Capítulo 7

Distribuições de Probabilidades Contínuas

1. DISTRIBUIÇÃO NORMAL

A distribuição normal é a distribuição mais importante do campo da Estatística, uma vez que:

- Serve de parâmetro de comparação;
- Muitas funções convergem para a normal (Poisson, Binomial);
- Muitos fenômenos são descritos pela distribuição normal.

Condições para que uma variável aleatória siga uma distribuição normal:

- Um grande número de fatores influencia a variável aleatória
- Cada fator tem, individualmente, um peso muito pequeno
- Efeito de cada fator é independente dos outros fatores
- Efeito dos fatores no resultado é adicionado.

A f.d. p. da variável aleatória normal X, com média μ e desvio padrão σ é dada por:

$$f(x) = N(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\prod}} e^{\left(\frac{-(X-\mu)^2}{2\sigma^2}\right)} \text{ se} -\infty < x < +\infty$$

A função acumulada ou Função Repartição é dada por :

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

Assim,

$$P(a < x < b) = \frac{1}{\sigma\sqrt{2\Pi}} \int_a^b e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)} dx$$

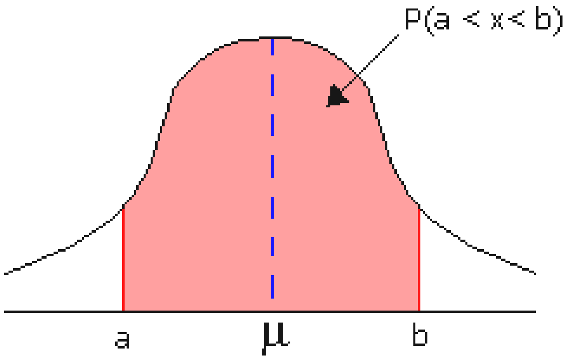


Figura 7.1

Uma vez especificados μ e σ, a curva normal pode ser completamente definida. Os parâmetros μ e σ são também chamados de parâmetros de LOCALIZAÇÃO e ESCALA, pois:

