Primeira avaliação (A1)

Disciplina: Inferência Estatística
Instrutor: Luiz Max Carvalho
Monitores: Jairon Nóia & Tiago Silva

24 de Setembro de 2022

- O tempo para realização da prova é de 3 horas;
- Leia a prova toda com calma antes de começar a responder;
- Responda todas as questões sucintamente;
- Marque a resposta final claramente com um quadrado, círculo ou figura geométrica de sua preferência;
- A prova vale 80 pontos. A pontuação restante é contada como bônus;
- Apenas tente resolver a questão bônus quando tiver resolvido todo o resto;
- Você tem direito a trazer **uma folha de "cola"** tamanho A4 frente e verso, que deverá ser entregue junto com as respostas da prova.

1. Treasure map.

Suponha que temos um modelo estatístico paramétrico, com f.d.p./f.m.p. $f_\theta(x)$, $\theta \in \Omega \subseteq \mathbb{R}^p$ com suporte em $\mathcal{X} \subseteq \mathbb{R}^d$. Dada uma observação $\boldsymbol{X} = \boldsymbol{x}$, o chamado estimador *maximum a posteriori*, MAP, é definido como

$$\delta_{\text{MAP}}(\boldsymbol{X}) = \arg\max_{\theta \in \Omega} \xi(\theta \mid \boldsymbol{x}).$$

a) (10 pontos) Mostre que quando $\Omega = \{\theta_1, \theta_2, \ldots, \theta_k\}$, $k \geq 2$, isto é, quando o espaço de parâmetros é discreto, δ_{MAP} é o estimador de Bayes sob a seguinte perda

$$L(\delta, \theta) = \begin{cases} 0, \delta = \theta, \\ 1, \text{caso contrário}, \end{cases}$$

também chamada de perda *zero-um* (*0-1* loss).

b) (10 pontos) Suponha que a proporção θ de itens defeituosos em uma linha de produção toma apenas os valores $0,1$ e $0,2$. Suponha ainda que n itens são inspecionados e x são defeituosos, $x \in \{0, 1, \ldots, n\}$. Mostre como encontrar o estimador de máxima verossimilhança para θ;

c) (10 pontos) Suponha que, *a priori*, $\Pr(\theta = 0,1) =: \pi(0,1) = 0,7$. Exiba a distribuição *a posteriori* de θ e mostre como encontrar o MAP para θ.

Conceitos trabalhados: estimador de Bayes; perda; estimador de máxima verossimilhança. **Nível de dificuldade**: fácil.

Resolução: Primeiro, vamos escrever a perda esperada:

$$\begin{aligned} E_{\theta|\boldsymbol{x}}[L(\delta, \theta)] &= \sum_{t \in \Omega} L(\delta, t)\xi(t \mid \boldsymbol{x}), \\ &= \sum_{t^\star \neq \delta} \xi(t^\star \mid \boldsymbol{x}), \\ &= P_{\theta|\boldsymbol{x}}(\theta \neq \delta). \end{aligned}$$

Deste modo, o estimador de Bayes é estimador que minimiza esta perda:

$$\begin{aligned} \delta_{\text{MAP}} &= \arg\min_{d \in \mathcal{D}} E_{\theta|\boldsymbol{x}}[L(d, \theta)], \\ &= \arg\min_{d \in \Omega} P_{\theta|\boldsymbol{x}}(\theta \neq d), \\ &= \arg\min_{d \in \Omega} \left\{1 - P_{\theta|\boldsymbol{x}}(\theta = d)\right\}, \\ &= \arg\max_{d \in \Omega} P_{\theta|\boldsymbol{x}}(\theta = d), \\ &= \arg\max_{d \in \Omega} \xi(d \mid \boldsymbol{x}), \end{aligned}$$

como queríamos demonstrar. Note que é natural fazer $\mathcal{D} = \Omega$ porque a perda é 1 para todo valor fora de Ω. Para encontrar o EMV pedido em b), primeiro vamos formular um modelo probabilístico para os dados. É razoável afirmar que o número de itens defeituosos tem distribuição binomial com n tentativas e probabilidade de sucesso $\theta \in \{0.1, 0.2\}$, isto é, com um espaço paramétrico discreto em lugar do usual $\Omega = (0, 1)$:

$$f(x \mid n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \theta \in \left\{ \frac{1}{10}, \frac{2}{10} \right\}, x = 0, 1, \ldots, n.$$

