MAE0311 - Inferência Estatística

Fernando Henrique Ferraz Pereira da Rosa
Vagner Aparecido Pedro Junior

18 de novembro de 2003

Lista 6[1]

1. Seja X uma variável aleatória com função de densidade $f(x|\theta) = \theta^2 x e^{-\theta x}$, $x > 0, \theta > 0$, queremos testar $H_0 : \theta = 1$ versus $H_1 : \theta = 2$.

   (a) Qual é a região crítica se $n = 5$ e $\alpha = 0.05$?
   Notemos que
   $$X \sim Gama(\lambda = \theta, r = 2)$$
   e que
   $$\begin{aligned} L_0(x) &= \prod_{i=1}^{n} f(x_i|\theta_0) = \prod_{i=1}^{n} f(x_i|\theta = 1) = \prod_{i=1}^{n} x_i e^{-x_i} \\ &= e^{-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} x_i \\ L_1(x) &= \prod_{i=1}^{n} f(x_i|\theta_1) = \prod_{i=1}^{n} f(x_i|\theta = 2) = \prod_{i=1}^{n} 2^2 x_i e^{-2x_i} \\ &= 4^n e^{-2\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} x_i \end{aligned}$$
   Pelo Lema de Neyman-Pearson, utilizando a razão de verossimilhanças simples, temos que o teste mais poderoso será aquele com região região crítica dada por
   $$A_1^* = \left\{ x; \frac{L_1(x)}{L_0(x)} \geq k \right\}$$
   Notemos que
   $$\frac{L_1(x)}{L_0(x)} = \frac{4^n e^{-2\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} x_i}{e^{-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} x_i} = 4^n e^{-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} x_i$$

[1] Powered by LaTeX 2_ε, R 1.8.0 and Gentoo 1.4

Que é equivalente a rejeitar H_0 quando $\sum_{i=1}^n x_i < -\log k/4^n = c$. Portanto a região crítica do teste M.P. é dada por

$$A_1^* = \left\{x; \sum_{i=1}^n x_i < c\right\}$$

Como queremos $\alpha = 0.05$ devemos exigir

$$P_{H_0}(X \in A_1^*) = 0.05$$

Notemos que $\sum_{i=1}^n X_i \sim Gama(\lambda = 1, r = 2n)$ e que portanto, é simples achar c que satisfaça a condição dada. Com o seguinte comando no R, obtemos esse valor:

```
> qgamma(0.05,10,1)
[1] 5.425406
```

Assim a região crítica para $\alpha = 0.05$ e $n = 5$ fica sendo

$$A_1^* = \left\{x; \sum_{i=1}^n x_i < 5.425406\right\}$$

Façamos uma simulação para verificar esse resultado.

```
# Geramos em primeiro lugar 100 mil amostras
# de tamanho 5 de uma gama (r=2,l=1)
sample <- matrix(rgamma(5*100000,2,1),nrow=5)
# Definamos uma função que nos indicará se, de acordo com a
# região critica encontrada, rejeitamos ou não H0
> rejeita <- function(x) {
+   if (sum(x) < qgamma(0.05,10,1)) {
+     1
+    }
+    else {
+     0
+  }
+ }
#Executemos essa funcao para cada amostra observada
#calculando a proporcao observada que tivemos de rejeicao
#de H0, H0 verdadeira (estimativa de alfa)
sum(apply(sample,2,rejeita))/100000
[1] 0.0508
```

Os resultados da simulação portanto sugerem que a região de rejeição está apropriada.

(b) Se $n = 1$, qual é o teste que minimiza $\alpha+\beta$? E qual o valor de $\alpha+\beta$?
Pelo Lema 6.3.1. sabemos que o o teste com região crítica

$$A_1^* = \left\{x; \frac{L_1(x)}{L_0(x)} \geq \frac{a}{b}\right\}$$

minimiza o valor de $a\alpha + b\beta$. Como queremos minimizar o valor de $\alpha + \beta$ basta fazermos $a = b = 1$ e acharmos a seguinte região crítica

$$A_1^* = \left\{x; \frac{L_1(x)}{L_0(x)} \geq 1\right\}$$

Se $n = 1$, temos

$$\frac{L_1(x)}{L_0(x)} = \frac{2^2 x e^{-2x}}{xe{-}x} = 4e^{-x} \geq 1 \Rightarrow e^{-x} \geq \frac{1}{4} \Rightarrow x < \log 4$$

