Lista 3 - Inferência Estatística

Livro Robert Keener

3.7.1.

a)
$$\begin{aligned} b(\theta,\delta) &= E_\theta[\delta(\mathbf{X})]-\theta \\ &= E_\theta\left[a\frac{X}{n}+(1-a)b\right]-\theta \\ &= a\,E_\theta[X]/n+(1-a)b-\theta \\ &= a\,\theta+(1-a)b-\theta \\ &= (1-a)(b-\theta) \end{aligned}$$

$$\begin{aligned} Var_\theta(\delta) &= Var\left(a\frac{X}{n}+(1-a)b\right) \\ &= a^2 Var(X)/n^2 \\ &= a^2\theta(1-\theta)/n \end{aligned}$$

b) Segundo a dica, considere a perda quadrática. Assim

$$\begin{aligned} R(\theta,\delta) &= E_\theta[(\delta-\theta)^2] \\ &= Var_\theta(\delta(\mathbf{X}))+b(\theta,\delta)^2 \\ &= \frac{a^2\theta(1-\theta)}{n}+(1-a)^2(b-\theta)^2 \\ &> \theta^2(1-\theta)/n, \end{aligned}$$

fazendo $a=1$. Assim, para $a>1$, $\delta_{a,b}$ é inadmissível.

c) Fixe $a\in[0,1)$. Para b fixo, a variância independe de b. Observe que se $b>1$, $b-\theta>1-\theta \Rightarrow (b-\theta)^2>(1-\theta)^2$ $\Rightarrow R(\theta,\delta_{a,b})>R(\theta,\delta_{a,1}) \Rightarrow \delta_{a,b}$ é inadmissível. Ademais, se $b<0$, $b-\theta<-\theta \Rightarrow (b-\theta)^2>\theta^2 \Rightarrow$ $R(\theta,\delta_{a,b})>R(\theta,\delta_{a,0}) \Rightarrow \delta_{a,b}$ é inadmissível.

d) Se $a<0$, note que $(1-a)^2 > 1 \Rightarrow b^2(\theta, \delta_{a,b})$ é maior que $b^2(\theta, \delta_{a,0})$. Além do mais, a variância também é maior. Portanto $R(\theta, \delta_{a,b}) > R(\theta, \delta_{0,b})$ e $\delta_{a,b}$ é inadmissível.

4.7.1. $X_i \overset{iid}{\sim} Exp(\lambda_x)$, $Y_i \overset{iid}{\sim} Exp(\lambda_y)$. Note que

a) $f_{\lambda_x,\lambda_y}(\mathbf{x},\mathbf{y}) = \lambda_x^m \lambda_y^n \exp\{-\lambda_x \sum_{i=1}^m x_i - \lambda_y \sum_{i=1}^n y_i\}$,

que pertence à família exponencial. Portanto,

$$T(\mathbf{x},\mathbf{y}) = \left(\sum_{i=1}^m x_i, \sum_{i=1}^n y_i\right)$$

é estatística suficiente completa para (λ_x, λ_y). Sabemos que $\sum_{i=1}^n y_i \sim Gamma(n, \lambda_y)$ e $\sum_{i=1}^m x_i \sim Gamma(m, \lambda_x)$

Portanto,

$$Z = \frac{\overbrace{\lambda_y \sum_{i=1}^n y_i}^{Gamma(n,1)}}{\lambda_x \sum_{i=1}^m x_i} \sim \beta'(n, m), \quad m \geq 2!$$

em que sabemos que $E[Z] = n/(m-1)$. Portanto

$$E\left[\frac{\lambda_x}{\lambda_y} \frac{(m-1)}{n} Z\right] = \frac{\lambda_x}{\lambda_y}.$$

Com isso $\delta(\mathbf{x},\mathbf{y}) = (m-1)/n \cdot \sum_{i=1}^n y_i / \sum_{i=1}^m x_i$ é estimador não enviesado. Por Lehmann-Scheffé, como δ é função de T, uma estatística suficiente completa, temos que δ é UMVUE.

b) $$\delta = c\frac{\bar{Y}}{\bar{X}} = c\frac{m}{n}\frac{\sum Y_i}{\sum X_i} = c\frac{m}{n}\frac{\lambda_x}{\lambda_y} Z$$

Com isso, podemos calcular viés e variância

$$b\left(\frac{\lambda_x}{\lambda_y}, \delta\right) = c\frac{m}{n}\frac{\delta_x}{\delta_y}E[Z] - \frac{\delta_x}{\delta_y} = c\frac{m}{\cancel{n}}\frac{\delta_x}{\delta_y}\frac{\cancel{n}}{m-1} - \frac{\delta_x}{\delta_y}$$

$$= \left(\frac{cm}{m-1} - 1\right)\frac{\delta_x}{\delta_y}$$

$$\mathrm{Var}_{\frac{\lambda_x}{\lambda_y}}(\delta) = c^2 \frac{m^2}{n^2}\frac{\lambda_x^2}{\lambda_y^2}\mathrm{Var}(Z)$$

$$= c^2 \frac{m^2}{n^{\cancel{2}}}\frac{\lambda_x^2}{\lambda_y^2}\frac{\cancel{n}(n+m-1)}{(m-1)^2(m-2)}.$$