- Para σ = constante

  μ = variando

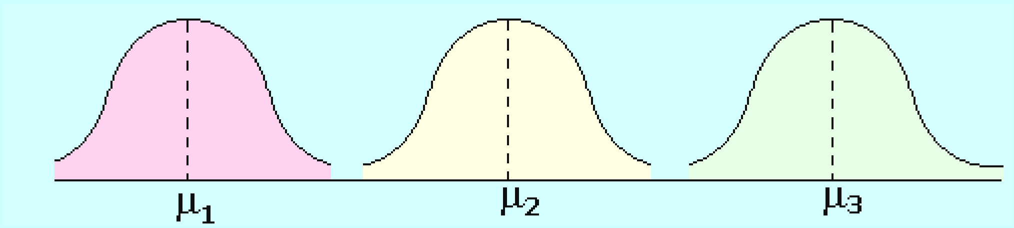


Figura 7.2

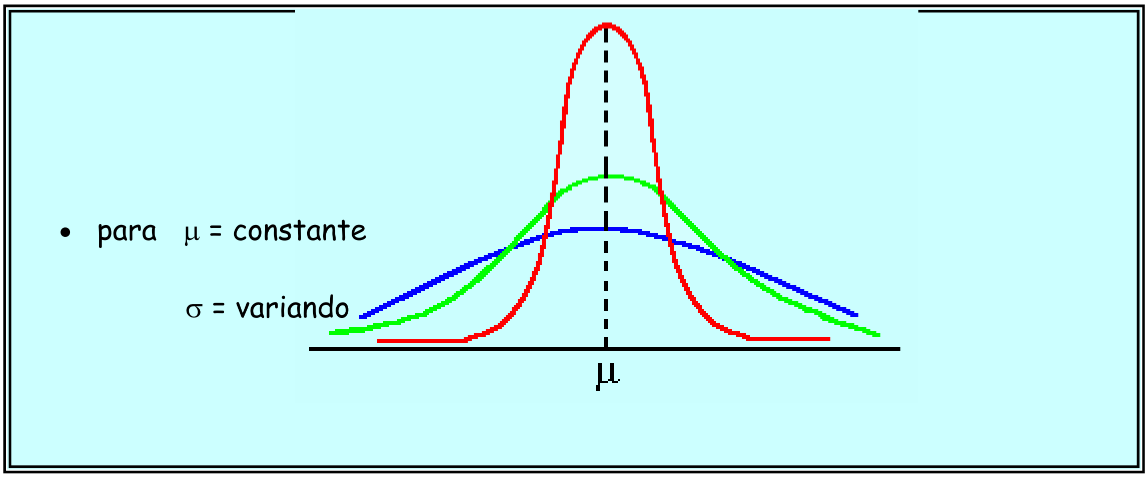


Figura 7.3

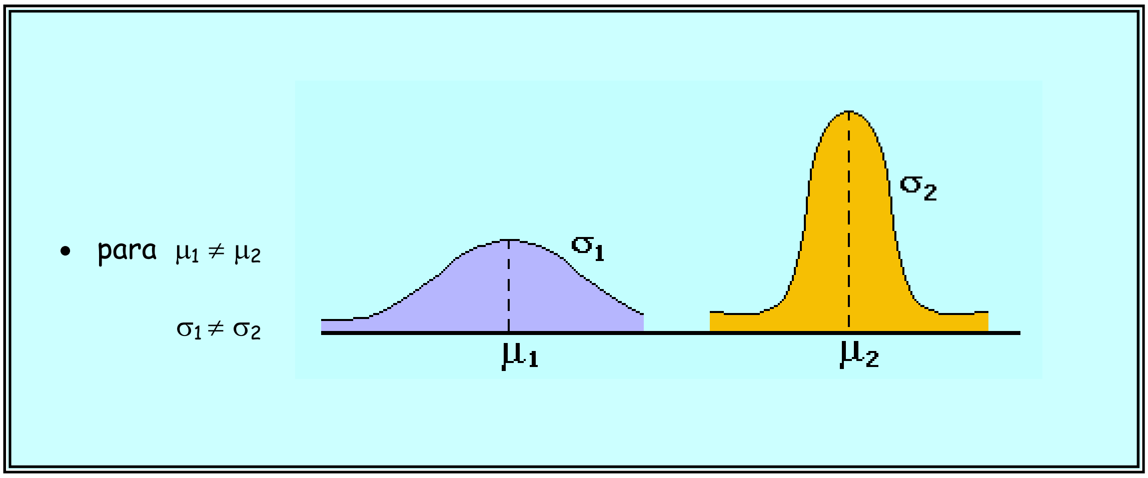


Figura 7.4

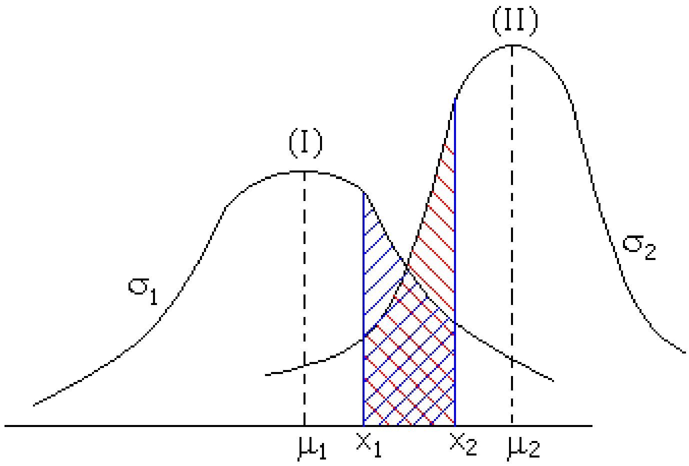


Figura 7.5

Atenção

$P(x_1 < X < x_2)_{I} \neq P(x_1 < X < x_2)_{II}$

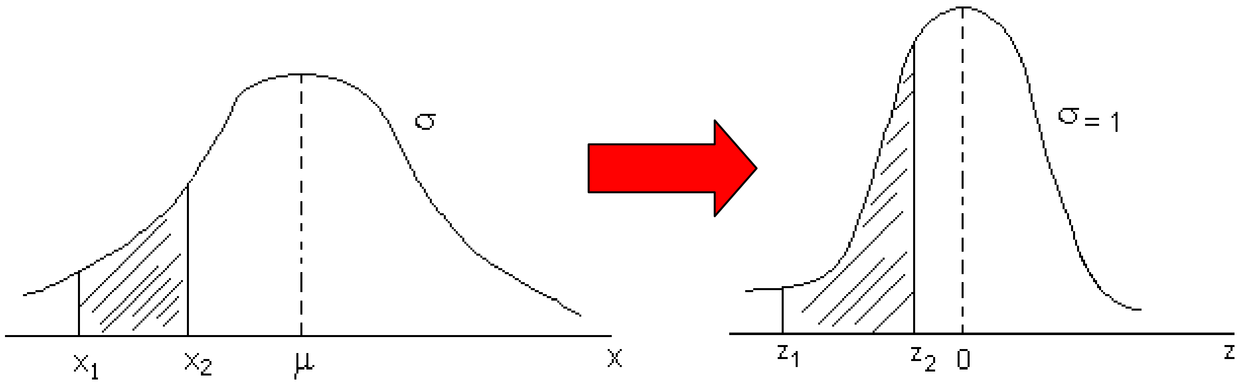
Pois as duas distribuições tem equações diferentes!

Propriedades da Distribuição normal

1. A Moda da distribuição ocorre em $x = \mu$;
2. A curva é simétrica em relação à μ.;
3. A curva tem seus pontos de inflexão em $x = \mu \pm \sigma$;
4. A curva normal se aproxima assintoticamente do eixo - x à medida que se afasta de μ;
5. A área total compreendida entre a curva e o eixo - x é igual a 1 (lógico: é uma f.d.p.!).

Cálculo da probabilidade

Uma vez que o f(x) é função de μ e σ, teremos inúmeras equações para diferentes valores de μ e σ .Para evitar cálculos laboriosos com a integração, criou-se uma tabela única - a da normal padronizada, com μ = 0 e σ = 1.



X ~ N(μ ,6)   z ~ N (1, 0)

TABELADA!

$$P(x_1 < x < x_2) = P(z_1 < z < z_2)$$

Figura 7.6

Transformo x em z !

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

Tabela A3 - Walpole (pags 681 e 682)

Exercício 7.1

1. Dada uma distribuição Normal com μ = 50 e σ = 10, ache a probabilidade da v.a. x assumir valores entre 45 e 62.

2. Em uma prova, a média das notas foi 74 e o desvio padrão 7. Se 12% dos alunos tiraram nota A e as notas seguem uma Distribuição Normal, qual será o menor valor de A e o maior valor de B?

3. Se a média das alturas dos poodles miniatura é 12", com desvio padrão de 1,8", qual a percentagem de poodles que excedem 14" em altura, assumindo que as alturas seguem a distribuição Normal?

4. Se $X \sim N$ (850, 45)
   a) $P(X < 900)$
   b) $P(700 < X < 1000)$
   c) $P(X > 750)$

5. Suponha que, baseado nos dados históricos, o total anual precipitado em uma bacia hidrográfica siga uma normal $X \sim N$ (60", 15").
   a) Qual a probabilidade de que nos anos futuros a precipitação anual fique entre 40 e 70 polegadas ?
   b) Qual a probabilidade da precipitação anual ser inferior a 30" ?

Teorema Central do Limite

O Teorema Central do Limite constitui o fundamento para a estimativa de parâmetros populacionais e para o teste de hipóteses.

Dado:

- A variável aleatória x segue uma distribuição de probabilidades qualquer, com média μ e desvio padrão σ
- Amostras de tamanho n são extraídas aleatoriamente desta população

Resultado:

- Na medida em que o tamanho da amostra aumenta, a distribuição das médias amostrais $\overline{x}$ tende a uma Distribuição Normal
- A média das médias amostrais será a média populacional μ
- O desvio padrão das médias amostrais será $\sigma/\sqrt{n}$

Atenção:

- Para amostras de tamanho $n > 30$, a distribuição das médias amostrais pode ser aproximada satisfatoriamente por uma Distribuição Normal. A aproximação melhora a medida que o tamanho da amostra n aumenta

- Se a distribuição original é Normal, então as médias amostrais seguirão uma Distribuição Normal para qualquer tamanho de n

2. DISTRIBUIÇÃO LOG-NORMAL

Muitos fenômenos hidrológicos exibem uma evidente assimetria, principalmente devido ao fato dos fenômenos naturais – precipitação, vazão, etc – somente assumirem valores maiores que zero, e estes valores poderem assumir teoricamente valores entre 0 e + ∞.

Em tais casos, a variável aleatória "x" não seguirá uma distribuição Normal, mas o seu logaritmo natural sim.

$$\boxed{Y = l_n\, x}$$

Então, $X \sim \Delta\,(\mu_Y, \sigma_Y)$, se $Y \sim N\,(\mu_Y, \sigma_Y)$

Neste caso, sua f.d.p (de y) pode ser facilmente determinada substituindo "y" pelo "x" na equação da Distribuição Normal:

$$\underline{\text{Normal}}: \left\{ \begin{array}{l} f(x) = \dfrac{1}{\sigma\sqrt{2\Pi}} \exp\left[-\dfrac{(x-\mu)^2}{2\sigma^2}\right] \\ \longrightarrow \boxed{f(y) = \dfrac{1}{\sigma_y\sqrt{2\Pi}} \exp\left[-\dfrac{(y-\mu_y)^2}{2\sigma_y^{\,2}}\right]} \end{array} \right.$$

Podemos calcular f(x) (lognormal) através de:

$$f(x) = f(y)\,.\,\frac{dy}{dx}$$

mas, $y = \ln x$

$$\therefore\ \frac{dy}{dx} = \frac{1}{x} \qquad x > 0$$

$$\boxed{f(x) = \frac{f(y)}{x}}$$

Assim:

<u>Normal:</u>

$$\begin{cases} f(x) = \dfrac{1}{\sigma_y x\sqrt{2\Pi}} \quad \exp\left[-\dfrac{(y-\mu_y)^2}{2\sigma^2{}_y}\right] \quad , \quad p/\ x \geq 0 \\ f(x) = \ 0 \qquad\qquad , \text{ caso contrário} \end{cases}$$

A equação anterior mostra que a equação é unilateral ou seja, só toma valores somente no intervalo positivo de y.(*).