Assim a região crítica do teste que minimiza $\alpha + \beta$ é dada por

$$A_1^* = \{x; x < \log 4\}$$

Calculando agora α e β:

$$\alpha = P_{H_0}(X \in A_1^*) = P(X < \log 4)$$

Onde $X \sim Gama(2, 1)$. Obtemos então α, através do comando

```
> pgamma(log(4),2,1)
[1] 0.2357868
```

Analogamente para β

$$\beta = P_{H_1}(X \notin A_1^*) = P(X \geq \log 4)$$

Onde $X \sim Gama(2, 2)$. Obtemos então, através do comando

```
> pgamma(log(4),2,2,lower.tail=F)
[1] 0.4034264
```

Logo $\alpha + \beta = 0.6392132$.

3. Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória X com função de densidade dada por

$$f(x|\theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0.$$

(a) Mostre que o teste mais poderoso para testar $H_0 : \theta = 1$ versus $H_1 : \theta = 2$ rejeita H_0, se e somente se, $\sum_{i=1}^{n} -\log x_i \leq a$, onde a é uma constante.

O teste mais poderoso será aquele da razão de verossimilhanças simples, cuja região crítica é dada por

$$A_1^* = \left\{ x; \frac{L_1(x)}{L_0(x)} \geq k \right\}$$

Desenvolvendo a razão de verossimilhanças

$$\begin{aligned} \frac{L_1(x)}{L_0(x)} &= \frac{2^n \prod_{i=1}^{n} x_i}{1} \geq k \Rightarrow \prod_{i=1}^{n} \geq \frac{k}{2n} \Rightarrow \sum_{i=1}^{n} \log x_i \geq \log \frac{k}{2n} \\ &\Rightarrow -\sum_{i=1}^{n} \log x_i \leq \log \frac{k}{2n} = a \end{aligned}$$

O que mostra a identidade pedida.

(b) Sendo $n = 2$ e $\alpha = (1 - \log 2)/2$, qual a região crítica? Para isso precisamos em primeiro lugar encontrar a distribuição de $-\sum_{i=1}^{n} \log X_i$. Analisemos primeiro cada parcela da soma, verificando a distribuição de $-\log X_i$. Sob H_0, temos que cada X_i é uniforme em $0, 1$, ou seja

$$f_X(x) = \underset{(0,1)}{I}(x)$$

Notando que

$$f_Y(y) = f_x(g^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right|$$

e encontrando as quantidades necessárias:

$$g^{-1}(y) = e^{-y} \qquad \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right| = \left| -e^{-y} \right| = e^{-y}$$

Temos que $Y = -\log X_i$ segue distribuição com função de densidade dada por:

$$f_Y(y) = e^{-y} \underset{(0,1)}{I}(e^{-y}) \quad = e^{-y} \quad , y > 0$$

Ou seja, uma variável aleatória com distribuição exponencial de parâmetro $\lambda = 1$. Mais que isso, temos ainda a distribuição de $\sum_{i=1}^{n} -\log X_i$, pois a soma de n exponenciais independentes com parâmetro $\lambda = 1$ segue uma $Gama(n, 1)$. Assim, para achar a região crítica para $\alpha = \frac{1-\log 2}{2}$, basta achar o quantil $\frac{1-\log 2}{2}$ de uma $Gama(2, 1)$,o que é feito a partir do comando

```
> qgamma((1-log(2))/2,2,1)
[1] 0.6931472
```

Logo, a Região Crítica ao nível de significância $\alpha = \frac{1-\log 2}{2}$ será dada por:

$$A_1^* = \left\{ -\sum_{i=1}^{n} \log x_i \leq 0.6931472 \right\}$$

8. Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória $X \sim N(0, \sigma^2)$.

