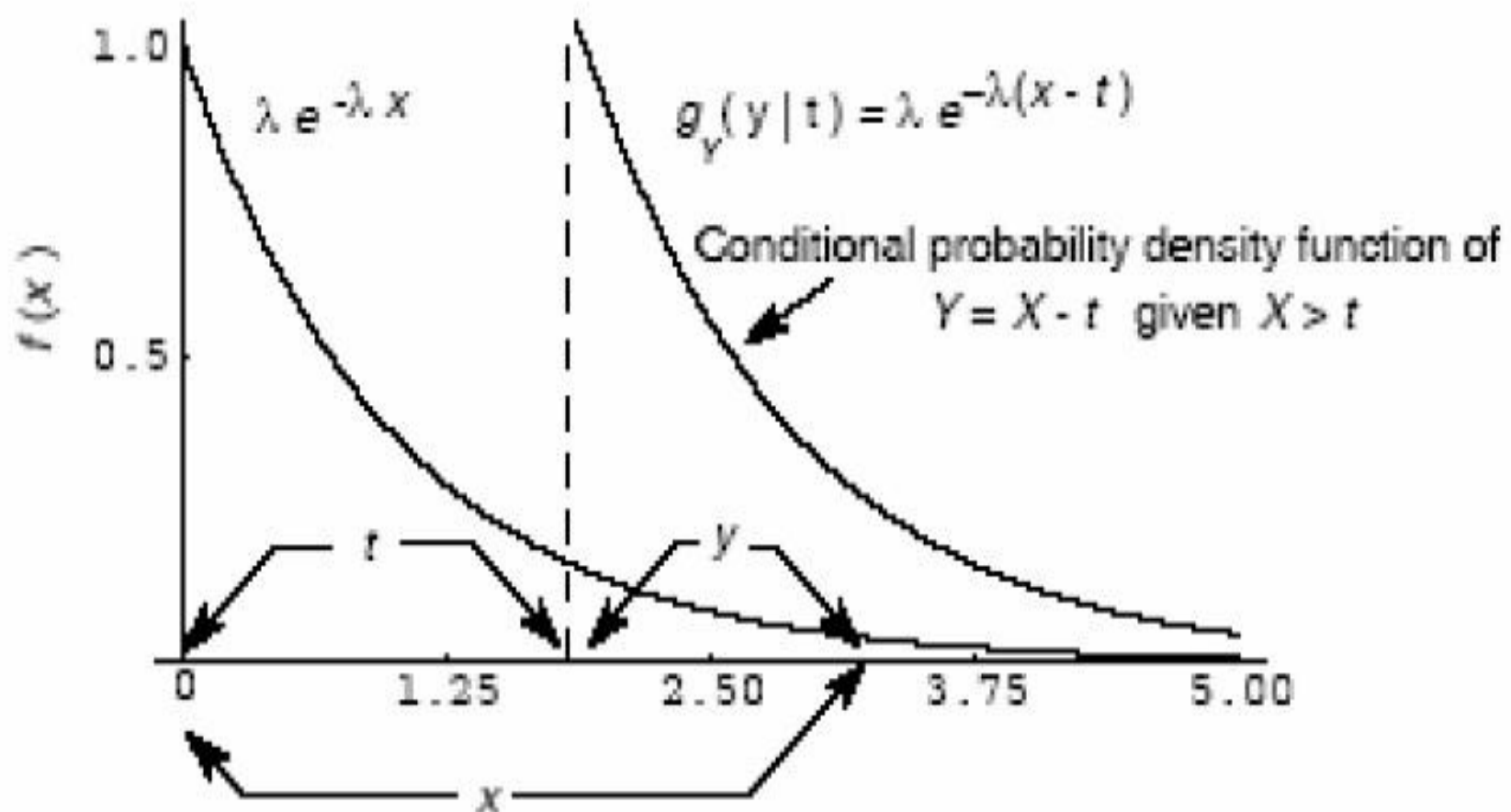
$$P[Y < y \mid X > t] = ?$$

Exponencial: propriedade memoryless

$$\begin{aligned} G_Y(y | t) &= P(Y \leq y | X > t) \\ &= P(X \leq y + t | X > t) \\ &= \frac{P(t < X \leq y + t)}{P(X > t)} = 1 - e^{-\lambda y} \end{aligned}$$