Assim, temos que

$$\begin{aligned}
\delta_{\text{EMV}} &= \underset{\theta \in \left\{\frac{1}{10}, \frac{2}{10}\right\}}{\arg\max} \binom{n}{x} \theta^x (1-\theta)^{n-x}, \\
&= \underset{\theta \in \left\{\frac{1}{10}, \frac{2}{10}\right\}}{\arg\max}\ \theta^x (1-\theta)^{n-x}.
\end{aligned}$$

Poderíamos parar por aqui, mas vamos explorar essa questão mais um pouco. Escreva $f_1(x) = f(x \mid n, 0.1)$ e $f_2(x) = f(x \mid n, 0.2)$ e note que

$$r(x) := \frac{f_1(x)}{f_2(x)} < 1 \iff \delta_{\text{EMV}} = 0.2.$$

Reescrevendo a verossimilhança como $f(x \mid n, \theta) \propto (\frac{\theta}{1-\theta})^x (1-\theta)^n$, temos

$$\begin{aligned}
\frac{(\frac{1}{9})^x (\frac{9}{10})^n}{(\frac{1}{4})^x (\frac{8}{10})^n} < 1 &\iff \delta_{\text{EMV}} = 0.2, \\
\left(\frac{4}{9}\right)^x \left(\frac{9}{8}\right)^n < 1 &\iff \delta_{\text{EMV}} = 0.2, \\
\frac{\log\left(\frac{4}{9}\right)}{\log\left(\frac{9}{8}\right)} < -\frac{n}{x} &\iff \delta_{\text{EMV}} = 0.2, \\
\frac{n}{x} > 6.884949 &\iff \delta_{\text{EMV}} = 0.2,
\end{aligned}$$

com o compromisso de que vamos definir $n/0 = \infty > 6.884949$. Agora vamos resolver c), computando primeiro a distribuição *a posteriori* de θ:

$$\begin{aligned}
\xi(\theta = 0.1 \mid \boldsymbol{x}) &= \frac{f_1(x)\pi(0.1)}{f_1(x)\pi(0.1) + f_2(x)[1 - \pi(0.1)]}, \\
&= \frac{0.7\left(\frac{1}{10}\right)^x \left(\frac{9}{10}\right)^n}{0.7\left(\frac{1}{10}\right)^x \left(\frac{9}{10}\right)^n + 0.3\left(\frac{2}{10}\right)^x \left(\frac{8}{10}\right)^n}.
\end{aligned}$$

Fazendo $\pi_1 = \pi(0.1)$ e $\pi_2 = \pi(0.2)$, temos que de modo análogo ao que foi feito para o EMV:

$$\begin{aligned}
\frac{\xi(\theta = 0.1 \mid \boldsymbol{x})}{\xi(\theta = 0.2 \mid \boldsymbol{x})} < 1 &\iff \delta_{\text{MAP}} = 0.2, \\
\frac{f_1(x)\pi_1}{f_2(x)\pi_2} < 1 &\iff \delta_{\text{MAP}} = 0.2, \\
\frac{f_1(x)}{f_2(x)} \frac{\pi_1}{1 - \pi_1} < 1 &\iff \delta_{\text{MAP}} = 0.2, \\
\left(\frac{4}{9}\right)^x \left(\frac{9}{8}\right)^n \times \frac{7}{3} < 1 &\iff \delta_{\text{MAP}} = 0.2,
\end{aligned}$$

ou seja, a razão de verossimilhanças precisa ser ajustada pela razão entre as probabilidades *a priori*, ou *chance a priori* neste caso ($k = 2$). Por exemplo, quando $x = 3$, $\delta_{\text{EMV}} = 0.2$, mas $\delta_{\text{MAP}} = 0.1$. ■

Comentário: Esta questão se baseia no exercício 2 da seção 7.2 de De Groot (recomendado!) e exercício de revisão feito em aula. Neste exercício nós aplicamos a definição de estimador de Bayes para de fato nos convencer de que, uma vez estabelecida uma função de perda e uma priori (e portanto, uma posteriori), somos sempre capazes de encontrar (pelo menos) um estimador de Bayes. Além disso, vimos que é possível criar um estimador bayesiano que maximiza a posteriori em vez de integrar com respeito a ela, bem ao feitio da estatística clássica. Note que adicionar uma distribuição sobre Ω pode e em geral vai levar a inferências diferentes do que se obteria puramente usando a verossimilhança.

