MAE0311 - Inferência Estatística

Fernando Henrique Ferraz Pereira da Rosa

1 de novembro de 2003

Lista 5[1]

4. Seja X uma única observação da densidade

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

(a) Mostre que $-\theta \log X$ é uma quantidade pivotal e use-a para construir um intervalo de confiança para θ com coeficiente de confiança $\gamma = 1 - \alpha$.

Encontremos em primeiro lugar a distribuição de $Y = g(X) = -\theta \log X$. Sabemos que

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

Encontrando as quantidades necessárias:

$$g^{-1}(y) = x \Rightarrow y = -\theta \log x \Rightarrow -\frac{y}{\theta} = \log x \Rightarrow g^{-1}(y) = x = e^{-\frac{y}{\theta}}$$

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

Assim $-\theta \log X$ tem densidade dada por

$$f_y(y) = \theta \left(e^{-\frac{y}{\theta}} \right)^{\theta-1} \frac{e^{-\frac{y}{\theta}}}{\theta} = \left(e^{-\frac{y}{\theta}} \right)^{\theta-1} e^{-\frac{y}{\theta}} = \left(e^{-\frac{y}{\theta}} \right)^{\theta} = e^{-y}, y > 0$$

O que nos diz que $-\theta \log X$ segue distribuição exponencial com parâmetro $\lambda = 1$, que não depende de θ. Portanto $Q(\theta; X) = -\theta \log X$ é uma quantidade pivotal para θ.

Dados q_1 e q_2 fixados tais que

$$P(q_1 \leq -\theta \log X \leq q_2) = \gamma = 1 - \alpha,$$

[1] Powered by $\LaTeX\,2_\varepsilon$, R 1.8.0 and Gentoo 1.4

e desenvolvendo essa última expressão

$$\gamma = P\left(\frac{q_1}{\log X} \geq -\theta \geq \frac{q_2}{\log X}\right) = P\left(-\frac{q_1}{\log X} \leq \theta \leq -\frac{q_2}{\log X}\right)$$

Concluimos que um intervalo de confiança para θ com coeficiente de confiança γ será dado por

$$\left[-\frac{q_1}{\log X}; -\frac{q_2}{\log X}\right].$$

(b) Seja $Y = (-\log X)^{-1}$. Encontre o coeficiente de confiança associado ao intervalo $(Y/2, Y)$.
O problema se resume a encontrar γ tal que

$$P\left(\frac{Y}{2} \leq \theta \leq Y\right) = \gamma$$

Desenvolvendo essa expressão obtemos

$$\begin{aligned}
\gamma &= P\left(\frac{1}{-2\log X} \leq \theta \leq \frac{1}{-\log X}\right) = P\left(-\frac{1}{2} \geq \theta \log X \geq -1\right) \\
&= P\left(\frac{1}{2} \leq -\theta \log X \leq 1\right) = F_y(1) - F_y\left(\frac{1}{2}\right) \\
&= (1 - e^{-1}) - (1 - e^{-\frac{1}{2}}) = \frac{1}{\sqrt{e}} - \frac{1}{e} \approx 0.2386
\end{aligned}$$

5. Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória $X \sim N(\theta, \theta)$. Sugira uma quantidade pivotal para construir um intervalo de confiança para θ.

Temos que

$$Q_1(X, \theta) = \frac{\overline{X} - \theta}{\sqrt{\theta/n}} \sim N(0, 1)$$

Temos ainda que como $Var(X) = \theta$, uma outra quantidade pivotal é dada por

$$Q_2(X, \theta) = \frac{(n-1)S^2}{\theta} \sim \chi^2_{n-1}$$

Segue que

$$Q(X, \theta) = \frac{\frac{\overline{X}-\theta}{\sqrt{\theta/n}}}{\frac{(n-1)S^2}{(n-1)\theta}} = \frac{\sqrt{n}(\overline{X} - \theta)}{\sqrt{\theta}} \frac{\sqrt{\theta}}{S} = \frac{\sqrt{n}(\overline{X} - \theta)}{S} \sim t_{n-1}$$

é possível uma quantidade pivotal para θ.

12. Considere o Exercício 4.9. Obtenha um intervalo de confiança Baysiano para θ com coeficiente de confiança $\gamma = 0.95$. Como fica seu intervalo se $x = 0.4$?