- $G_Y(y/t)$ é independente de t e é igual a distribuição da v.a. X (exponencial)
- A distribuição do tempo que resta para ocorrer um evento não depende de quanto tempo já passou

Exponencial: propriedade memoryless



Exponencial x Poisson



- A discrete rv N_t : number of jobs arriving to a file server in the interval $(0, t]$
- N_t be *Poisson* distributed (parameter = λt)
- X : time to next arrival.

$$\begin{aligned} P(X > t) &= P(N_t = 0) \\ &= \frac{e^{-\lambda t} (\lambda t)^0}{0!} = e^{-\lambda t} \end{aligned}$$

- Therefore,

$$F_X(t) = 1 - e^{-\lambda t}$$

- X is exponentially distributed with parameter λ

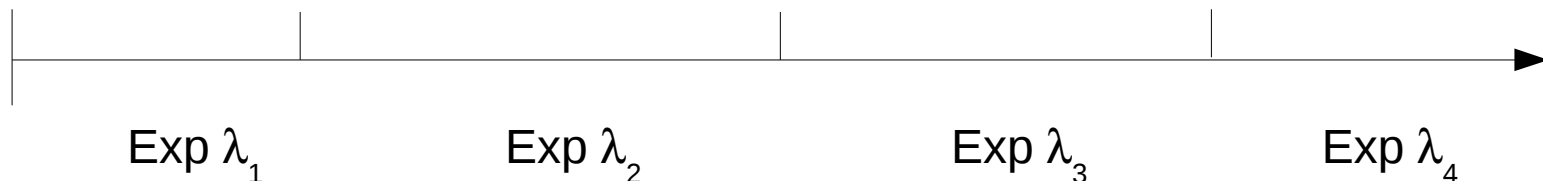
Variável Aleatória Hypoexponencial

- HypoExp: multiple Exp stages in series.
- 2-stage HypoExp denoted as $HYPQ(\lambda_1, \lambda_2)$. The density, distribution

$$f(t) = \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} (e^{-\lambda_1 t} - e^{-\lambda_2 t}), \quad t > 0$$

$$F(t) = 1 - \frac{\lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_2 - \lambda_1} e^{-\lambda_2 t}, \quad t \geq 0$$

- Disk service time may be modeled as a 3-stage Hypoexponential as the overall time is the sum of the seek, the latency and the transfer time.

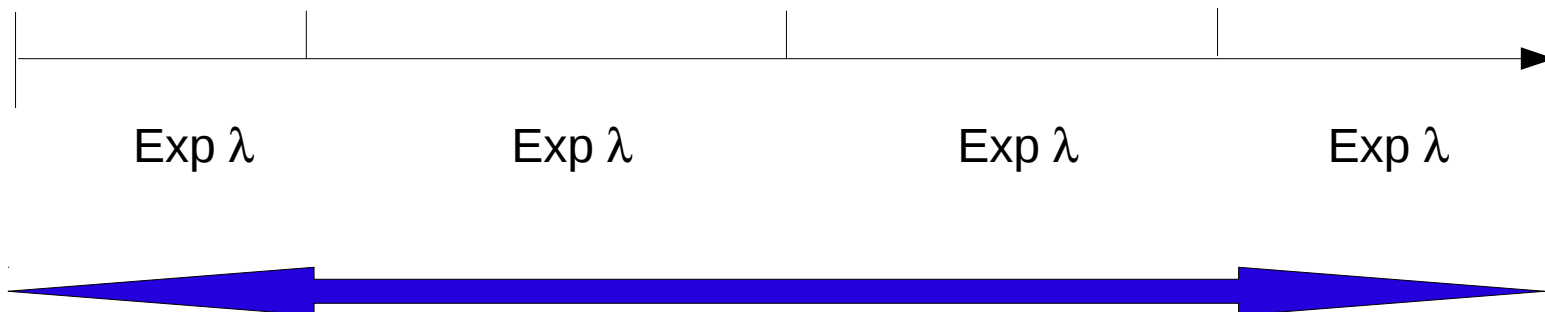


Variável Aleatória Erlang

- Special case of HYPO: All stages have same rate.

$$f(t) = \frac{\lambda^r t^{r-1} e^{-\lambda t}}{(r-1)!}, \quad t > 0, \quad \lambda > 0, \quad r = 1, 2, \dots$$

$$F(t) = 1 - \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad t \geq 0, \quad \lambda > 0, \quad r = 1, 2, \dots$$