2. Now, Dinah, tell me the truth.

Tome $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{N}^n$ um conjunto de realizações de uma variável aleatória X distribuída conforme a distribuição de Poisson, com f.m.p.

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

com taxa $\theta \in \mathbb{R}_+$ desconhecida.

a) (10 pontos) Defina $\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$. Verifique que os estimadores

$$\delta_1(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} X_i \text{ e } \delta_2(\mathbf{X}) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2,$$

são não viesados de θ.

b) (10 pontos) Mostre que δ_1 é eficiente. Ele é consistente?

c) (10 pontos) Suponha que $n = 2$. Mostre que δ_2 é **inadmissível**.

Dica: Se X e Y são variáveis aleatórias com distribuição de Poisson com média θ, então

$$E_\theta \left[\frac{(X - Y)^4}{4} \right] = 3\theta^2 + \frac{\theta}{2}.$$

d) (10 pontos) A informação de Fisher quantifica a informação sobre um parâmetro contida em uma amostra aleatória. Compute a informação de Fisher, $I_X(\theta)$, em $\mathbf{X}$.

e) (10 pontos) A parametrização é crucial para a informação de Fisher; existem transformações de variáveis que mudam substancialmente a sua interpretação. Sendo assim, prove que para o modelo Poisson, a informação de Fisher em $\mathbf{X}$ sobre $\eta = \sqrt{\theta}$ é constante.

Dica: Se $\eta = g(\theta)$, então $I_X(\theta) = I_X(\eta)|g'(\theta)|^2$.

Conceitos trabalhados: informação de Fisher; estimador de máxima verossimilhança; reparametrização; admissibilidade; eficiência.
Nível de dificuldade: médio.
Resolução: Para facilitar, vamos estabelecer alguns fatos antes de começar a resolução. Primeiro, se W é uma variável aleatória com distribuição Poisson com taxa λ, então

$$\begin{aligned} E_\theta[W^2] &= \mathrm{Var}_\theta(W) + (E_\theta[W])^2, \\ &= \theta + \theta^2, \\ &= \theta(1+\theta). \end{aligned}$$

Além disso, se $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ é um vetor de observações com média amostral $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, então

$$\begin{aligned} S_2 := \sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2, \\ &= \sum_{i=1}^n X_i^2 - 2\bar{X}_n \sum_{i=1}^n X_i + n\left(\bar{X}_n\right)^2, \\ &= \sum_{i=1}^n X_i^2 - n\left(\bar{X}_n\right)^2, \end{aligned}$$

onde a penúltima linha segue do fato de que $S_n = \sum_{i=1}^n X_i = n\bar{X}_n$. Armados destes fatos, vamos responder a). Primeiro,

$$\begin{aligned} E_\theta[\delta_1(\boldsymbol{X})] &= \frac{1}{n}\sum_{i=1}^n X_i, \\ &= \frac{n\theta}{n} = \theta, \end{aligned}$$

portanto δ_1, que é o estimador de máxima verossimilhança de θ, é não-viesado. Agora,

$$\begin{aligned} E_\theta[\delta_2(\boldsymbol{X})] &= \frac{1}{n-1}\left\{\sum_{i=1}^n X_i^2 - n\left(\bar{X}_n\right)^2\right\}, \\ &= \frac{1}{n-1}\left\{n\theta(1+\theta) - nE_\theta\left[\left(\bar{X}_n\right)^2\right]\right\}, \\ &= \frac{1}{n-1}\left\{n\theta(1+\theta) - \frac{n}{n^2}E_\theta\left[\left(\sum_{i=1}^n X_i\right)^2\right]\right\}, \\ &= \frac{1}{n-1}\left\{n\theta(1+\theta) - \frac{n}{n^2}n\theta(1+n\theta)\right\}, \\ &= \frac{\theta}{n-1}\left\{n(1+\theta) - (1+n\theta)\right\}, \\ &= \theta. \end{aligned}$$