   (a) Encontre o teste UMP para testar $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$. Definamos o teste alternativo $H_0' : \sigma^2 = \sigma_0^2$ versus $H_1' : \sigma^2 = \sigma_1^2 \quad (\sigma_1^2 > \sigma_0^2)$ e verifiquemos se sua região crítica depende de σ_0^2. O teste mais poderoso para esse teste alternativo será o de região crítica dada por

$$A_1^* = \left\{ x; \frac{L_1(x)}{L_0(x)} \geq k \right\}$$

Desenvolvendo a razão de verossimilhanças

$$\begin{aligned}
\frac{L_1}{L_0} &= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{\frac{-x_i^2}{2\sigma_1^2}}}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{\frac{-x_i^2}{2\sigma_0^2}}} = \left(\frac{\sigma_0^2}{\sigma_1^2}\right)^{n/2} e^{\sum_{i=1}^{n} x_i^2 \left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2}\right)} \geq k \\
&\Rightarrow \quad e^{\sum_{i=1}^{n} x_i^2 \left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2}\right)} \geq k \left(\frac{\sigma_1^2}{\sigma_0^2}\right)^{n/2} \\
&\Rightarrow \quad \sum_{i=1}^{n} x_i^2 \left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2}\right) \geq \log \left[k \left(\frac{\sigma_1^2}{\sigma_0^2}\right)^{n/2} \right] \\
&\Rightarrow \quad \sum_{i=1}^{n} x_i^2 \geq \log \left[k \left(\frac{\sigma_1^2}{\sigma_0^2}\right)^{n/2} \right] \left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2}\right)^{-1} = c
\end{aligned}$$

Assim, o nosso teste MP para nossas hipóteses alternativas terá região crítica dada por

$$A_1^* = \left\{ \sum_{i=1}^{n} x_i^2 \geq c \right\}$$

Como o teste acima vale para qualquer valor de σ_1^2, temos pela definição 6.4.1., que ele é também o teste UMP para a hipótese composta inicial.

(b) Seja $\alpha = 0.05$, $n = 9$ e $\sigma_0^2 = 9$, faça o gráfico da função poder.

Precisamos primeiro encontrar a distribuição de $\sum_{i=1}^{n} X_i^2$. Dados $n = 9$ e $\sigma_0^2 = 9$, temos que

$$\sum_{i=1}^{9} \left(\frac{X_i}{3}\right)^2 \sim \chi^2_{(9)}$$

Considerando $\alpha = 0.05$ e dada H_0

$$0.05 = P_{H_0}\left(\sum_{i=1}^{n} X_i^2 > c\right) = P\left(\frac{1}{9}\sum_{i=1}^{n} X_i^2 > \frac{c}{9}\right) = P\left(\sum_{i=1}^{n} \left(\frac{X_i}{3}\right)^2 > \frac{c}{9}\right)$$

Dessa forma, basta encontrarmos $\frac{c}{9}$ com o comando:

```
> qchisq(0.95,9)
[1] 16.91898
```

E portanto $\frac{c}{9} = 16.91898 \Rightarrow c = 152.2708$.

A região crítica do teste UMP com nível de significância $\alpha = 0.05$ para as hipóteses especificadas fica sendo

$$A_1^* = \left\{\sum_{i=1}^{n} x_i^2 \geq 152.2708\right\}$$

A função poder é definida por

$$\pi(\sigma_2) = P_{\sigma_2}\left(\sum_{i=1}^{9} X_i^2 \in A_1^*\right)$$

Como por definição $Var(X_i) = \sigma^2$ temos

$$\pi(\sigma_2) = P_{\sigma_2}\left(\frac{1}{\sigma^2}\sum_{i=1}^{9} X_i^2 \geq \frac{152.2708}{\sigma^2}\right) = P\left(W \geq \frac{152.2708}{\sigma^2}\right)$$

onde $W \sim \chi^2_{(9)}$. Assim a função poder se iguala a 1 menos a função de distribuição de uma qui-quadrado com 9 graus de liberdade. O seu gráfico se encontra na figura 1.

11. Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória $X \sim Poisson(\theta_1)$ e sejam $Y_1, \ldots, Y_m$ uma amostra aleatória da variável aleatória $Y \sim Poisson(\theta_2)$ sendo as amostras independentes.