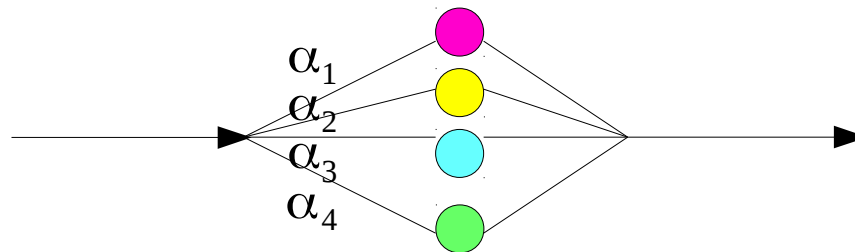
r = número de estágios

Variável Aleatória Hyperexponencial

- Hypo or Erlang have sequential Exp() stages.
- When there are alternate Exp() stages it becomes Hyperexponential.

$$f(t) = \sum_{i=1}^k \alpha_i \lambda_i e^{-\lambda_i t}, \quad t > 0, \quad \lambda_i > 0, \quad \alpha_i > 0, \quad \sum_{i=1}^k \alpha_i = 1$$
$$F(t) = \sum_i \alpha_i (1 - e^{-\lambda_i t}), \quad t \geq 0$$

- CPU service time may be modeled by HyperExp.



Variável Aleatória Gaussiana ou Normal

- A basic distribution of statistics. Many applications arise from **central limit theorem** (average of values of n observations approaches normal distribution, irrespective of form of original distribution under quite general conditions).
- Consequently, appropriate model for many, but not all, physical phenomena.
- Example: Distribution of physical measurements on living organisms, intelligence test scores, product dimensions, average temperatures, and so on.
- Many methods of statistical analysis presume normal distribution.
- One of the first applications: C.F. Gauss used it in 1809 to model observational errors in astronomy

Variável Aleatória Normal x Teorema do Limite Central

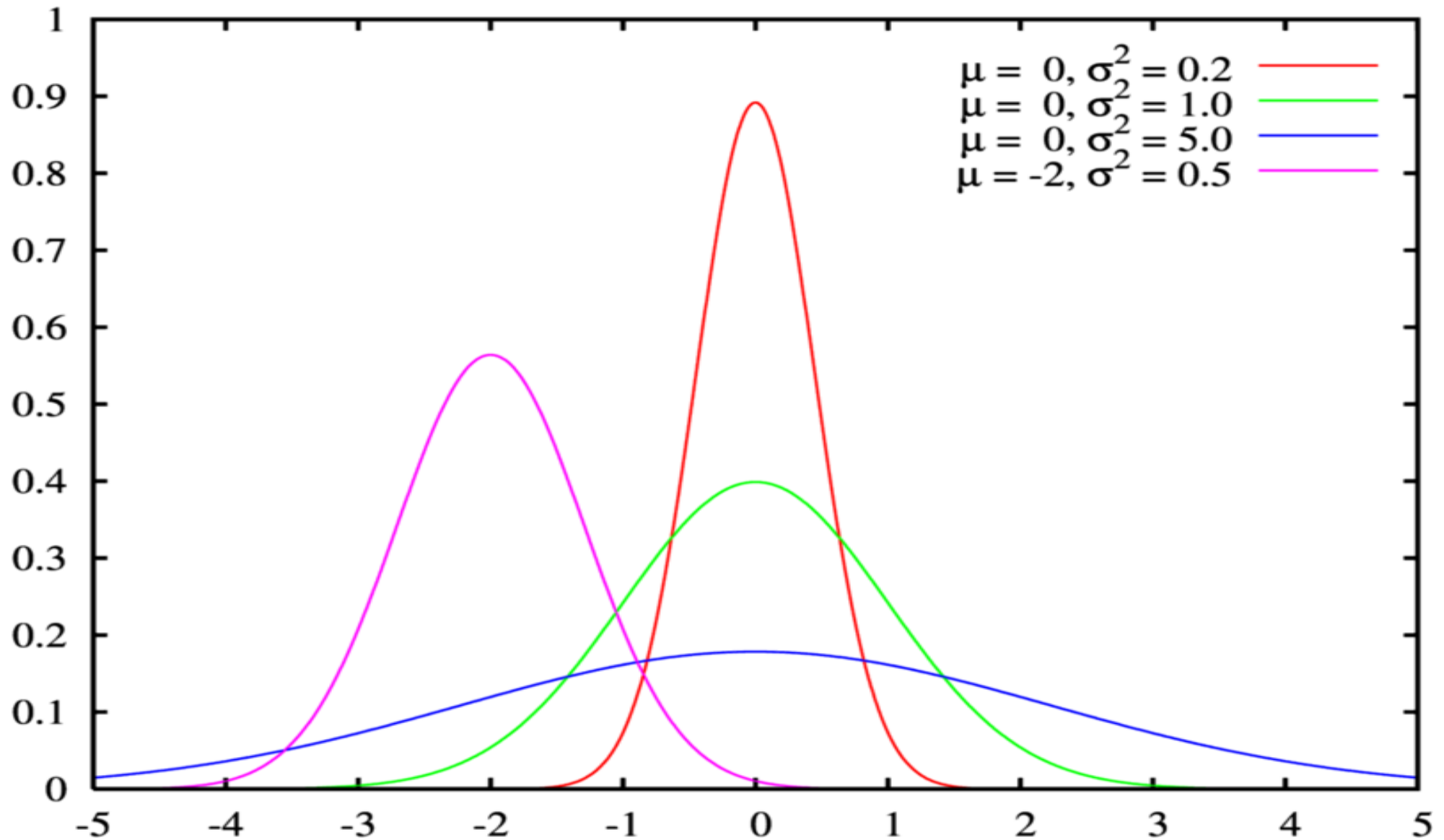
- Seja $(X_1, X_2, \dots, X_n)$ uma sequência de n variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), com média μ e variância $\sigma^2 > 0$, ambas finitas.