Para b) e d), precisamos calcular $I_n(\theta) = nI(\theta)$, e podemos fazer assim porque nossa amostra é i.i.d. Assim,

$$\begin{aligned}
I(\theta) &= E_\theta\left[-\frac{\partial^2\theta}{\partial\theta^2}\log f_n(\boldsymbol{x} \mid \theta)\right],\\
&= E_\theta\left[-\frac{-\sum_{i=1}^n X_i}{\theta^2}\right],\\
&= \frac{1}{\theta},
\end{aligned}$$

de modo que $I_n(\theta) = n/\theta$ e portanto o limite inferior de Cramér-Rao para a variância de estimadores não-viesados de θ é $1/I_n(\theta) = \theta/n$. Precisamos verificar se a variância de δ_1 "encaixa" nessa cota.

$$\begin{aligned}
\operatorname{Var}_\theta(\delta_1(\boldsymbol{X})) &= \operatorname{Var}_\theta(\bar{X}_n),\\
&= \frac{\operatorname{Var}_\theta(\sum_{i=1}^n X_i)}{n^2},\\
&= \frac{n\theta}{n^2},
\end{aligned}$$

o que mostra que a cota inferior é alcançada e, portanto, δ_1 é eficiente. Note também que $\lim_{n\to\infty}\bar{X}_n = \theta$, portanto δ_1 também é consistente. Outra maneira de ver isso é perceber que o viés de δ_1 é zero e sua variância converge para zero assintoticamente[1]. Agora, vamos resolver c), o que implica computar a variância de δ_2, $\operatorname{Var}_\theta(\delta_2(\boldsymbol{X})) = E_\theta\left[\{\delta_2(\boldsymbol{X})\}^2\right] - \theta^2$. Vamos reescrever δ_2 para nos facilitar a vida:

$$\begin{aligned}
\delta_2(\boldsymbol{X}) &= \left(X_1 - \frac{X_1+X_2}{2}\right)^2 + \left(X_2 - \frac{X_1+X_2}{2}\right)^2,\\
&= \left(\frac{X_1-X_2}{2}\right)^2 + \left(\frac{X_2-X_1}{2}\right)^2,\\
&= 2\frac{(X_1-X_2)^2}{4} = \frac{(X_1-X_2)^2}{2}.
\end{aligned}$$

Estamos em posição de calcular

$$\begin{aligned}
\operatorname{Var}_\theta(\delta_2(\boldsymbol{X})) &= E_\theta\left[\{\delta_2(\boldsymbol{X})\}^2\right] - \theta^2,\\
&= E_\theta\left[\left\{\frac{(X_1-X_2)^2}{2}\right\}^2\right] - \theta^2,\\
&= 3\theta^2 + \frac{\theta}{2} - \theta^2,\\
&= 2\theta^2 + \frac{\theta}{2},
\end{aligned}$$

onde a penúltima igualdade segue da dica dada. Como $2\theta^2$ é positivo necessariamente, concluímos que δ_2 não é eficiente; como demonstramos que existe outro estimador não-viesado que é de fato eficiente, somos forçados a concluir que $R(\delta_2,\theta) > R(\delta_1,\theta)$ para todo θ e, que portanto, δ_2 não é admissível. A Figura 1 mostra um esboço da distribuição de δ_1 e δ_2 para $n = 2$ e $\theta_0 = \zeta(3) \approx 1.2021$. Para finalizar e responder e), vamos usar a dica para calcular $g'(\theta) = 1/2\sqrt{\theta}$, o

[1] Lembre-se: convergência em média quadrática implica convergência em probabilidade, como vimos na revisão de Probabilidade no início do curso.