Utilizando os dados do enunciado do Exercício 4.9, temos que

$$f(x,\theta) = \frac{e^{-\theta}\theta^x}{x!}e^{-\theta} = \frac{e^{-2\theta}\theta^x}{x!}$$

Encontrando então a distribuição marginal de X em relação à essa conjunta

$$\begin{aligned} g(x) &= \int_{\Theta} f(x,\theta)\mathrm{d}\theta = \int_0^{\infty} \frac{e^{-2\theta}\theta^x}{x!}\mathrm{d}\theta = \int_0^{\infty} \frac{1}{\Gamma(x+1)}\theta^x e^{-2\theta}\mathrm{d}\theta \\ &= \frac{1}{2^{x+1}}\int_0^{\infty} \underbrace{\frac{2^{x+1}}{\Gamma(x+1)}\theta^x e^{-2\theta}}_{f(\theta),\theta\sim Gama(\lambda=2,r=x+1)} \mathrm{d}\theta = \frac{1}{2^{x+1}} \end{aligned}$$

Assim a densidade da distribuição a posteriori será dada por

$$\pi(\theta|x) = \frac{\frac{e^{-2\theta}\theta^x}{x!}}{\frac{1}{2^{x+1}}} = \frac{2^{x+1}}{\Gamma(x)}\theta^x e^{-2\theta} \Rightarrow \theta|X \sim Gama(\lambda = 2, r = x+1).$$

Para construirmos um intervalo de confiança com coeficiente de confiança γ basta escolhermos q_1 e q_2 tais que:

$$P(q_1 \leq \theta|X \leq q_2) = \gamma$$

No presente caso, em que $x = 0.4$, e $\gamma = 0.95$, escolhemos tomar um intervalo simétrico, de tal forma que $P(\theta|X \leq q_1) = P(\theta|X \geq q_2) = 0.025$. Considerando tal x, a distribuição de θ X fica sendo uma Gama com parâmetros $\lambda = 2$ e $r = 1.4$. Com o auxílio do pacote estatístico R, obtemos:

```
 > qgamma(0.975,shape=1.4,rate=2) #q2
[1] 2.242941
> qgamma(0.025,shape=1.4,rate=2)  #q1
[1] 0.04340487
```

Assim um intervalo de confiança para θ com coeficiente de confiança 0.95 é dado por

$$[0.0434, 2.2429].$$

14. Usando a amostra de tamanho $n = 20$ no Exemplo 3.1.6, construa um intervalo aproximado para θ, onde $f(x|\theta)$ é dada em (3.1.8).

Para construirmos um intervalo de confiança precisamos em primeiro lugar de uma quantidade pivotal. A partir da distribuição assintótica dos estimadores de máxima verossimilhança, sabemos que

$$Q(X,\theta) = \frac{\hat{\theta} - \theta}{\sqrt{(nI_F(\hat{\theta}))^{-1}}}$$

segue aproximadamente uma distribuição normal padrão.

Assim,

$$P\left(-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sqrt{(nI_F(\hat{\theta}))^{-1}}} \leq z_{\alpha/2}\right) \approx \gamma = 1 - \alpha$$

Desenvolvendo essa expressão:

$$\begin{aligned}
\gamma &\approx P\left(\frac{-z_{\alpha/2}}{\sqrt{(nI_F(\hat{\theta}))^{-1}}} \leq \hat{\theta} - \theta \leq \frac{z_{\alpha/2}}{\sqrt{(nI_F(\hat{\theta}))^{-1}}}\right) \\
&= P\left(\frac{-z_{\alpha/2}}{\sqrt{(nI_F(\hat{\theta}))^{-1}}} - \hat{\theta} \leq -\theta \leq \frac{z_{\alpha/2}}{\sqrt{(nI_F(\hat{\theta}))^{-1}}} - \hat{\theta}\right) \\
&= P\left(\hat{\theta} + \frac{z_{\alpha/2}}{\sqrt{(nI_F(\hat{\theta}))^{-1}}} \geq \theta \geq \hat{\theta} - \frac{z_{\alpha/2}}{\sqrt{(nI_F(\hat{\theta}))^{-1}}}\right) \\
&= P\left(\hat{\theta} - \frac{z_{\alpha/2}}{\sqrt{(nI_F(\hat{\theta}))^{-1}}} \leq \theta \leq \hat{\theta} + \frac{z_{\alpha/2}}{\sqrt{(nI_F(\hat{\theta}))^{-1}}}\right)
\end{aligned}$$