Então a variável aleatória

$$Z_n = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}},$$

converge em distribuição para a distribuição normal padrão $N(0,1)$.

Variável Aleatória Normal



Variável Aleatória Normal Padrão

- $N(0,1)$ é uma variável aleatória Normal com média igual 0 e desvio padrão igual 1, é chamada de v.a. Normal padrão
- Não existe fórmula fechada para cálculo da função distribuição da Normal
- Uso de tabelas após transformar uma v.a. Normal em uma Normal padrão
- Seja Z uma v.a. Normal(0,1) e X uma v.a. Normal(μ, σ)

$$F_Z(z) = P(Z \leq z) = P\left(\frac{X - \mu}{\sigma} \leq z\right) = P(X \leq \mu + z \sigma)$$

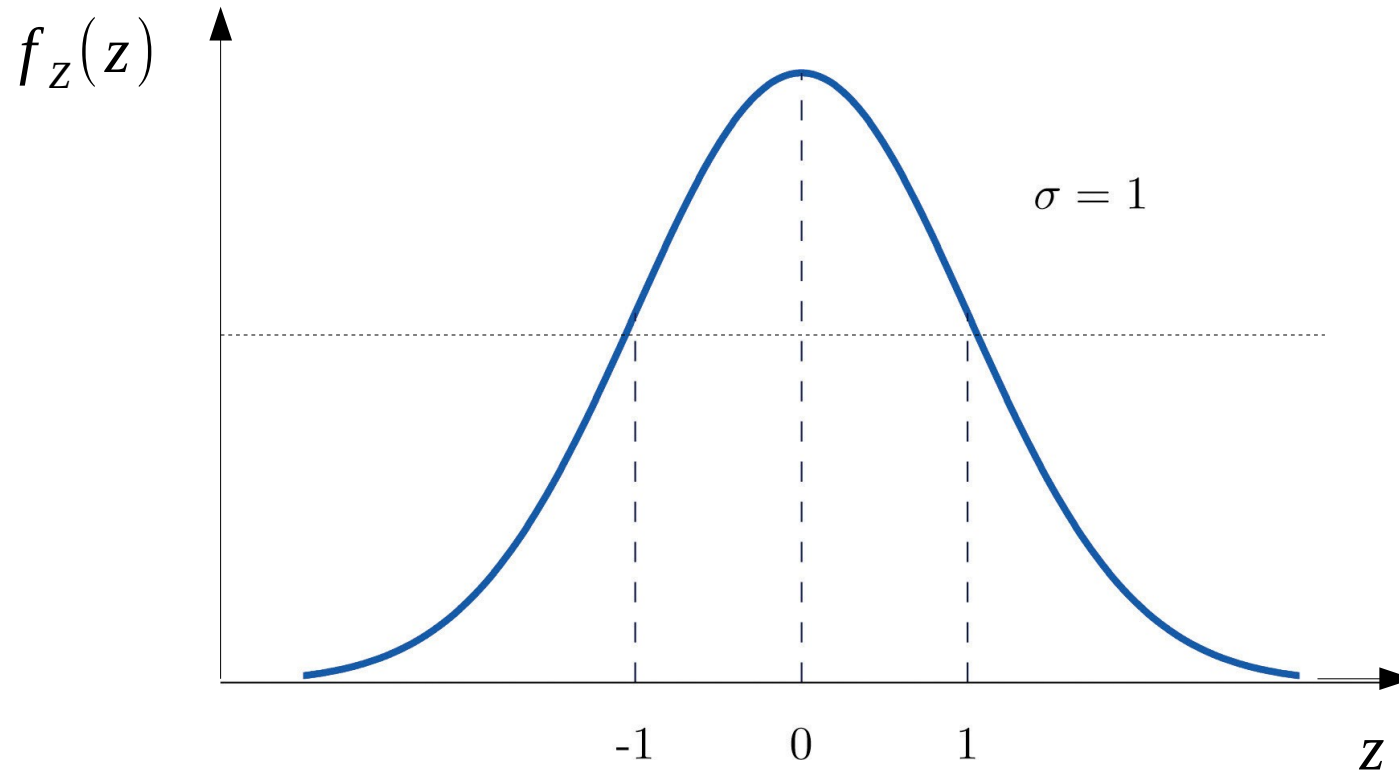
$$\text{Logo, } F_Z(z) = F_X(\mu + z \sigma)$$