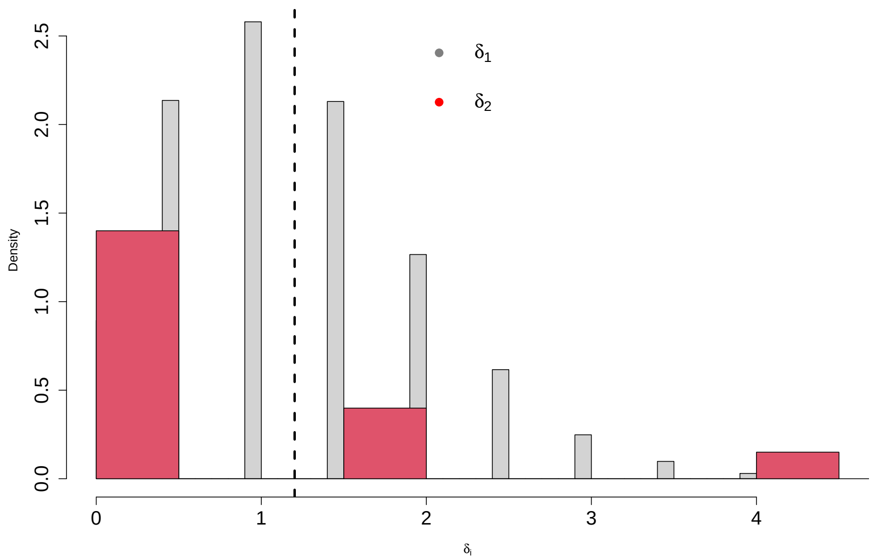

Figura 1: **Distribuição amostral dos estimadores δ_1 e δ_2 no caso Poisson.** Mostramos os histogramas para $N = 5000$ simulações de Monte Carlo com $\theta_0 = 1.2021$ – marcado pela linha tracejada vertical. Para estas simulações, $\mathrm{Var}(\delta_1) \approx 0.60$, enquanto $\mathrm{Var}(\delta_2) \approx 3.49$.

que nos leva a

$$\frac{1}{\theta} = \frac{I_X(\eta)}{4\theta} \implies I_X(\eta) = 4,\ \forall \theta \in (0, \infty),$$

o que de fato é constante com relação ao parâmetro, θ. Para ter certeza, vamos reparametrizar a p.m.f. e proceder aos cálculos. Note que $\theta = \eta^2$, portanto,

$$f_\eta(x) = \frac{\eta^{2x} e^{\eta^2}}{x!},$$

de modo que

$$\log f_\eta(x) = 2x \log(\eta) - \eta^2,$$

e

$$\frac{\partial^2}{\partial \eta^2} \log f_\eta(x) = -\frac{2x}{\eta^2} - 2.$$

Tomando menos a esperança da expressão acima, temos

$$I_X(\eta) = \frac{2E_\eta[x]}{\eta^2} + 2 = \frac{2\eta^2}{\eta^2} + 2 = 4.$$

■

Comentário: Extraído do exercício 3, Seção 8.8, do DeGroot, e dos exemplos 8.7.5 e 8.8.8. A Seção 3.7 do livro The Bayesian Choice contém os detalhes do item (e). O interessante aqui é que como na Poisson a média e a variância coincidem em termos do valor do parâmetro, então podemos criar dois estimadores de momentos, baseados na média e variância amostrais, respectivamente, para tentar estimar θ. No entanto, apenas um deles será eficiente, o que nesse caso coincide com o estimador de máxima verossimilhança. Vimos também que a parametrização influencia fortemente a forma da informação de Fisher, e este é um fato que pode ser explorado. Mais à frente no curso veremos as chamadas transformações estabilizadoras da variância e o método Delta, que ilustram esse ponto.

3. *Actually*, Beta wolves don't exist in the wild[2].

Em várias aplicações estatísticas os dados nos são apresentados na forma de proporções. Um bom exemplo é a proporção de óleo bruto que é convertida em gasolina depois da destilação e fracionamento. Para modelar estes dados é preciso escolher um modelo apropriado. A distribuição Beta é uma família de distribuições contínuas com suporte em $(0, 1)$, cuja densidade (no suporte) vale

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1},$$

para $a, b > 0$. Para esta distribuição, sabemos que $E_\theta[X] = a/(a+b)$ e $\text{Var}_\theta(X) = ab/[(a+b)^2(a+b+1)]$. Tome $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ uma amostra aleatória de uma distribuição Beta com parâmetros a e b.

a) (10 pontos) Encontre uma estatística suficiente para a quando b é conhecida.

b) (10 pontos) Encontre o estimador de máxima verossimilhança para $E_\theta[X]$ quando $b = 1$, conhecido.