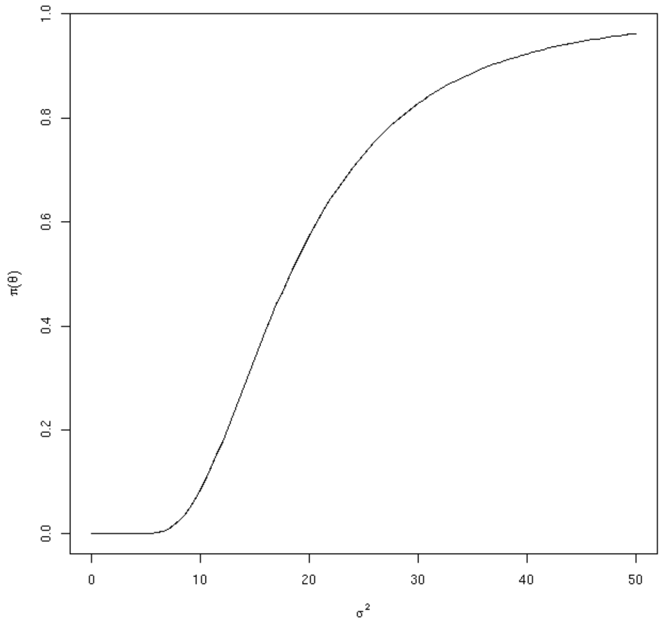

Figura 1: Exericício 8: Função poder para o teste UMP

(a) Encontre o teste da RVG (aproximado) para testar $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$. i

Encontremos em primeiro lugar o estimador de máxima verossimilhança para (θ_1, θ_2):

$$L(\theta_1, \theta_2; x, y) = \prod_{i=1}^{n} \frac{e^{-\theta_1}{\theta_1}^{x_i}}{x_i!} \prod_{j=1}^{m} \frac{e^{-\theta_2}{\theta_2}^{y}_{j}}{y_j!} = \frac{e^{-n\theta_1}\theta_1^{\sum x_i} e^{-m\theta_2}\theta_2^{\sum y_j}}{\prod_{i=1}^{n} x_i! \prod_{j=1}^{m} y_j!}$$

Que implica:

$$l(\theta_1, \theta_2; x, y) = -\log \prod x_i! - \log \prod y_j!$$

Calculando as derivadas parciais e igualando a zero obtemos imediatamente:

$$(\hat{\theta_1}, \hat{\theta_2}) = (\overline{X}, \overline{Y})$$

que é o estimador de máxima verossimilhança para (θ_1, θ_2) em Θ. Notando que o EMV de (θ_1, θ_2) em $(\theta_1 = \theta_2 = \theta)$, fica dado por

$$\hat{\theta} = \frac{n\overline{X} + m\overline{Y}}{n+m}$$

Temos que

$$\lambda(x,y)^{-1} \quad = \quad e^{-n\overline{X}-m\overline{Y}}\hat{\theta}^{-n\overline{X}-m\overline{Y}}\overline{X}^{-n\overline{X}}\overline{Y}^{-m\overline{Y}}$$

Pelo TRVG, usando 6.5.1 rejeitamos H_0 quando:

$$\lambda(x,y) \leq c \iff -2\log\lambda(x,y) > c$$

Simplificando $\lambda(x,y)$ e aplicando a transformação acima:

$$\begin{aligned}
\lambda(x,y) \leq c \quad &\iff \quad -2\log\lambda(x,y) > c \\
&= \quad -2\log\left(\left[k\hat{\theta}^{-n\overline{X}-m\overline{Y}}\right]^{-1}\right) \\
&= \quad 2\left[\log k - (n\overline{X} + m\overline{Y})\log\hat{\theta}\right] \\
&= \quad 2\left[\log\overline{X}^{n\overline{X}}\overline{Y}^{m\overline{Y}} - (n\overline{X} + m\overline{Y})\log\frac{n\overline{X} + m\overline{Y}}{n+m}\right]
\end{aligned}$$

A RC do teste é

$$A_1^* = \{(x,y) : -2\log\lambda(x,y) > c_1\}$$

onde c_1 é tal que $P(\chi^2_{(1)} > c_1) = \alpha$.

(b) Sendo $n = 5$, $\sum x_i = 3.8$, $m = 8$, $\sum y_i = 4.8$, qual a sua conclusão a um nível de significância de 5%?
Temos que $\overline{X} = 0.76$ e $\overline{Y} = 0.6$, assim:

$$-2\log\lambda(x,y) = 2\left[\log 0.76^{5\cdot 0.76}0.6^{8\cdot 0.6} - (3.8+4.8)\log\frac{3.8+4.8}{13}\right] = 0.11$$

Calculando o valor de c_1 no R

```
> qchisq(0.95,1)
[1] 3.841459
```

Como $0.11 < 3.84$, não há evidências que sugiram a rejeição de H_0.

Sobre

A versão eletrônica desse arquivo pode ser obtida em http://www.feferraz.net