$$\text{Da mesma forma, } X = Z \sigma + \mu, P[X \leq x] = P[Z \sigma + \mu \leq x] = P\left[Z \leq \frac{x - \mu}{\sigma}\right]$$

$$F_X(x) = F_Z\left(\frac{x - \mu}{\sigma}\right)$$

Variável Aleatória Normal Padrão

- $N(0,1)$ é simétrica:
$$f_Z(z) = f_Z(-z)$$
$$F_Z(-z) = 1 - F_Z(z)$$



Parâmetros de CDF's

- Uma distribuição pode ser deslocada e reescalada por uma transformação do tipo: $y = (x - \mu) / \sigma$ correspondendo a uma mudança de origem e unidade
- Os parâmetros location e scale são usados para fazer esta transformação (ex: normal)
 - Location: fornece o posicionamento da distribuição com relação ao eixo das abcissas
 - Scale: define a amplitude da distribuição
- Algumas distribuições podem ter a sua forma alterada através do parâmetro shape que define a forma da distribuição (ex: gamma, weibull)

Variável Aleatória Gamma

- Modela a soma de exponenciais onde o número de estágios é contínuo
- Erlang, exponencial e chi-square são casos particulares da distribuição Gamma

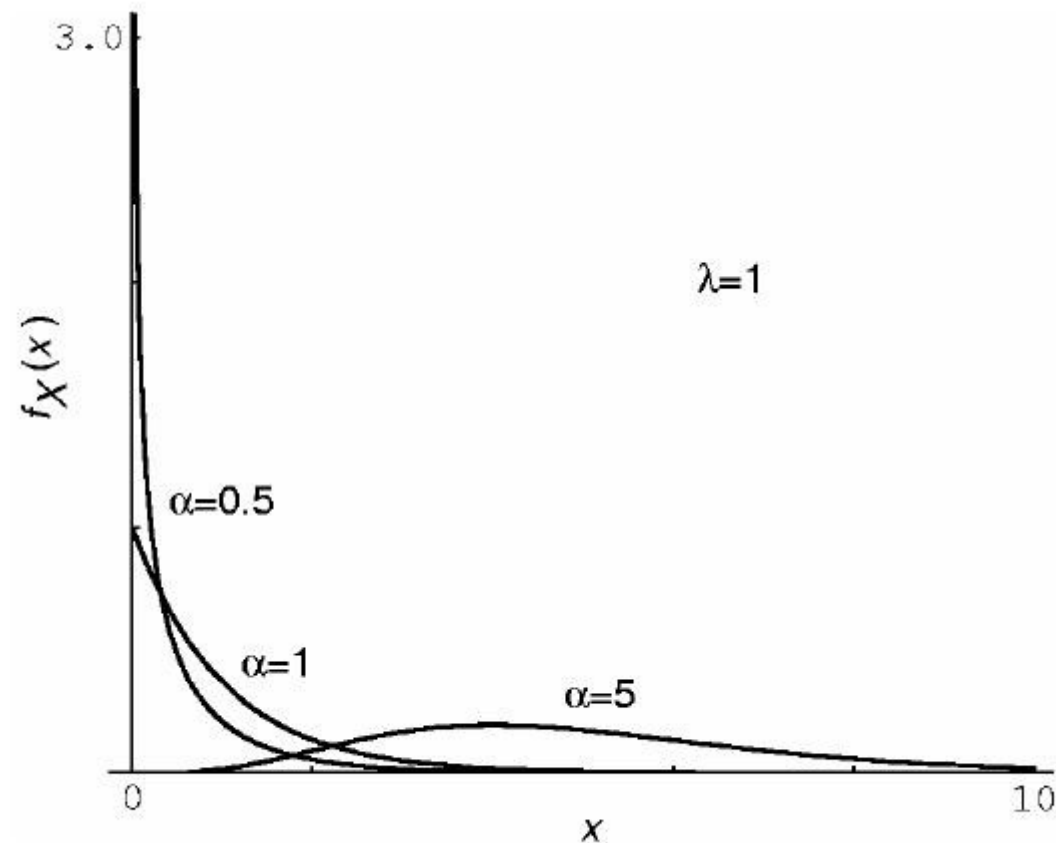
Variável Aleatória Gamma

- Gamma density function is,

$$f(t) = \frac{\lambda^\alpha t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)}, \left(\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \right), \alpha > 0, t > 0$$

- α (shape), λ (scale)

Variável Aleatória Gamma



Variável Aleatória Weibull - 1

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

- É usada para descrever tempo para falha devido a fadiga de um componente
- Também tem sido usada para modelar um usuário acessando a web (tempo em ON e intervalo entre requisições)
- Assume diversas formas variando-se os parâmetros α (shape) e λ (scale)

