Inferência em Modelos de Posição e Escala com Dados Censurados

Francisco Felipe de Queiroz

Centro de Ciências Exatas e da Terra
Universidade Federal do Rio Grande do Norte - UFRN
Departamento de Estatística
Seminário de Tópicos Especiais-STE
Professora: Dione Maria Valença

13 de maio de 2015

Contents

Sumário

Introdução

No estudo de probabilidade e estatística é muito comum encontrarmos famílias de distribuições que possuem características importantes.

Um exemplo destas famílias é a Família Exponencial que é muito utilizada em Inferência Estatística. Esta família de distribuições nos ajuda a encontrar, de maneira razoavelmente simples, estatísticas suficientes, Testes de hipóteses UMP (pelo critério do TRVG), enfim...Uma outra família de distribuições (a qual estudaremos) é a Família de Posição e Escala.

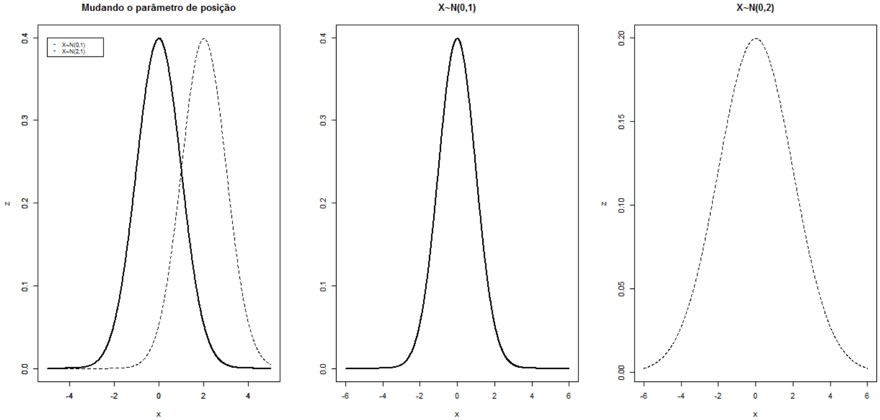
Sumário

Família de Posição e Escala

Consideremos uma variável aleatória Y, se conseguirmos escrever a f.d.p de Y como sendo

$$f_Y(y) = \frac{1}{\sigma} f_0\left(\frac{y-\mu}{\sigma}\right), \tag{1}$$

em que $\mu(-\infty < \mu < \infty)$ é o parâmetro de posição, $\sigma(\sigma > 0)$ é o parâmetro de escala e $f_0(\cdot)$ é uma f.d.p., diremos, então, que a distribuição da v.a. Y pertence a família de Posição e Escala.



Mudando o parâmetro de posição
X~N(0,1)
X~N(2,1)
X~N(0,1)
X~N(0,2)
z
x

Sumário

Exemplo 1

Consideremos o exemplo clássico em que $Y \sim N(\mu, \sigma^2)$, para verificarmos se a distribuição Normal pertence a família de posição e escala precisamos escrever a f.d.p de Y, $f_Y(y)$, na forma descrita em 1.

Para tanto, note que

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\dfrac{(y-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma} f_0\left(\frac{y-\mu}{\sigma}\right), \quad -\infty \leq y \leq \infty$$

em que,

$$f_0(y) = \frac{1}{\sqrt{2\pi}}\, e^{\dfrac{y^2}{2}}$$

Notemos, ainda, que $f_0(y)$ denota a f.d.p de uma v.a Normal Padrão, isto é, $N(0,1)$. Com isso conseguimos escrever a f.d.p de Y como descrito em 1 e verificamos que $f_0(y)$ é uma f.d.p. Portanto, mostramos que a distribuição Normal pertence a família de posição e escala.

Sumário

Função Acumulada e Função de Sobrevivência

Consideremos uma v.a. Y com f.d.p dada em 1, em que $f_0(\cdot)$ é uma f.d.p.. Por definição, sabemos que

$$F_Y(y) = P(Y \leq y) \quad e \quad S(y) = P(Y > y) = 1 - F_Y(y),$$

onde $F_Y(y)$ denota a função acumulada da v.a. Y e $S(y)$ a função de sobrevivência.