Note-se que os parâmetros μ_y e σ_y , que aparecem na f.d.p. são a média e o desvio padrão de y, ou seja, de ln(x) e não de x.

(*) Além disso, a f(x) toma várias formas diferentes para valores distintas de μ_y e σ_y Como se vê na figura abaixo, a f.d.p. é assimétrica a direita, tornando-se a assimetria mais pronunciada a medida que σ_y aumenta Ver gráficos Paulo Miranda

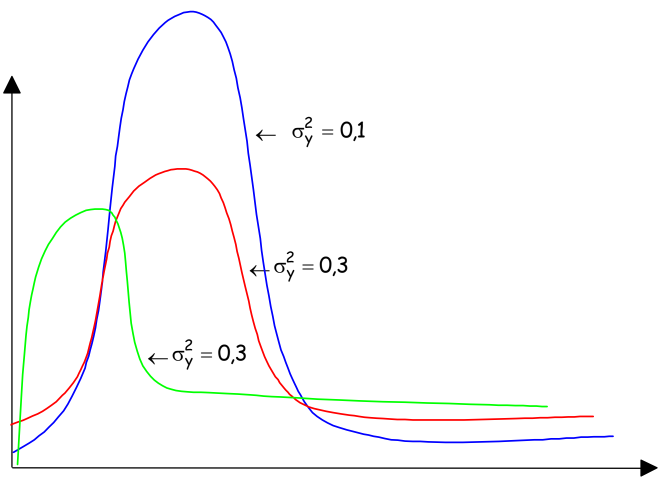


Figura 7.7. Distribuição Lognormal com $\boxed{\mu_y = 0}$ e vários valores de σ_y^2. (SOONG)

Estimativas dos parâmetros da distribuição lognormal.

1. Transformando os X_i''s em Y_i''s e calculando.

$$Y_i = l\,x_i$$

$$\mu_y = \frac{\sum y_i}{n}$$

$$\sigma_y = \sqrt{\frac{\left(\sum y_i^2 - n\,.\,\bar{y}^2\right)}{(n-1)}}$$

2. Relação entre ν_x e μ_x e μ_y e σ_y (sem calcular os logarítmos)

$$\boxed{\bar{x} = \mu_x = \exp\left(\mu_y + \frac{1}{2}\sigma_y^2\right)}$$

$$\nu_x = \frac{\sigma_x}{\mu_x} \qquad \text{ou} \qquad \text{seja} \qquad \nu_x^2 = \frac{\sigma_x^2}{\mu_x^2}$$

mas $\boxed{\nu_x^2 = e^{\sigma_y^2} - 1}$

Exercício 7.2

Um curso d'água tem vazões diárias que se supõe seguirem uma Distribuição Lognormal. Em um período de 30 anos, encontrou-se a vazão média igual a 2.300cfs e distribuição padrão 360cfs. Qual o valor dos parâmetros μ_y e σ_y.

$$\begin{cases} \mu_x = \exp\left(\mu_y + \frac{1}{2}\sigma_y^2\right) \\ \nu_x^2 = e^{\sigma_y^2} - 1 \end{cases}$$

<u>Solução</u>: $\mu_x = 2.300$

$\sigma_x = 360$ mas, $\nu_x^2 = \frac{\sigma x^2}{\mu x^2} = \frac{360^2}{(2300)^2} = 0{,}0245$

Sabemos que:

$$e^{\sigma y^2} = \nu_x^2 + 1 =$$

$$e^{\sigma y^2} = 1{,}0245 \qquad (\ln)$$

$$\boxed{\sigma_y^2 = 0{,}0242}$$

$\boxed{\sigma_y = 0{,}1556}$

mas

$$\ln(2300) = \mu_y + \frac{1}{2}\sigma_y^2$$

$\boxed{\mu_y = 7{,}73}$

b) Qual o valor da cheia milenar ?

c) Qual o valor da cheia decamilenar?

d) Qual a probabilidade dos valores diários ultrapassarem 4.000cfs?

Exercício 7.3 (Haan)

Dados as vazões de pico do Rio Kentucky (USA) entre 1895 a 1960:

Tabela 7.1 - Peak discharges (cfs), Kentucky River, near Salvisa, Kentucky (McCabe, 1962)

1895	47,300	1917	111,000	1939	84,300
96	54,400	18	71,700	40	45,000
97	87,200	19	96,100	41	28,400
98	65,700	20	92,500	42	46,000
99	91,500	21	34,100	43	80,400
1900	53,500	22	69,000	44	55,000
01	67,800	23	73,400	45	72,900
02	70,000	24	99,100	46	71,200
03	66,900	25	79,200	47	46,800
04	34,700	26	62,600	48	84,100
05	58,000	27	93,700	49	61,300
06	47,000	28	68,700	50	87,100
07	66,300	29	80,100	51	70,500
08	80,900	30	32,300	52	77,700
09	80,000	31	43,100	53	44,200
10	52,300	32	77,000	54	20,000
11	58,000	33	53,000	55	85,000
12	67,200	34	70,800	56	82,900
13	115,000	35	89,400	57	88,700
14	46,100	36	62,600	58	60,200
15	52,400	37	112,000	59	40,300
16	94,300	38	44,000	60	50,500

$\mu_x = 67{,}660$

$\sigma_x = 20{,}885$

Supondo que o modelo log.normal seja válido, calcular:

a) P(x > 100.000 cfs)

b) A magnitude da cheia centenária

c) P(x < 50.000 cfs)

Solução

b) magnitude da cheia centenária.

$$\begin{cases} \mu_x = 67.660 \\ \sigma_x = 20.885 \end{cases}$$

Tr = 100 $\rightarrow$ P(X > x) = 0,01

X $\rightarrow$ é só achar P(X < x) = 0,99

$$P(X < x) = P\left(\underbrace{\ln X}_{Y} < \ln x\right) = P\left(Z < \frac{\ln x - \mu_y}{\sigma_y}\right) = 0{,}99$$

$$\boxed{z = 2{,}32}$$

mas $$Z = \frac{\ln x - \mu_y}{\sigma_y}$$

Determinados no problema anterior

$$\begin{cases} \mu_y = 11{,}0737 \\ \sigma_y = 0{,}30395 \end{cases}$$

$\ln x = \boxed{\text{“k”}}$

$x = e^{\text{“k”}}$

x= 130.700 cfs

b) As cheias para Tr = 50 e Tr = 1000 anos.

Exercício 7.5 (profa.Terezinha, pag. 20)

A descarga média anual do Rio Piabinha (Petrópolis) no posto hidrométrico em Areal, durante o período de 1933 – 1948, foi de 14.500 l/s, com desvio padrão de 2.100 l/s. Supondo a validade do modelo lognormal, pede-se a probabilidade de que, em um determinado ano, a descarga x esteja entre 10.700 l/s e 16.900 l/s.