E portanto um intervalo de confiança aproximado para θ será dado por

$$\left[\hat{\theta} - \frac{z_{\alpha/2}}{\sqrt{(nI_F(\hat{\theta}))^{-1}}}; \hat{\theta} + \frac{z_{\alpha/2}}{\sqrt{(nI_F(\hat{\theta}))^{-1}}}\right] \tag{1}$$

Não iremos utilizar a amostra de tamanho 20 fornecida no Exemplo 3.1.6 pois ela está com problemas. Uma das observações é -3.4919, mas a variável aleatória só está definida em $[-1, 1]$.

Vamos entretanto implementar um conjunto de funções em R que nos permitam trabalhar com a função de densidade definida em 3.1.8:

```
> rbolfarine

function (n, theta)
{
    sample <- runif(n)
    (-1 + 2 * sqrt(0.25 - theta * (0.5 - 0.25 * theta - sample)))/theta
}
```

Essa função simula valores da variável aleatória definida em 3.1.8.

```
> inffisher

function (theta.j)
{
    1/(2 * theta.j^3) * (log((1 + theta.j)/(1 - theta.j)) - 2 *
        theta.j)
}
```

```
> mmvapprox2

function (x, teta.0 = mean(x), maxeps = 1e-04)
{
    eps <- maxeps + 1
    while (eps > maxeps) {
        teta.i <- teta.0 + sum(x/(1 + teta.0 * x))/sum(x^2/(1 +
            teta.0 * x)^2)
        eps <- abs(teta.i - teta.0)
        teta.0 <- teta.i
    }
    teta.i
}
```

As duas funções acima, respectivamente, calculam a informação de Fisher para um dado teta e aproximam um valor de teta através de iterações sucessivas.

```
> IC.complete

function (x, conf = 0.95, teta = NULL)
{
    teta.hat <- mmvapprox2(x, maxeps = 1e-07)
    lower <- teta.hat + qnorm((1 - conf)/2)/sqrt(length(x) *
        inffisher(teta.hat))
    upper <- teta.hat - qnorm((1 - conf)/2)/sqrt(length(x) *
```

```
        inffisher(teta.hat))
    if (is.null(teta)) {
        c(lower, upper)
    }
    else {
        if (teta >= lower && teta <= upper) {
            1
        }
        else {
            0
        }
    }
}
```

Por fim, a função acima gera um intervalo de confiança, baseado na expressão 1. Os parâmetros defininem o coeficiente de confiança e a amostra a ser levada em conta. Há ainda um parâmetro adicional, `teta`, que define o comportamento da função: se `teta` é algum valor não nulo então a função retorna 1 se `teta` está no intervalo de confiança criado, ou caso contrário retorna 0. Caso `teta` não seja especificado, IC retorna um intervalo de confiança.

Geremos então uma amostra de tamanho $n = 50$ e $\theta = 0.4$ e vamos construir um intervalo de confiança com $\gamma = 0.95$:

```
> sample <- rbolfarine(50, 0.4)
> IC.complete(sample)

[1] -0.2796992  0.6694328
```

Vamos agora fazer uma simulação para verificar se os intervalos de confiança que estamos gerando estão corretos.

```
> sum(apply(matrix(rbolfarine(1000 * 1000, 0.4), ncol = 1000),
+     2, IC.complete, conf = 0.25, teta = 0.4))/1000

[1] 0.258
```

Nesse primeiro exemplo simulamos 1000 amostras de tamanho 1000 e construimos 1000 intervalos de confiança, com coeficiente de confiança 0.25. O resultado mostrado é a proporção observada dos intervalos que continham o parâmetro θ.

```
> sum(apply(matrix(rbolfarine(1000 * 1000, 0.4), ncol = 1000),
+     2, IC.complete, conf = 0.95, teta = 0.4))/1000
```

```
[1] 0.957
```

Mesma simulação, mas com coeficiente de confiança 0.95. Como os valores estão próximos, concluimos que nosso intervalo de confiança foi construido corretamente.

Sobre

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