MAE0311 - Inferência Estatística

Fernando Henrique Ferraz Pereira da Rosa

23 de novembro de 2003

Testes de Hipóteses

1 Testes

1. Seja $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória $X \sim Exp(\theta)$. Queremos testar $H_0 : \theta = 2$ versus $H_1 : \theta = 1$. Qual o teste mais poderoso para $\alpha = 0.05$ e $n = 10$?

   Pelo Lema de Neyman-Pearson, temos que o teste mais poderoso, dado α fixado, terá região crítica dada por

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

   Notemos então que

   $$\frac{L_1(\theta; x)}{L_0(\theta; x)} = \frac{e^{-n \sum x_i}}{2^n e^{-2n \sum x_i}} = 2^{-n} e^{n \sum x_i} \geq k \tag{1}$$

   Que é equivalente a rejeitar H_0 quando $\sum x_i \geq \frac{\log k 2^n}{n} = c$. Portanto a região crítica do teste MP será dada por:

   $$A_1^* = \left\{ x, \sum x_i \geq c \right\}$$

   Tomando $\alpha = 0.05$, c é tal que

   $$P_{H_0} \left(\sum X_i \geq c \right) = 0.05$$

   Notando que sob H_0, $\sum X_i \sim Gama(10, 2)$, obtemos $c = 7.8526$. Ou seja, a região crítica será dada por

   $$A_1^* = \left\{ x, \sum x_i \geq 7.8526 \right\}$$

   Analogamente obtemos $\beta = P_{H_1}\{x \notin A_1^*\} = 0.2653$.

2. Mantenhamos agora o problema exatamente da mesma forma. Só que troquemos as hipóteses, fazendo $H_0 : \theta = 1$ e $H_1 : \theta = 2$.

   Não é difícil ver que as contas permanecem praticamente as mesmas. Invertendo a fração do lado esquerdo da equação 1 temos:

$$\frac{L_1(\theta; x)}{L_0(\theta; x)} = 2^n e^{-n \sum x_i} \geq k$$

   Que é equivalente a rejeitar H_0 quando $\sum x_i \leq \frac{\log k 2^n}{n} = c$. Portanto a região crítica do teste MP será dada por:

$$A_1^* = \left\{ x, \sum x_i \leq c \right\}$$

   Notando agora que sob H_0, $\sum X_i \sim Gama(10, 1)$ e sob H_1, $\sum X_i \sim Gama(10, 2)$ temos que a região crítica do teste MP para $\alpha = 0.05$ e $n = 10$ será dada por

$$A_1^* = \left\{ x, \sum x_i \leq 5.4254 \right\}$$

   Analogamente obtemos $\beta = P_{H_1}\{x \notin A_1^*\} = 0.3567$.

2 Interpretação das Regiões Críticas Obtidas

Na figura 1 temos os gráficos de densidade de probabilidade para $\sum X_i$ e as respectivas regiões críticas para o primeiro e para o segundo testes propostos.

Estamos testando em ambos os casos se a amostra retirada veio de uma população exponencial com parâmetro $\theta = 1$ (designemos essa população de população A) ou $\theta = 2$ (população B), o que implica que a média para a população A será 1 e para a B será $\frac{1}{2}$.

Analisando a primeira região crítica encontrada, ela faz sentido de acordo com a interpretação intuitiva: rejeitamos $H_0 : \theta = 2$, quando a soma dos valores observados *exceder* um determinado valor fixado. É natural que seja assim, se levarmos em conta que a amostra veio da população A, ela vai ter uma média maior, e portanto, quanto maior a soma dos valores, maior a chance de que $H_1 : \theta = 1$ é mais apropriada e portanto maiores as evidências para rejeitar H_0.

A segunda região crítica encontrada, rejeita H_0 quando a soma dos valores obtidos na amostra *não atinge* um determinado valor. Isso também faz sentido, porque agora temos que se H_0 for verdadeira, a amostra veio da população A e portanto deve ter uma soma maior do que se tivesse vindo da população B. Assim, quanto menor for a soma dos valores, maiores são as evidências contra H_0 e a favor de H_1.

Por fim as disparidades ficam com relação aos erros tipo II: como uma $Gama(10, 2)$ é mais leptocúrtica que uma $Gama(10, 1)$, e no segundo caso a região crítica fica um pouco à direita do ponto de máximo da densidade, é também esperado que o erro tipo II seja maior para o segundo caso.

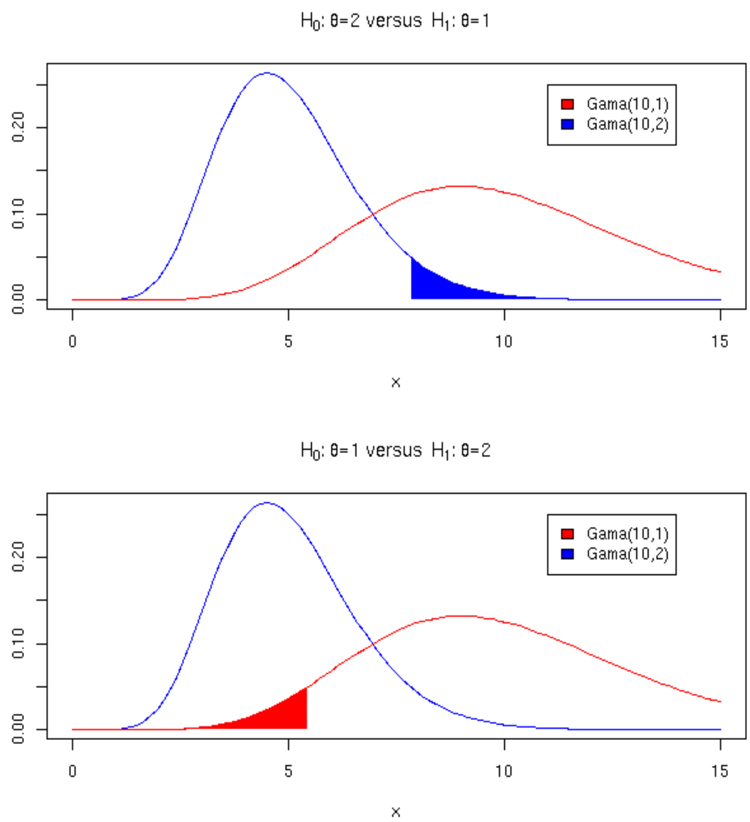

Figura 1: Comparação das regiões críticas

Sobre

A versão eletrônica desse arquivo pode ser obtida em http://www.fernandohrosa.com.br