<u>Solução</u>: μ = 14.500 P(10.700 $\leq$ x $\leq$ 16.900)

σ = 2.100

$$\nu_x^2 = \frac{\sigma_x^2}{\mu_x^2} = \frac{(2.100)^2}{(14.500)^2} = 0{,}0209$$

$$\nu_x^2 = e^{\sigma y^2} - 1 \therefore\ e^{\sigma y^2} = \nu_x^2 + 1$$

$$\ln\left(e^{\sigma y^2}\right) = \ln(\nu^2 + 1)$$

$$\sigma_y^2 = 0{,}0207 \qquad \boxed{\sigma_y = 0{,}144}$$

$$\mu_x = \exp(\mu_y + \frac{1}{2}\sigma_y^2)$$

$$\ln \mu_x = \mu_y + \frac{1}{2}\sigma_y^2 \quad \therefore \quad \mu_y = \ln\mu_x - \frac{1}{2}\sigma_y^2$$

$$= 9{,}58 - 0{,}01035$$

$$\boxed{\mu y = 9{,}56}$$

assim,

$= P\ (10.700 \leq X \leq 16.900)$

$= P\ (\ln 10.700 \leq Y \leq \ln 16.900)$

$$= P\left(\frac{\ln(10.700) - 9{,}56}{0{,}144}\right) \leq Z \leq \frac{\ln(16.900) - 9{,}56}{0{,}144}$$

$= P\ (-1{,}96 \leq Z \leq 1{,}21)$

$= P\ (z \leq 1{,}21) + P(z \leq 1{,}86) = 0{,}3869 + 0{,}4750$

$=$ <u>0,8619</u>

3. DISTRIBUIÇÃO DE PROBABILIDADES EXPONENCIAL

(Haan, pg. 79)

O processo de Poisson é um processo discreto numa escala de tempo contínua. Então, a distribuição de probabilidade do **número de eventos num tempo t** é uma distribuição discreta, enquanto que a distribuição de probabilidade para **o tempo entre os eventos** e o **tempo para que o k-ésimo evento ocorra** é uma distribuição contínua.

3.1. DEDUÇÃO DA FÓRMULA (Soong, 195)

Suponha que o evento seja a chegada de barcos em um porto. Seja X um v.a. que representa o n° de chegadas no intervalo de tempo (0, t) e que segue uma distribuição de Poisson ($X \sim p(x; \lambda t)$). Mas o que nos interessa aqui é o <u>tempo entre duas chegadas</u>, que, naturalmente, também é uma v.a. , no caso , <u>contínua</u>.

Denotemos por T esse tempo entre as chegadas. Suponha que chegue pelo menos 1 barco neste tempo t. A FDP da variável aleatória T é, por definição:

$$FDP = P(T \le t) = 1 - \underbrace{P(T > t)}$$

→ Equivale a "não há chegadas no tempo t" , ou seja, $P(0) \to p(0; \lambda t)$.

Sabemos que a distribuição de Poisson é dada por:

$$p(x, \lambda) = \frac{(\lambda t)^x \cdot e^{-\lambda t}}{x!}$$

Assim, $p(0; \lambda t) = \dfrac{(\lambda t)^0 \cdot e^{-\lambda t}}{0!} = e^{-\lambda t}$

Assim, a FDP da exponencial é dada por:

$\boxed{FDP = 1 - e^{-\lambda t}}$, onde λ = taxa média de chegadas

Se a FDP é dada por $1 - e^{-\lambda t}$, a função densidade de probabilidade (fdp) da Exponencial é dada por:

Sabemos que, por definição:

$$\boxed{FDP(x) = \int_{-\infty}^{x} f(x).dx}$$

No caso da Exponencial, a variável aleatória é "T", o tempo entre chegadas (não existe t < 0).

Assim:

$$\boxed{FDP(T) = \int_0^t f(t)dt \ = 1 - e^{-\lambda t}}$$

Se a $FDP = \int_0^t f(t)dt,$ então $f(t) = \frac{d}{dt}(FDP)$

Assim,

$$f(t) = \frac{d}{dt}\left(1 - e^{-\lambda t}\right) = \frac{d(1)}{dt} - \frac{d}{dt}\left(e^{-\lambda t}\right) = 0 - e^{-\lambda t}(-\lambda) = \lambda e^{-\lambda t}$$

$$\boxed{f(t) = \lambda e^{-\lambda t}}$$

Concluindo,

A função densidade de probabilidade (fdp) e a função distribuição de probabilidades (FDP) da Exponencial são dadas por:

$f(t) = \lambda e^{-\lambda t}$ $\qquad t \leq 0$

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

Sendo a média e a variância dadas por:

$$\boxed{\mu = \frac{1}{\lambda}} \qquad \boxed{\sigma^2 = \frac{1}{\lambda^2}}$$

Exercício 7.6 (ABRH, vol.4, pág.130)

Se ocorrem 3 chuvas catastróficas com duração de 1 hora a cada 10 anos, qual a probabilidade de que leve menos de 1 ano até a próxima ocorrência?

Solução: λ = 3/10 = 0,3 chuva/ano

t = 1 ano

Usando a FDP diretamente:

$P(T < 1) = 1 - e^{-\lambda t}$

para $t = 1$

$\lambda = 0{,}3$

$P(T < 1) = 1 - e^{-0{,}3} = 0{,}26$

(Ver mais em Haan, pág. 78)

ATENÇÃO: Possível pergunta de um aluno

Então, em um estudo de duração de secas, eu não preciso dos dados individuais (durações de cada seca) ? Somente o nº de secas e o nº total de anos observados?

Depende:

Se eu tiver certeza que os dados se ajustam a uma Exponencial, basta ter o λ (nº de ocorrências / nº de anos observados) - parâmetro da Poisson. O "λ", é na verdade, o inverso da média das durações das secas (média da exponencial = nº de anos obs. / nº de ocorrências).

λ=3 ocorrências/ 10 anos

Média da exponencial = 1/λ= 10 anos/ 3 ocorrências = 3,3

Se eu não tiver certeza que os dados se ajustam à uma Exponencial, eu precisarei dos dados históricos – duração de cada seca – testo a aderência pelo Teste do Chi-Quadrado (χ^2). Se a Exponencial passar no teste, basta calcular o parâmetro λ (inverso da "duração média").

Relação Exponencial x Poisson

Poisson

3 chuvas catastróficas em 10 anos

Qual a probabilidade de nenhuma chuva catastrófica em 30 anos ?

<u>Solução</u>:

λ = 3 chuvas/10 anos

t = 30 anos

x = 0

Assim, λt= 9 chuvas

$$p(x;\lambda t)=\frac{(\lambda t)^x . e^{-\lambda t}}{0!}=\frac{9^0 .\ e^{-9}}{0!}=1\ \text{x}\ 0{,}0001$$

=0,0001

Explicação: o nº esperado de chuvas catastróficas em 30 anos seria de 9 chuvas.

Por isso a probabilidade de não ocorrer nenhuma deu tão pequena!

Exponencial

3 chuvas catastróficas em 10 anos

Qual a probabilidade de passar mais de 30 anos para que ocorra a próxima chuva catastrófica ?

<u>Solução</u>: $P(t>30)=1-P(t\le 30)$ Por definição = $1-e^{-\lambda t}$

$$P(t>30)=1-\left(1-e^{-\lambda t}\right)=e^{-\lambda t}$$

$$P(t>30)=e^{-9}=0{,}0001$$

$$\begin{cases}\lambda=\dfrac{3}{10}\\ t=30\end{cases}\qquad \lambda t=9$$

Explicação: Espera-se que a cada 3,33 anos ocorra uma chuva catastrófica. Por isso a probabilidade de que se passe mais de 30 anos para ocorrer esta chuva é tão pequena.