Se Y pertence a família de posição e escala, então:

Propriedades

- $F_Y(y) = F_0\left(\dfrac{y-\mu}{\sigma}\right)$
- $S(y) = S_0\left(\dfrac{y-\mu}{\sigma}\right)$

Prova:
Encontraremos função Acumulada de Y.

$$F_Y(y) = P(Y \leq y) = \int_{-\infty}^{y} \frac{1}{\sigma} f_0\left(\frac{t-\mu}{\sigma}\right) dt = \frac{1}{\sigma}\int_{-\infty}^{y} f_0\left(\frac{t-\mu}{\sigma}\right) dt =$$

$$\Rightarrow F_Y(y) = F_0\left(\frac{t-\mu}{\sigma}\right)\Big|_{-\infty}^{y} = F_0\left(\frac{y-\mu}{\sigma}\right).$$

Da mesma forma,

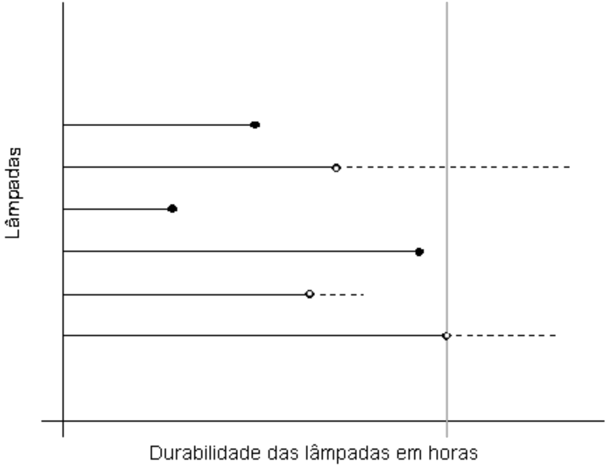
$$S(y) = S_0\left(\frac{y-\mu}{\sigma}\right)$$

Note que conseguimos escrever a função acamulada e função de sobrevivência da v.a. Y como função da densidade $f_0(\cdot)$, que é, comumente chamada de Densidade Padrão da v.a. Y (isso deve-se ao fato de que a função densidade $f_0(\cdot)$, é um caso particular da densidade de Y quando $\mu = 0$ e $\sigma = 1$) e, algumas vezes, é mais fácil trabalhar com a densidade $f_0(\cdot)$ do que com a densidade de Y.

Sumário

Censura

Para compreendermos o conceito de censura em Análise de Sobrevivência, suponhamos o exemplo à seguir.
Deseja-se estudar a durabilidade de lâmpadas em um período de 720 horas. Algumas das observações podem ser vistas no gráfico a seguir.


Lâmpadas
Durabilidade das lâmpadas em horas

Sumário

Preliminares

Definiremos, agora, duas v.a.(independentes) T e C, tal que:

- T_i : tempo até "falha", $i = 1, ..., n$
- C_i : tempo até censura, $i = 1, ..., n$

Pelo gráfico anterior, não é difícil verificar que em um problema prático, dada uma determinada lâmpada, só observaremos o menor tempo. Isto é, o menor entre os tempos de "falha" e censura. Com isso, definiremos uma nova v.a. T^*, de modo que

$$T_i^* = min\{T_i, C_i\},$$

ainda, definiremos uma função indicadora, dada por:

$$\delta_i = \begin{cases} 1, & \text{se} \quad T_i \leq C_i \\ 0, & \text{se} \quad T_i > C_i \end{cases}$$

Ainda, suponhamos que

$$T_i \sim f_\theta(t_i) \quad e \quad P(T_i > t_i) = S_\theta(t_i)$$

e

$$C_i \sim g(t_i) \quad e \quad P(C_i > t_i) = G(t_i)$$

Pelas considerações anteriores podemos perceber que os dados observados serão da forma $(t_i^*, \delta_i), i = 1, ..., n$. Pode-se mostrar que, para uma amostra aleatória observada, $D = (t_i^*, \delta_i), i = 1, ..., n$, a verossimilhança será dada por:

$$L(\theta; D) = \prod_{i=1}^{n} f_\theta(t_i)^{\delta_i} S_\theta(t_i)^{(1-\delta_i)}$$

Função de verossimilhança

Definiremos agora uma nova v.a Y_i, tal que $Y_i = \log T_i$, de modo que Y_i pertence a família de posição e escala. Então, para uma amostra aleatória observada (com censura), $D = (\log t_i, \delta_i)$, a Função de Verossimilhança para $(\mu, \sigma)^t$ é dada por

$$L(\mu, \sigma; D) = \prod_{i=1}^{n} \left[\frac{1}{\sigma} f_0\left(\frac{y_i - \mu}{\sigma}\right)\right]^{\delta_i} S_0\left(\frac{y_i - \mu}{\sigma}\right)^{1-\delta_i} \qquad (2)$$

onde μ é o parâmetro de posição, σ o parâmetro de escala e $y_i = \log t_i$.