Conceitos trabalhados: Teorema da fatorização; estimador de máxima verossimilhança; invariância; suficiência.
Nível de dificuldade: médio.
Resolução: Vamos resolver a) escrevendo a verossimilhança e depois aplicando o Teorema da Fatorização. Assim, como nossa amostra é aleatória

$$\begin{aligned}
f_n(\boldsymbol{x} \mid a, b) &= \prod_{i=1}^n f(x_i; a, b), \\
&= \left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right)^n \left(\prod_{i=1}^n x_i\right)^{a-1} \left(\prod_{i=1}^n 1 - x_i\right)^{b-1}, \\
&\propto \left(\frac{\Gamma(a+b)}{\Gamma(a)}\right)^n \left(\prod_{i=1}^n x_i\right)^{a},
\end{aligned}$$

de onde conseguimos facilmente identificar $T(\boldsymbol{X}) = \prod_{i=1}^n X_i$ como estatística suficiente para a quando b é conhecida. Para b) queremos o EMV para $\mu =$

[2] https://sciencenorway.no/ulv/wolf-packs-dont-actually-have-alpha-males-and-alpha-females-the-idea-is-based 1850514

$E_\theta[X] = a/(a+b)$ quando $b = 1$, isto é $\mu = a/(a+1)$. A primeira providência é escrever

$$L_n(a) \propto \left(\frac{\Gamma(a+1)}{\Gamma(a)}\right)^n \left(\prod_{i=1}^n x_i\right)^a,$$
$$= a^n \left(\prod_{i=1}^n x_i\right)^a.$$

Temos algumas opções para computar esta quantidade por máxima verossimilhança: podemos reparametrizar a distribuição Beta em termos da sua esperança μ e maximizar a verossimilhança resultante ou podemos obter $\hat{a}_{\text{EMV}}$ e usar a invariância do EMV para escrever $\hat{\mu}_{\text{EMV}} = \hat{a}_{\text{EMV}}/(\hat{a}_{\text{EMV}}+1)$. Aqui vamos fazer os dois caminhos, mas primeiro o mais fácil. Tomando o log da verossimilhança e diferenciando, temos

$$\frac{\partial}{\partial a} \log L_n(a) = \frac{n}{a} + \sum_{i=1}^n \log(x_i),$$

de modo que podemos escrever[3]

$$\hat{a}_{\text{EMV}} = -\frac{n}{\sum_{i=1}^n \log(X_i)}.$$

Agora, vamos ver como fica verossimilhança reparametrizada:

$$L_n(\mu) = \left(\frac{\mu}{1-\mu}\right)^n \left(\prod_{i=1}^n x_i\right)^{\left(\frac{\mu}{1-\mu}\right)}.$$

Assim,

$$\frac{\partial}{\partial \mu} \log L_n(\mu) = -\frac{\left(n - \sum_{i=1}^n \log(x_i)\right)\mu - n}{(1-\mu)^2 \mu}.$$

Como o denominador é positivo, para que $\frac{\partial}{\partial \mu} \log L_n(\mu) = 0$, é preciso que

$$-\left(n - \sum_{i=1}^n \log(x_i)\right)\mu - n = 0,$$
$$\implies \hat{\mu}_{\text{EMV}} = -\frac{n}{n - \sum_{i=1}^n \log(x_i)}.$$

o que de fato coincide com o que já calculamos – confira se quiser. ■

Comentário: Esta questão se baseia no exercício 7 da seção 7.7 de DeGroot (recomendado!). Aplicações do princípio da invariância do EMV foram vistos em sala e, por exemplo, no exercício 12 da seção 7.8, também recomendado.

Bônus: Lindex!

A função de perda LINEX (LINear–EXponential) é uma função de perda que trata assimetrias de maneira suave. Essa função é definida como:

$$L(\theta, a) = e^{c(a-\theta)} - c(a-\theta) - 1,$$

[3]Note que $\frac{\partial^2}{\partial a^2} \log L_n(a) = -\frac{n}{a^2} < 0$.

onde $c > 0$. Quando c varia, a função de perda varia de muito assimétrica para quase simétrica.

a) (10 pontos) Mostre que o estimador de Bayes para θ é dado por

$$\delta(\boldsymbol{X}) = -\frac{1}{c}\log\left(E\left[e^{-c\theta} \mid \boldsymbol{X})\right]\right).$$

b) (10 pontos) Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição $N(\mu, \sigma^2)$, com σ^2 conhecido. Suponha ainda que a priori é não informativa, ou seja $p(\mu) \propto 1$. Mostre que o estimador de Bayes utilizando a perda LINEX é

$$\widehat{\theta}_L = \overline{X}_n - \frac{c\sigma^2}{2n}.$$

Dica: Se Z é uma variável aleatória com distribuição normal de média m e variância v,

$$E[\exp\{kZ\}] = \exp\left(\frac{k^2 v + 2km}{2}\right),$$

para $k \in \mathbb{R}$.