Exercício 7.8 (Haan, 99 – Ex.6.2)

Haan e Jonhson estudaram as características físicas de depressões no setor centro-oeste de Iowa. Achar a probabilidade de uma depressão ter área maior que 2,25 acres (supor seguir uma exponencial).

Área (acres)	nº de dep.
0 ⊢ 1/2	106
1/2 ⊢ 1	36
1 ⊢ 1/2	18
$1^{1/2}$ ⊢ 2	9
2 ⊢ $2^{1/2}$	12
$2^{1/2}$ ⊢ 3	2
3 ⊢ $3^{1/2}$	5
$3^{1/2}$ ⊢ 4	1
4 ⊢ $4^{1/2}$	4
$4^{1/2}$ ⊢ 5	5
5 ⊢ $5^{1/2}$	2
$5^{1/2}$ ⊢ 6	6
6 ⊢ $6^{1/2}$	3
$6^{1/2}$ ⊢ 7	1
7 ⊢ $7^{1/5}$	1
$7^{1/5}$ ⊢ 8	1
	212

$$\bar{x} = \frac{270{,}5}{212} = 1{,}27$$

$$\bar{x} = \frac{\sum x_i \cdot F_i}{n}$$

(média da Exponencial

Sabemos que: $\boxed{\text{média} = 1/\lambda}$

$\lambda = 0{,}7837 \rightarrow$parâmetros de Poisson

$P(x > 2{,}25) = 1 - P(x \leq 2{,}25) = 1 - (1 - e^{-(0{,}7837)\cdot 2{,}25}) = 0{,}171$

mas $P(T \leq t) = 1 - e^{-\lambda t}$

Exercício 7.9 (Ang and Tang, pág, 121, ex.3.18)

Dados históricos de terremotos em San Francisco, Ca, mostram que no período de 1836 – 1916, ocorreram 16 terremotos de grande intensidade. Se a ocorrência de terremotos desta intensidade segue uma distribuição de Poisson, qual a probabilidade de ocorrerem terremotos nos próximos dois anos ?

Solução: $\lambda = 16/125 = 0{,}128$ terremotos por ano

$$P(T \leq 2) = 1 - e^{--\lambda t} = 1 - e^{-(0,128)\,.2} = 0{,}226$$

b) E a probabilidade de nenhum terremoto desta intensidade ocorrer nos próximos 10 anos.

Solução: $P(T > 10) = 1 - \underbrace{P(T \leq 10)}_{1 - e^{-\lambda t}} = e^{--\lambda t} = e^{-(0,128)\,.\,10} = 0{,}278$

Exercício 7.10 (Bedient e Huber, 196 – Hydrology e Floodplain Analysis)

Durante um ano, cerca de 110 episódios independentes de tempestade ocorreram em Gainsville, Flórida. Sua duração média foi de 5,3 horas. Considerando um ano com 8760 horas, o tempo médio entre tempestades foi de:

Tempo c/ tempestade = 110 x 5,3 = 583 hs

Tempo total $= \dfrac{8760 - 583}{110} = 74{,}3\,\text{horas}$ (entre tempestades)

(suponha a validade do modelo exponencial)

a) Qual a probabilidade de que no mínimo se passe 4 dias entre tempestades?

Solução: Sabemos que: $P(T \leq t) = 1 - e^{-\lambda t}$

no mínimo 4 = E(4) + E(5) + E(6) ...

$$= P(T \geq 4) = 1 - \underbrace{P(T < 4)}_{(1 - e^{-\lambda t})}$$

$$= P(T \geq 4) = e^{-\lambda t}$$

Sabemos que:

exp. $\rightarrow$ média = $1/\lambda$ $\therefore$ $\lambda = 0{,}0135$

t = 4 dias x 24hs = 96hs

$= P(T \geq 4) = e^{-0,0135 \cdot 96}$

$= P(T \geq 4) = 0,2747$

b) Qual a probabilidade do intervalo entre tempestade ser de exatamente 12 horas?

$\int_{12}^{12} f(t)\,.\,dt = 0$

c) Qual a probabilidade do intervalo entre tempestade seja menor ou igual a 12 horas?

$P(T \leq 12) = 1 - e^{-\lambda t} = 1 - e^{-0,0135 \cdot 12} = 0,1496$

4. DISTRIBUIÇÃO DE PROBABILIDADES GAMA

Seja X uma variável aleatória que representa o número de chegadas no intervalo de tempo (0, t) e que segue uma distribuição de Poisson. Suponha que o tempo a v-ésima chegada seja dada pela f.d.p.

$$\boxed{f(t) = \lambda^{\nu}\,.\,t^{(\nu-1)}\,.\,\frac{e^{-\lambda t}}{(\nu-1)!}}$$

para valores inteiros de ν.

Veja que quando $\nu = 1$....

$$\boxed{f(t) = \lambda\, e^{-\lambda t}} \rightarrow \text{Exponencial}$$

Ou seja, a Distribuição Exponencial é um caso particular da Distribuição Gama.

Exercício 7.11 Haan, 79

Os barcos chegam a uma eclusa numa média de 4 a cada hora. Se a eclusa só pode operar 4 barcos de cada vez, qual a probabilidade do primeiro barco ter que esperar pelo menos 1 hora antes de ser embarcado?

Solução: É a mesma coisa de : Qual a probabilidade do 4° barco demorar mais de 1 hora (após o 1° ter chegado) ?

$\boxed{1}$ $\boxed{1}$ $\boxed{2}$ $\boxed{3}$

t = 0

Se o tempo começa a contar a partir da chegada do 1° barco, eu quero saber qual a probabilidade do terceiro barco que falta demorar mais que uma hora.

$P(T_3 > 1) = 1 - P(T_3 \leq 1)$

$$= 1 - \int_0^1 \frac{\lambda^\nu \cdot t^{(\nu-1)} \cdot e^{-\lambda t}}{(\nu - 1)!} \, dt$$

Sabemos que: λ=4 e ν=3

$$= 1 - \boxed{\int_0^1 \frac{4^3 \cdot t^2 \cdot e^{-4t}}{2} \cdot dt} \quad = \quad 1 - \frac{4^3}{2} \underbrace{\int_0^1 t^2 \cdot e^{-4t} \cdot dt}$$

(*) Pode ser resolvida por integração por partes

(*) $\int u dv = uv - \int v du$

Se: $\begin{cases} u = t^2 & du = 2t \, dt \\ dv = e^{-4t} \, dt \dfrac{(-4)}{(-4)} & v = e^{-4t} \end{cases}$ $\quad d(e^u) = e^u \, du$

$$= t^2 \cdot e^{-4t} \Big|_0^1 - \int_0^1 e^{-4t} \cdot 2t \, dt$$

(resolver!) . . . = 0,762

Assim, $P(T_3 > 1) = 1 - 0{,}762 =$ **0,238** → Distribuição Acumulada Inversa

Pode-se ainda usar a Tabela (1 – FDP) da Gama (Haan, 342 – 347)