Com isso

$$R\left(\frac{\lambda_x}{\lambda_y}, \delta\right) = \frac{\lambda_x^2}{\lambda_y^2}\left[c^2\left(\frac{m^2}{n}\frac{(n+m-1)}{(m-2)(m-1)^2} + \frac{m^2}{(m-1)^2}\right) - \frac{2cm}{m-1} + 1\right]$$

que é minimizado quando

$$c = \frac{\frac{\cancel{m}}{\cancel{m-1}}}{\frac{m^{\cancel{2}}(n+m-1)}{n(m-2)(m-1)^{\cancel{2}}} + \frac{m^{\cancel{2}}}{(m-1)^{\cancel{2}}}} = \frac{\cancel{(m-1)}(m-2)n}{\cancel{(m-1)}\,m(n+1)}$$

c) $\mathbb{P}(X_1 > 1) = \mathbb{E}[\mathbb{1}\{X_1 > 1\}]$. Logo o estimador

$$\psi(x) = \mathbb{1}\{X_i > 1\}$$

é não enviesado. Com isso, o estimador de Rao-Blackwell

$$\delta(t) = \mathbb{E}[\psi(x) \mid \textstyle\sum x_i = t]$$

é UMVUE. Veja que

$$\delta(t) = \mathbb{P}(X_1 > 1 \mid \textstyle\sum x_i = t).$$

$$f_{X_1|T}(x|t) = \frac{f_{X_1, T-X_1}(x, t-x)}{f_T(t)}.$$

$X_2 + \dots + X_n \sim \text{Gamma}(n-1, \lambda_x)$ e $T \sim \text{Gamma}(n, \lambda_x)$

$$f_{X_1|T}(x|t) = \frac{\frac{\cancel{(\lambda_x)^n}}{\Gamma(n-1)}(t-x)^{n-2}\cancel{e^{-\lambda_x(t-x)}}\cancel{e^{-\lambda_x x}}}{\cancel{(\lambda_x)^n}\, t^{n-1}\cancel{e^{-\lambda_x t}}/\Gamma(n)}$$

$$= (n-1)(t-x)^{n-2}/t^{n-1} = (n-1)\left(1-\frac{x}{t}\right)^{n-2}/t, \quad x < t.$$

Logo, $S(x) = \mathbb{P}(X_1 > 1 \mid \sum X_i = t)$

$$= \int_1^t \frac{(n-1)(1-\frac{x}{t})^{n-2}}{t}\,dx$$

$$= (n-1)\int_0^{\frac{t-1}{t}} v^{n-2}\,dv$$

$$= \left(\frac{t-1}{t}\right)^{n-1} \mathbb{1}\{T \geq t\}$$

4.7.5. $X_1, \dots, X_n \overset{iid}{\sim} Q_\theta$.

a) $$E[S^2] = \frac{n}{n-1} E[(X_1 - \bar{X})^2]$$

$$= \frac{n}{n-1}\left(\mathbb{E}[X_1^2] - 2\mathbb{E}[X_1\bar{X}] + \mathbb{E}[\bar{X}^2]\right)$$

$$= \frac{n}{n-1}\left(E[X_1^2] + Var(\bar{X}) + E[\bar{X}]^2 - \frac{2}{n}\sum_{j=1}^n E[X_1X_j]\right)$$

$$= \frac{n}{n-1}\left(E[X_1^2] + \frac{1}{n}Var(X_1) + E[X_1]^2 - \frac{2}{n}E[X_1^2] - \frac{2(n-1)}{n}E[X_1]^2\right)$$

$$= \frac{n}{n-1}\left(\cancel{E[X_1^2]} + \cancel{\frac{E[X_1^2]}{n}} - \frac{E[X_1]^2}{n} + E[X_1]^2 - \frac{2}{\cancel{n}}\cancel{E}[X_1^2] - \frac{2(n-1)}{n}E[X_1]^2\right)$$

$$= \frac{n}{n-1}\left(\frac{n-1}{n}E[X_1^2] - \frac{n-1}{n}E[X_1]^2\right)$$

$$= Var_\theta(X_1)$$

b) $S^2 = (n-1)^{-1}\left(\sum_{i=1}^n X_i^2 - \bar{X}^2\right)$. No caso Bernoulli, $X_i^2 = X_i$. Logo, definindo $T(x) = \sum_{i=1}^n X_i$,

$$S^2 = (n-1)^{-1}(T - T^2/n^2)$$

é função de T, que é estatística suficiente completa, pois

$$f_\theta(x_1, \dots, x_n) = \theta^T(1-\theta)^{n-T}$$
$$= \exp\{T(\log \theta/1-\theta)\}(1-\theta)^n$$

pertence à família exponencial, com $\mathbb{R}$ sendo o espaço dos parâmetros.