Sumário

Nota

Quando nos referirmos a θ estamos nos referindo a $\theta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}$

Preliminares

Para a v.a. $Y_i = \log T_i$, pertencente a família de posição e escala, com função de verossimilhança dada em 2, o $\log L(\mu, \sigma)$ é dado por

$$\log L(\mu, \sigma) = -r \log \sigma + \sum_{i=1}^{n} \left[\delta_i \log f_0(z_i) + (1 - \delta_i) \log S_0(z_i)\right], \tag{3}$$

em que $z_i = \left(\dfrac{y_i - \mu}{\sigma}\right)$, $r = \sum_{i=1}^{n} \delta_i$ e, como mensionado anteriormente, $y_i = \log t_i$.

As derivadas parciais, e segundas derivadas são dadas pelas expressões a seguir.
Temos que $\dfrac{\partial z_i}{\partial u} = -\sigma^{-1}$ e $\dfrac{\partial z_i}{\partial \sigma} = -z_i\sigma^{-1}$, então:

$$\frac{\partial \log L}{\partial u} = -\frac{1}{\sigma}\sum_{i=1}^{n}\left[\delta_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1-\delta_i)\frac{\partial \log S_0(z_i)}{\partial z_i}\right]$$

e,

$$\frac{\partial \log L}{\partial \sigma} = -\frac{r}{\sigma} - \frac{1}{\sigma}\sum_{i=1}^{n}\left[\delta_i z_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1-\delta_i) z_i \frac{\partial \log S_0(z_i)}{\partial z_i}\right].$$

As segundas derivadas são dadas por:

$$\frac{\partial^2 \log L}{\partial u^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[\delta_i \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1-\delta_i)\frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right],$$

$$\frac{\partial^2 \log L}{\partial \sigma^2} = \frac{r}{\sigma^2} + \frac{2}{\sigma^2}\sum_{i=1}^{n}\left[\delta_i z_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1-\delta_i) z_i \frac{\partial \log S_0(z_i)}{\partial z_i}\right] + \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[\delta_i z_i^2 \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1-\delta_i) z_i^2 \frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right]$$

E, por último:

$$\frac{\partial^2 \log L}{\partial u \partial \sigma} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[\delta_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1-\delta_i)\frac{\partial \log S_0(z_i)}{\partial z_i}\right]$$
$$+\frac{1}{\sigma^2}\sum_{i=1}^{n}\left[\delta_i z_i \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1-\delta_i) z_i \frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right]$$

Notemos que, na família de posição e escala, temos dois parametros envolvidos, $\theta = (\mu, \sigma)$. Então, temos que

$$U(\theta) = \frac{\partial}{\partial \theta} \log L(\theta; \mathbf{x}) = \begin{pmatrix} \dfrac{\partial}{\partial \mu} \log L(\theta; \mathbf{x}) \\ \\ \dfrac{\partial}{\partial \sigma} \log L(\theta; \mathbf{x}) \end{pmatrix} = \begin{pmatrix} U_\mu(\theta) \\ U_\sigma(\theta) \end{pmatrix},$$

Para encontrar os estimadores para θ utilizaremos o método da máxima verossimilhança.
Como θ é um vetor com apenas dois parâmetros, o EMV para θ pode ser obtido como solução do sistema:

$$\begin{cases} U_\mu(\hat{\theta}) = 0 \\ U_\sigma(\hat{\theta}) = 0 \end{cases} \Rightarrow \begin{cases} \left.\dfrac{\partial}{\partial \mu} \log L(\theta; \mathbf{x})\right|_{\theta=\hat{\mu}} = 0 \\ \left.\dfrac{\partial}{\partial \sigma} \log L(\theta; \mathbf{x})\right|_{\theta=\hat{\sigma}} = 0 \end{cases}$$