Entrar na tabela com: $\begin{cases} \chi^2 = 2\lambda t \\ \nu = 2v \end{cases}$

No problema anterior

$$\begin{cases} \lambda = 4 \\ v = 3 \\ t = 1 \end{cases}$$

Então,

$$\left.\begin{array}{l} \chi^2 = 2 \cdot 4 \cdot 1 = 8 \\ \nu = 2 \cdot 3 = 6 \end{array}\right\} \quad \text{tabela} = \underline{0{,}23810}$$

Diretamente ! Pois a tabela da P(T > t)

Comparação entre diversas distribuições (ver Paul Meyer, pag 233):

1. Admita-se que provas independentes de Bernoulli estão sendo realizadas:

a) variável aleatória = número de ocorrências do evento A em um número fixo de provas;

Distribuição: Binomial

b) variável aleatória = número de provas de Bernoulli necessárias para obter a primeira ocorrência do evento A;

Distribuição: **Geométrica**

c) variável aleatória = número de provas de Bernoulli para se obter a r-ésima ocorrência do evento A.

Distribuição: Binomial Negativa (ou Pascal)

2. Admita-se um processo de Poisson

- número médio de sucessos num dado intervalo de tempo – conhecido
- p de 1 sucesso ~ comprimento do intervalo de tempo, etc.
- número de falhas impossível saber!

a) variável aleatória = número de ocorrências do evento A, durante um intervalo de tempo fixo.

Distribuição: Poisson

b) variável aleatória = tempo necessário até a primeira ocorrência do evento A;

Distribuição: Exponencial

c) variável aleatória = tempo necessário até a v-ésima ocorrência de A;

Distribuição: Gama

4.1. Aplicações da Distribuição Gama na Hidrologia (ver ABRH, vol 4 pag.134)

A Distribuição Gama tem grande aplicação na Hidrologia, devido a aspectos de natureza morfológica unicamente. Nesse caso, o valor de "ν" poderá não ser inteiro e o fatorial (v – 1)! deve ser computado pela Função Gama, que dá seu nome a distribuição.

$$\text{Função Gama}: \quad \Gamma(v) = \int_0^\infty x^{v-1}.e^{x}.dx$$

Se "v" for inteiro positivo, $\Gamma(v) = (v - 1)!$

Para v não inteiro, $\Gamma(v)$ está tabelado (Haan, 351)

Então,

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

Atenção:

Normalmente, a Distribuição Gama é apresentada sob a seguinte estrutura matemática:

$$\begin{cases} t = x \\ \lambda = \beta \\ \nu = \alpha \end{cases}$$

Assim,

$$\boxed{f(x) = \frac{\beta^{\alpha} . x^{(\alpha-1)} . e^{-\beta x}}{\Gamma(\alpha)}}$$

$$\boxed{\text{média} = \frac{\alpha}{\beta}} \qquad \boxed{\text{variância} = \frac{\alpha}{\beta^2}}$$

(Semelhante ao Haan)

Atenção:

O Walpole considera:

$$\begin{cases} t = x \\ \lambda = \dfrac{1}{\beta} \\ \nu = \alpha \end{cases}$$

Assim,

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

$$\boxed{f(x) = \frac{\beta^{\alpha} . x^{(\alpha-1)} . e^{-x/\beta}}{\beta^{\alpha} . \ \Gamma(\alpha)}}$$

$$\boxed{\text{média} = \alpha\beta} \qquad \boxed{\text{variância} = \alpha\beta^2}$$

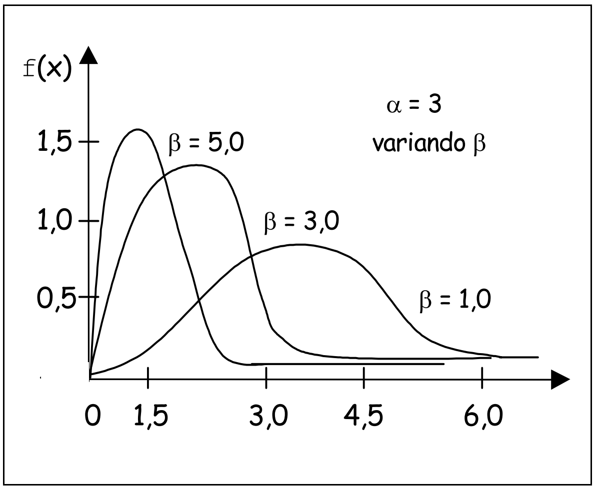
Semelhante ao Walpole
$\alpha > 0, \ \beta > 0$

4.2. OS PARÂMETROS DA DISTRIBUIÇÃO GAMA

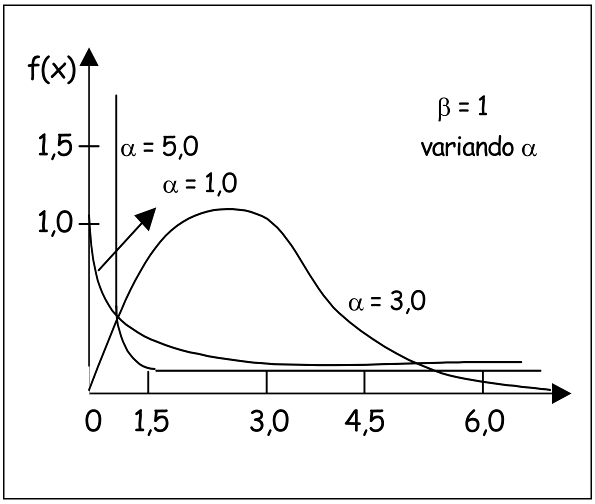
(Metodologia do Haan. Ver também Soong, 193)

Os parâmetros associados à Distribuição Gama são α e β e supõem-se ambos positivos. Como a distribuição Gama é unilateral, serve freqüentemente de modelo para quantidades físicas que só tomam valores positivos. Além disto, graças à sua versatilidade, torna-se um modelo útil, já que variando os valores de "α" e "β", podemos obter uma ampla variedade de formas da f.d.p Gama.

Variando α:



Variando β:



Podemos observar que "α" determina a forma da distribuição, sendo portanto um parâmetro de forma; enquanto "β" é o parâmetro de escala da distribuição.

Em geral a f.d.p. é unimodal com pico em x = 0 para $\alpha \le 1$ (caso da exponencial e da "J-shaped") é em $x = \frac{(\alpha - 1)}{\beta}$ para $\alpha > 1$.