4.7.28. $X_1, \dots, X_n \sim \mathcal{U}nif(0, \theta)$

a) O limite inferior para $\mathrm{Var}_\theta(\delta)$, com $E_\theta[\delta(\mathbf{X})] = \theta$, é

$$\frac{(\theta + \Delta - \theta)^2}{E_\theta\left[\frac{p_{\theta+\Delta}(x)}{p_\theta(x)} - 1\right]^2},$$

com o fato que $p_\theta(x) = 0 \Rightarrow p_{\theta+\Delta}(x) = 0$ e, portanto $\Delta < 0$.
Note que

$$\frac{p_{\theta+\Delta}(x)}{p_\theta(x)} - 1 = \frac{(\theta+\Delta)^{-n} \mathbb{1}\{x_{(n)} < \theta + \Delta\}}{\theta^{-n}} - 1$$

$$= \frac{\theta^n \mathbb{1}\{x_{(n)} < \theta + \Delta\} - (\theta+\Delta)^n}{(\theta+\Delta)^n}$$

$$= \begin{cases} \theta^n/(\theta+\Delta)^n - 1, & \text{se } x_{(n)} < \theta + \Delta \\ -1, & \text{c.c.} \end{cases}$$

Com isso, a esperança do quadrado é

$$\left(\frac{\theta^n}{(\theta+\Delta)^n} - 1\right)^2 P(x_{(n)} < \theta + \Delta) + P(X_{(n)} > \theta + \Delta).$$

Temos que $P(x_{(n)} < \theta + \Delta) = \prod_{i=1}^n P(X_i < \theta + \Delta)$
$= \left(\frac{\theta+\Delta}{\theta}\right)^n.$

Com isso, o limite inferior é

$$\frac{\Delta^2 \cancel{(\theta+\Delta)^n}}{\frac{\theta^n}{(\theta+\Delta)^n} - \cancel{(\theta+\Delta)^n}\, 1},$$

com $\Delta \in (-\theta, 0)$.

$$\frac{c^2\theta^2/n^2}{\frac{\theta^n}{\left(\theta - \frac{c\theta}{n}\right)^n} - 1}$$

b) Faça $\Delta = -c\theta/n$. Assim, o limite inferior é

$$\frac{c^2\theta^2/n^2}{(1 - c/n)^{-n} - 1} = \frac{\theta^2}{n^2}\left[\frac{c^2}{(1-c/n)^{-n} - 1}\right]$$

$$g(c) = \lim_{n\to\infty} g_n(c) = c^2/e^c - 1$$

c) $g(c_0) := \max_{x \in (0,1)} \frac{x^2}{e^x - 1}$,

$$\frac{d}{dx} g(x) = \frac{(e^x-1)2\cancel{x} - x^{\cancel{2}} e^x}{(e^x-1)^2} = 0$$

$$\Rightarrow e^x(2-x) = 2$$

$$\Rightarrow 1 - x/2 = e^{-x}.$$

Essa igualdade não é satisfeita para nenhum $x \in (0,1)$. Além disso, $g'(x) \geq 0 \Rightarrow c_0 = 1$.

Livro Casella-Berger

6.36. a) Note que $T_2 = f(T_1)$. Assim,

$$\begin{aligned} E[U_1 | T_2] &= E[E[U | T_1] | T_2] \\ &= E[U | T_2] \\ &= U_2. \end{aligned}$$

Propriedade da Torre

b)
$$\begin{aligned} Var(U_1) &= E[Var(U_1|T_2)] + Var(E[U_1|T_2]) \\ &= E[Var(U_1|T_2)] + Var(U_2) \\ &\geq Var(U_2), \end{aligned}$$

pois $Var(U_1|T_2) \geq 0$.

Extra

1. Resolvido em 4.7.5.

2. $E[S_c(\mathbf{X})] = c(n-1)\sigma^2$

Note que $c^{-1}S_c(\mathbf{X})/\sigma^2 \sim \chi^2_{n-1}$. Assim

$Var(S_c(\mathbf{X})) = c^2\sigma^4 \, Var(S_c(\mathbf{x})/c\sigma^2) = 2(n-1)c^2\sigma^4.$

Portanto

$$R(\sigma^2, \delta_c) = 2(n-1)\, c^2 \sigma^4 + [c(n-1) - 1]^2 \sigma^4$$
$$= \left(c^2 (n^2-1) - 2c(n-1) + 1 \right) \sigma^4$$

Minimizamos R em $c = \dfrac{(n-1)}{n^2-1} = \dfrac{1}{n+1}$.

Mas o viés é não nulo.

3. Baseado no livro de Robert Keener.