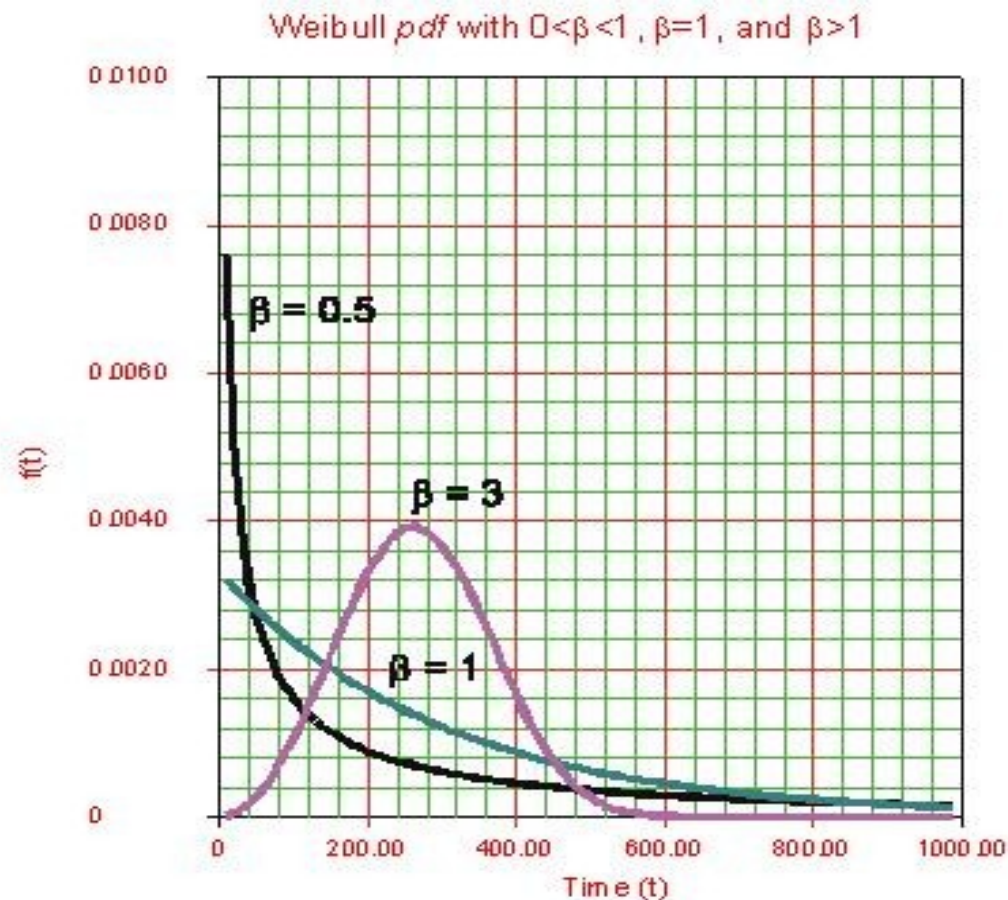
Variável Aleatória Weibull - 2

$$F(t) = 1 - e^{-\lambda(t-\theta)^\alpha}, \quad t \geq \theta$$

- Esta forma possui três parâmetros:
- α (shape), λ (scale) e θ (location)