(OBS: Ver também comentário em Yevjevich, pág 147 sobre Lognormal e Gama)

Exercício 7.12 (ABRH vol. 14, pág. 135)

Dados os valores máximos anuais de vazões no rio Mãe Luzia em Forquilhinha, em m^3/s.

$$\bar{x} = 311,,35 \, m^3/s$$

$$\sigma = 169,6 \; m^3/s$$

a) Estimar os parâmetros da distribuição da distribuição Gama

(Método dos Momentos)

$$\mu = \frac{\alpha}{\beta} \longrightarrow \alpha = 311,35\,\beta$$

$$\sigma^2 = \frac{\alpha}{\beta^2} \longrightarrow (169,6)^2 = \; = \frac{311,35}{(169,6)^2} = 0,01$$

$$\begin{cases} \beta = 0,01 \\ \alpha = 3,37 \end{cases}$$

b) Calcular o período de retorno para a vazão máxima observada ($Q = 880m^3/s$)

<u>Solução</u>: Por definição $P(Q > 880) = \dfrac{1}{Tr}$

(Na Tabela Haan, pág. 342)

$v = 2\alpha = 2 \times 3,37 = 6,74$

$\chi^2 = 2\beta x = 2 \times \cdot 0,01 \times 880 = 17,6$

$P(X > x) = 0,01444$

$\therefore \; Tr = \dfrac{1}{0,01444} = 69,25 \qquad Tr \sim 70$ anos

5. DISTRIBUIÇÃO DE PROBABILIDADE CHI-QUADRADO

Um outro importante caso particular da distribuição Gamma é obtido fazendo $\alpha = \dfrac{\nu}{2}$ e $\beta = \dfrac{1}{2}$, onde ν é um inteiro positivo. A distribuição tem um único parâmetro – "ν" - chamado de "graus de liberdade".

$$f(x) = \frac{x^{\left(\frac{\nu}{2}-1\right)} \cdot e^{-x/2}}{2^{\left(\frac{\nu}{2}\right)} \cdot \Gamma(\nu/2)} \qquad , \; x > 0$$

$$\text{média} = \nu \qquad\qquad \text{variância} = 2\nu$$

A distribuição x^2 é um dos principais instrumentos na área da inferência estatística e teste de hipóteses.

6. LOG-NORMAL A 3 PARÂMETROS

(ver Yevjevich, pág.138)

Quando o limite inferior da distribuição não é zero, é necessário modificar a função distribuição de probabilidade da log-normal, introduzindo um terceiro parâmetro – o **limite inferior**.

A f.d.p. da distribuição log-normal a 3 parâmetros é:

$$f(x) = \frac{1}{(x-\delta)\ \sigma_y\ \sqrt{2\pi}} \cdot \exp\left[-\frac{(\ln(x-\delta)-\mu_y)^2}{2\sigma_y^{\ 2}}\right]$$

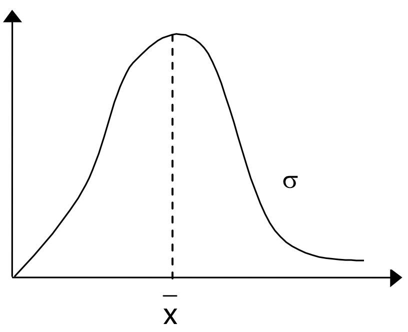
Esta equação é a mesma da Log-nomal, apenas com a substituição de **x** por $(x - \delta)$:

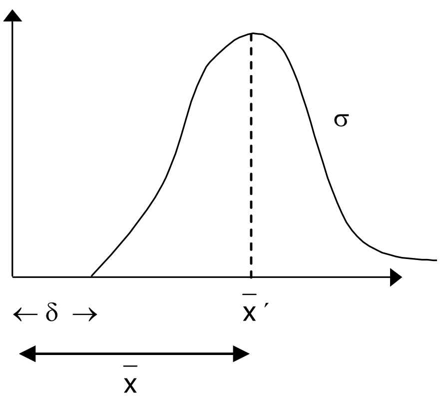
$$f(x) = \frac{1}{x\ \sigma_y\ \sqrt{2\pi}} \cdot \exp\left[-\frac{(\ln(x)-\mu_y)^2}{2\sigma_y^{\ 2}}\right]$$

A adição (ou subtração) de uma constante de uma variável não muda sua variância, mas muda sua média. Assim, a mudança de μ_x implica na modificação de seu coeficiente de variação ν_x. Aqueles parâmetros que dependem de ν_x, como σ_y também mudam, uma vez que:

$$\boxed{\upsilon_x^2 = e^{\sigma_y^2} - 1}$$

A maneira mais fácil de expressar os parâmetros δ, μ_y e σ_y é usando os momentos de **x**; onde o de primeira ordem define μ_x e o de segunda ordem, σ_x . Como já foi dito anteriormente, μ_x e σ_x são diferentes dos parâmetros correspondentes à função lognormal a 2 parâmetros.





Onde x= x´ + δ

Exercício 7.13 (Lista Terezinha pág. 13, ex. 7)

X ~ Λ (μ_y, σ_y, δ)

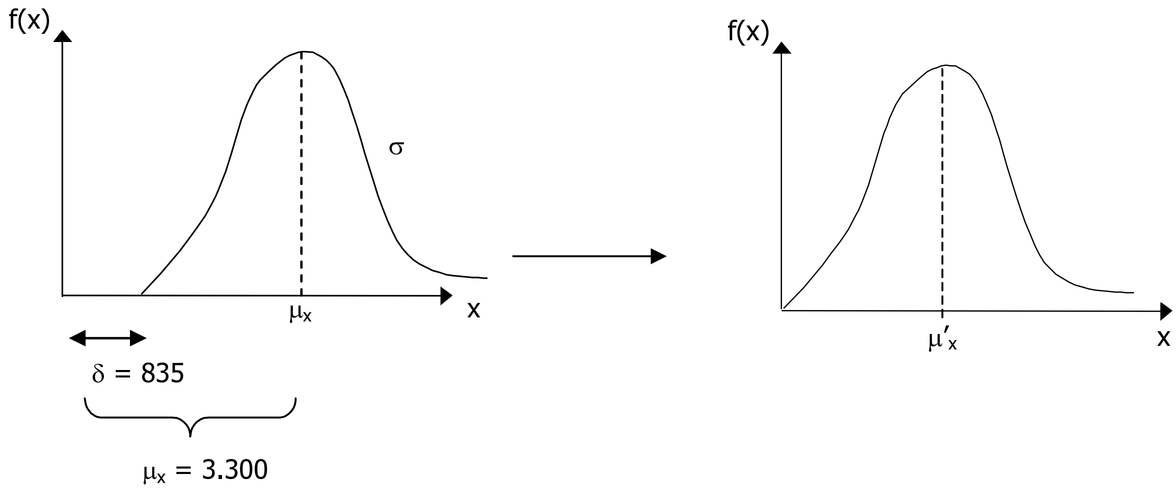
Sabe-se que a média e o desvio padrão das vazões máximas são:

μ_x = 3.300 m^3/s

σ_x = 470 m^3/s

Considerar que a vazão mínima possível é 835 m^3/s. Calcular os parâmetros da Log 3

Solução:



Sabemos que:

$\mu_x' = 3.300 - 835 = 2.465\ m^3/s$

$\sigma_x = 470\ m^3/s$

Mas,

$$\mu_x' = \exp\left(\mu_y + \frac{1}{2}\,\sigma_y^2\right)$$

$$\upsilon_x^2 = e^{\sigma_y^2} - 1 \longrightarrow \upsilon_x^2 = \frac{\sigma x^2}{\mu'_x{}^2} = \frac{(470)^2}{(2.465)^2} = 0{,}0364$$

$$(0{,}0364) = e^{\sigma y^2} - 1 \longrightarrow \sigma_y{}^2 = l_n(1{,}0364) = 0{,}0358$$

$$\boxed{\sigma y = 0{,}1891}$$

Então,

$$2.465 = \exp\left(\mu_y + \frac{1}{2}\,\underbrace{\sigma y^2}_{0{,}0358}\right)$$

$$\mu_y = l_n(2.465) - \frac{1}{2}(0{,}358))$$

$$\boxed{\mu y = 7{,}79}$$

Assim, X ~ Λ(7,79; 0,1891; 835)