Como o temos apenas dois parâmtros, temos que a matriz da informação de Fisher é dada por

$$E[\mathbb{I}_F(\theta)] = -E[H(\theta)]$$

em que

$$H(\theta) = \begin{bmatrix} \dfrac{\partial^2}{\partial \mu^2} \log L(\theta; \mathbf{x}) & \dfrac{\partial^2}{\partial \mu \partial \sigma} \log L(\theta; \mathbf{x}) \\ \dfrac{\partial^2}{\partial \mu \partial \sigma} \log L(\theta; \mathbf{x}) & \dfrac{\partial^2}{\partial \sigma^2} \log L(\theta; \mathbf{x}) \end{bmatrix}$$

No entanto, em casos que temos censura, não é possível obter a matriz da informação de Fisher.

Com isso, definimos a matriz da informação observada, que é dada por

$$\mathbb{I}_{obs}(\hat{\theta}) = -H(\hat{\theta}),$$

ainda, definiremos uma nova matriz, V, tal que

$$V(\hat{\theta}) = \mathbb{I}_{obs}(\hat{\theta})^{-1} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

Intervalo de Confiança

Agora, consideremos $\hat{\theta} = \begin{pmatrix} \hat{\mu} \\ \hat{\sigma} \end{pmatrix}$, então,

$$\hat{\theta} \dot{\sim} N_2\left(\theta, \mathbb{I}_{obs}(\hat{\theta})^{-1}\right), \quad n \text{ grande}$$

Assim, para construir um $IC(\mu, \gamma)$ e $IC(\sigma, \gamma)$ podemos utilizar, respectivamente, as seguintes quantidades pivotais:

$$Q(\mathbf{X}, \mu) = \frac{\hat{\mu} - \mu}{ep(\hat{\mu})} \quad e \quad Q(\mathbf{X}, \sigma) = \frac{\hat{\sigma} - \sigma}{ep(\hat{\sigma})}$$

Em que ambas quantidades pivotais tem distribuição assintótica $N(0,1)$, $ep(\hat{\mu})$ e $ep(\hat{\sigma})$ são dados, respectivamente, por $\sqrt{\hat{V}_{11}}$ e $\sqrt{\hat{V}_{22}}$

Assim, os respectivos intervalos de confiança para μ e σ são dados por:

$$IC(\mu;\gamma) : \left[\hat{\mu} \pm z_{\frac{\alpha}{2}} ep(\hat{\mu})\right] \quad (4)$$

$$IC(\sigma;\gamma) : \left[\hat{\sigma} \pm z_{\frac{\alpha}{2}} ep(\hat{\sigma})\right], \quad (5)$$

em que $\gamma = 1 - \alpha$, e $z_{\frac{\alpha}{2}}$ é o quantil da distribuição Normal Padrão, tal que $P\left(Z \leq z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$

Teste de Hipóteses

Para testar as hipóteses

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \quad ou \quad \begin{cases} H_0 : \sigma = \sigma_0 \\ H_1 : \sigma \neq \sigma_0 \end{cases}$$

Sabemos que, para testar este tipo de hipótese podemos utilizar o TRVG, em que

$$\lambda(\mathbf{x}) = \left[\frac{L(\tilde{\theta}_0; \mathbf{x})}{L(\hat{\theta}; \mathbf{x})} \right],$$

e a região de rejeição é dada por $A_1 = \{\mathbf{x}; \lambda(\mathbf{x}) \leq k\}$.

No entanto, podemos utilizar o resultado assintótico do TRVG, com isso, a estatística do TRVG é dada por:

$$\zeta_{RV} = -2 log \lambda(\mathbf{x}),$$

que, sob H_0, tem distribuição assintótica $\chi_{(1)}$.
Vale lembrar que este resultado assintótico só vale para testar hipótese "$=$" contra "$\neq$".

Sumário

Conclusões

Embora não consigamos encontrar, analiticamente, as derivadas da função de verossimilhança para a família de posição e escala para dados com censura, pode-se encontrar, facilmente, formas fechadas para as expressões destas derivadas, podendo, assim, serem implementadas em um software.

Sumário

Referências

- Lawless, Jerald F., **Statistical Models and Methods for Lifetime Data**. Ontario: Wiley, 2002.
- Bolfarine, H. Sandoval, M. C., **Introdução a Inferência Estatística**. São Paulo: SBM, 2000.