Variável Aleatória Weibull

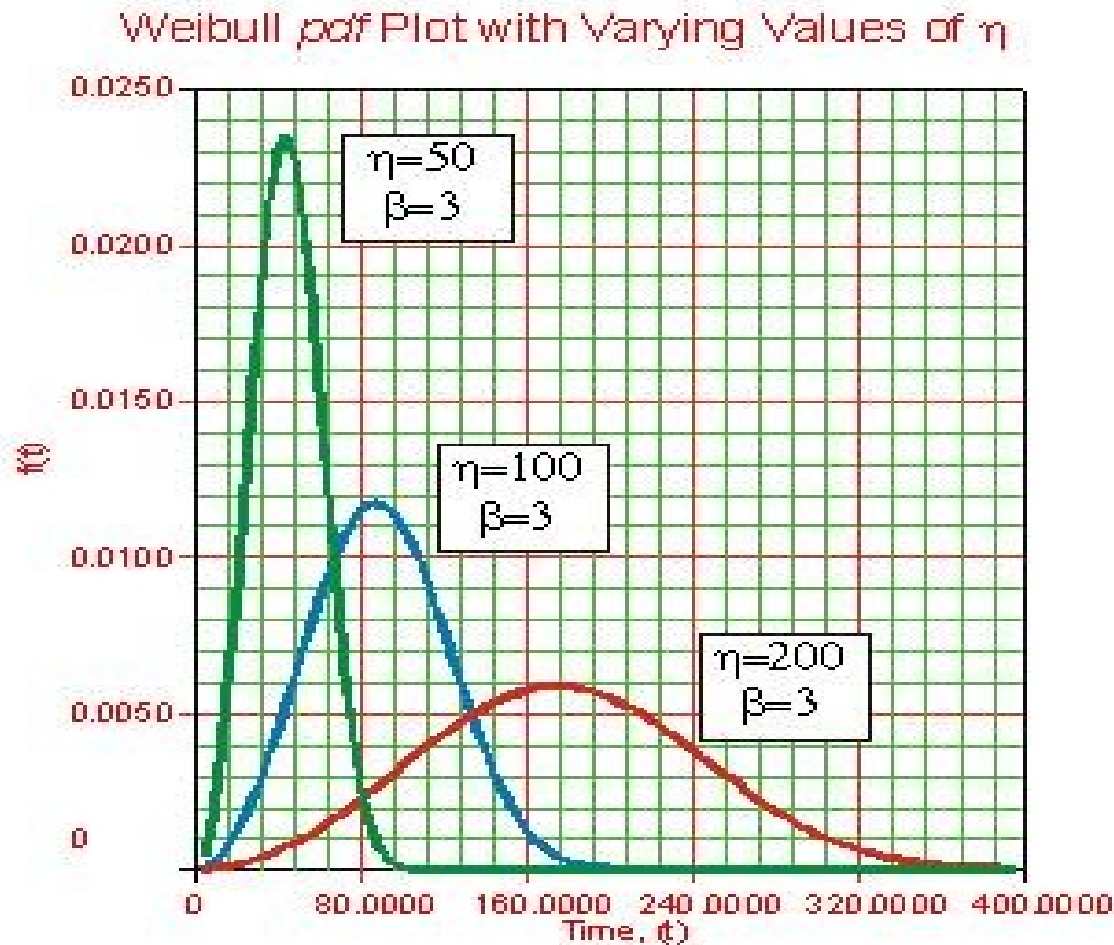
■ Efeito do parâmetro shape



Na figura ao lado β é o parâmetro shape

Variável Aleatória Weibull

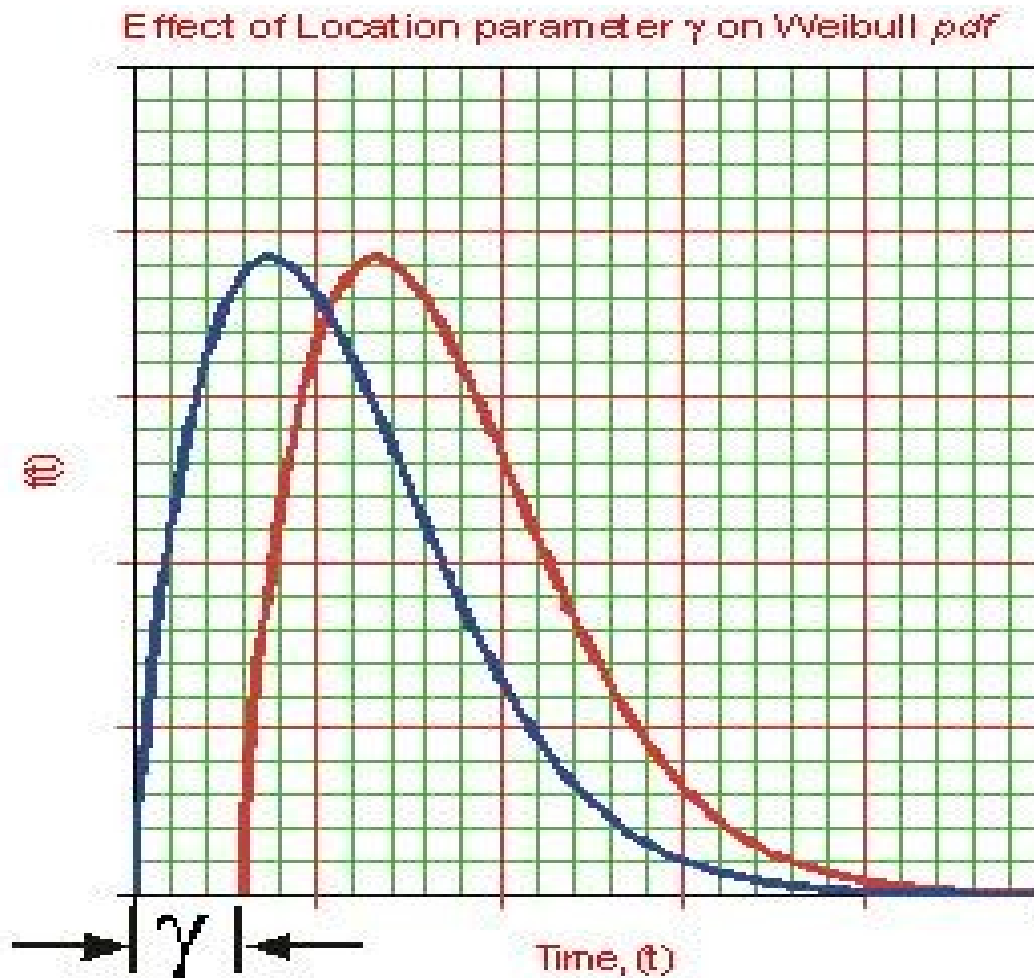
- Efeito do parâmetro scale



Na figura ao lado β é o parâmetro shape e η é o parâmetro scale

Variável Aleatória Weibull

- Efeito do parâmetro location



Na figura ao lado γ
é o parâmetro location

Variável Aleatória Weibull: Exemplo

- Suponha que o tempo de vida de um componente tenha uma distribuição Weibull com parâmetro $\alpha = 2$
- A seguinte observação foi feita: 15% dos componentes falham antes de 100 horas mas duram mais que 90 horas.
- X é a variável aleatória que representa o tempo de vida do componente

$$P[X < 100 | X > 90] = 0.15$$

Variável Aleatória Weibull: Exemplo

- Ache o parâmetro “scale” lambda para essa distribuição

$$P[X < 100 | X > 90] = \frac{P[90 < X < 100]}{P[X > 90]}$$

$$P[X < 100 | X > 90] = \frac{F_X(100) - F_X(90)}{1 - F_X(90)} = \frac{e^{-\lambda(90)^2} - e^{-\lambda(100)^2}}{1 - e^{-\lambda(90)^2}} = 0.15$$

$$\lambda = 0.00008554$$