Conceitos trabalhados: Estimador de Bayes.

Nível de dificuldade: médio.

Resolução: De maneira similar ao que fizemos na questão 2), vamos escrever o risco explicitamente:

$$\begin{aligned}
R(\delta, \theta) &= E_{\theta|\boldsymbol{x}}\left[\exp\left\{c(\delta - \theta)\right\} - c(\delta - \theta) - 1\right], \\
&= \exp(c\delta) E_{\theta|\boldsymbol{x}}\left[\exp\{-c\theta\}\right] - c\delta - cE_{\theta|\boldsymbol{x}}\left[\theta\right] - 1.
\end{aligned}$$

Para facilitar a notação, vamos escrever

$$\begin{aligned}
\omega &:= E_{\theta|\boldsymbol{x}}\left[\exp\{-c\theta\}\right], \\
\mu_{\boldsymbol{x}} &:= E_{\theta|\boldsymbol{x}}\left[\theta\right],
\end{aligned}$$

de modo que

$$R(\delta, \theta) = \exp(c\delta)\omega - c\delta - c\mu_{\boldsymbol{x}} - 1.$$

Agora vamos verificar que a LINEX é convexa, e, no processo, encontrar o estimador de Bayes.

$$\frac{\partial}{\partial \delta} R(\delta, \theta) = c(\omega \exp(c\delta) - 1), \tag{1}$$

$$\frac{\partial^2}{\partial \delta^2} R(\delta, \theta) = c^2 \omega \exp(c\delta). \tag{2}$$

Concluímos que igualar (1) a zero vai nos dar um ponto de mínimo, visto que a derivada segunda é maior que zero. Deste modo,

$$\begin{aligned}
c(\omega \exp(c\delta) - 1) = 0 &\implies \omega \exp(c\delta) = 1, \\
&\implies \delta^\star = -\frac{\log(\omega)}{c},
\end{aligned}$$

como queríamos demonstrar. Agora, vamos especializar este resultado para o caso normal com uma priori imprópria sobre μ. Para começar,

$$
\begin{aligned}
\xi(\mu \mid \boldsymbol{x}) &\propto f_n(\boldsymbol{x} \mid x), \\
&\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right), \\
&\propto \exp\left(-\frac{n}{2\sigma^2}(\mu - \bar{x}_n)^2\right),
\end{aligned}
$$

onde a última expressão segue de identidade conhecida e discutida em sala. Agora, notamos que a posteriori $\xi(\mu \mid \boldsymbol{x})$ é a densidade de uma distribuição normal com média $\bar{x}_n$ e variância σ^2/n. Agora, sabemos que o estimador LINEX é

$$
\delta_{\text{LINEX}} = -\frac{1}{c}\log\left(E_{\theta|\boldsymbol{x}}\left[\exp\{-c\mu\}\right]\right).
$$

Seguindo a dica e fazendo $k = -c$, $m = \bar{x}_n$ e $v = \sigma^2/n$, temos

$$
\begin{aligned}
\delta_{\text{LINEX}} &= -\frac{1}{c}\log\left(\exp\left\{\frac{c^2\frac{\sigma^2}{n} - 2c\bar{x}_n}{2}\right\}\right), \\
&= -\frac{1}{c}\left\{\frac{c^2\frac{\sigma^2}{n} - 2c\bar{x}_n}{2}\right\},
\end{aligned}
$$

de onde segue a resposta de b). Note que δ_{LINEX} é inadmissível sob perda quadrática, tendo viés $b(\mu) = -c\sigma^2/(2n)$, independente do valor de μ e indo para zero quando $n \to \infty$. ■

Comentário: Nesta questão nós aplicamos os princípios da teoria da Teoria da Decisão que dão suporte à inferência bayesiana. Empregamos uma perda cuja assimetria é possível controlar através de um parâmetro c e descobrimos que o estimador resultante é viesado, mas esse é o preço a se pagar por um estimador que é ótimo sob qualquer nível de assimetria induzido pela perda LINEX.