Exercício 7.14 (Lista Terezinha pág.6)

Consideremos a seguinte seqüência de dados hidrológicos (ordenados)

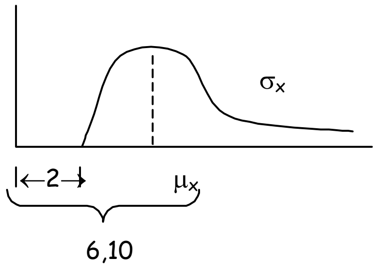
2,50	4,30	4,65	5,31	6,30	7,14	9,43	
2,70	4,32	4,70	5,50	6,36	7,36	9,69	
3,80	4,50	4,73	5,78	6,43	7,80	10,20	
4,05	4,57	4,85	5,82	6,66	8,10	10,68	
4,10	4,60	4,90	5,88	6,89	8,56	11,50	
4,13	4,61	5,10	5,93	6,92	8,77		41 valores

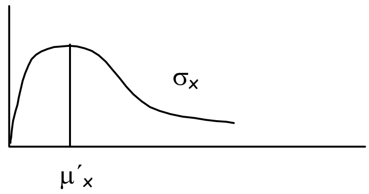
Ajustar uma Log-normal a 3 parâmetros. Supor o menor valor possível para qualquer observação é x = 2,0.

$\mu_x = 6{,}10$

$\sigma_x = 2{,}14$

$$\begin{cases} \mu_x' = \mu_x - \delta = 6{,}10 - 2 = 4{,}10 \\ \sigma_x' = \sigma_x = 2{,}14 \end{cases}$$





$$\upsilon_x^2 = e^{\sigma_y^2} - 1$$

$$\upsilon_x^2 = \frac{\sigma_x'^2}{\mu_x'^2} = 0{,}27$$

$$\therefore\ e^{\sigma y^2} = 1{,}27$$

$$\sigma_y^2 = l_n(1{,}27)$$

$$\sigma_y^2 = 0{,}239$$

$$\mu'_x = \exp\left(\mu_y + \frac{1}{2}\sigma_y^2\right)$$

$$\rho_n(4{,}10) = \mu_y + \frac{1}{2}\ \sigma_y^2$$

$$\mu_y = 1{,}411 - 0{,}11$$

$$\boxed{\mu_y = 1{,}29}$$

$$\boxed{x \sim \Lambda(1{,}29;\ 0{,}49;\ 2)}$$

7. GAMA A 3 PARÂMETROS

Similarmente a Lognormal a 3 parâmetros, se x for substituído por y = (x-δ), temos que a f.d.p. será dada por:

$$f(x) = \frac{\beta^{\alpha} \cdot (x - \delta)^{(\alpha-1)} \cdot e^{-\beta(x-\delta)}}{\Gamma(\alpha)}$$

onde α, β e δ são parâmetros e $\Gamma(\alpha)$ é a Função Gama.

Apenas a média é alterada, passando neste caso para:

$$\mu'_x = \mu_x - \delta$$

Exercício 7.15

Ajustar a distribuição do exemplo anterior (Terezinha, pág.6) a uma Gama III.

$$\begin{cases} \mu_x = 6{,}10 \\ \sigma_x = 2{,}14 \\ \delta = 2 \end{cases}$$

$$\mu'_x = 4{,}10 = \frac{\alpha}{\beta}$$

$$\sigma_x^2 = \frac{\alpha}{\beta^2} \quad \rightarrow \quad 4{,}56 = \frac{4{,}10\,\beta}{\beta^2} \qquad \beta = \frac{4{,}10}{4{,}56} \qquad \beta = 0{,}9$$

$$\alpha = 4{,}10 \times 0{,}9 \qquad \alpha = 3{,}68$$

$$\boxed{x \sim \Gamma(3{,}68;\ 0{,}9;\ 2)}$$

8. DISTRIBUIÇÃO DE EXTREMOS DO TIPO I ou DISTRIBUIÇÃO DE GUMBEL

O Método de Gumbel baseia-se em uma distribuição de valores extremos. A distribuição é dada por:

P (X≤ x) = exp(-exp(-y))

onde P é a probabilidade de um dado valor de vazão ser igualado ou excedido e y é a variável reduzida dada por:

$$y = (x - x_f)\frac{S_n}{S_x}$$

$$x_f = \bar{x} - S_x\left(\frac{\bar{y}_n}{S_n}\right)$$

onde x_f é a moda dos valores extremos, S_n é o desvio padrão da variável reduzida Y, S_x é o desvio padrão da variável x, e $\bar{x}$ e $\bar{y}$, as médias das variáveis x e y, respectivamente.

A aplicação do método de Gumbel no cálculo da vazão é mostrada nos passos seguintes:

1. Determinar a medida $(\bar{x})$ e o desvio-padrão (S_x) da série de dados históricos.

2. Em função do número de dados (n), extrair da Tabela 7.4 os valores esperados da medida $(\bar{y}_n)$ e desvio-padrão (s_n), associados a variável reduzida.

Tabela 7.4 - Valores esperados da média (Y_n') e desvio-padrão (S_n) da variável reduzida (y) em função do número de dados (n). (Fonte: VILLELA, 1975).

n	$\bar{y}_n$	S_n	n	$\bar{y}_n$	S_n
20	0,52	1,06	80	0,56	1,19
30	0,54	1,11	90	0,56	1,20
40	0,54	1,14	100	0,56	1,21
50	0,55	1,16	150	0,56	1,23
60	0,55	1,17	200	0,57	1,24
70	0,55	1,19	∞	0,57	1,28

3. Determinar a moda dos valores extremos, pela expressão seguinte:

$$x_f = \bar{x} - S_x\left(\frac{\bar{Y}_n}{S_n}\right)$$

4. Em função do período de retorno (T_r), extrair da tabela 7.5, o valor da variável reduzida (y) – ou usar a fórmula

Tabela 7.5 – Variável reduzida, Probabilidade e período de retorno. (Fonte: VILLELA, 1975).

Variável Reduzida (y)	Período de Retorno (T_r)	Probabilidade (1 – P)	Probabilidade (P)
0,000	1,58	0,632	0,368
0,367	2,00	0,500	0,500
0,579	2,33	0,429	0,571
1,500	5,00	0,200	0,800
2,250	10,0	0,100	0,900
2,970	20,0	0,050	0,950
3,395	30,0	0,033	0,967
3,902	50,0	0,020	0,980
4,600	100	0,010	0,990
5,296	200	0,005	0,995
5,808	300	0,003	0,997
6,214	500	0,002	0,998
6,907	1000	0,001	0,999

5. Determinar a vazão de projeto (x), aplicando elementos obtidos nos passos precedentes à equação:

$$y = (x - x_f)\frac{S_n}{S_x}$$

Ou seja,

$$x = x_f + y\frac{S_x}{S_n}$$