Lista de Exercícios 1

Estatística Bayesiana

1.17 (Berger and Wolpert (1988, p. 21)) Consider x with support $\{1, 2, 3\}$ and distribution $f(\cdot \mid 0)$ or $f(\cdot \mid 1)$, where

	x 1	2	3
$f(x\|0)$	0.9	0.05	0.05
$f(x\|1)$	0.1	0.05	0.85

Show that the procedure that rejects the hypothesis $H_0 : \theta = 0$ (to accept $H_1 : \theta = 1$) when $x = 2, 3$ has a probability 0.9 to be correct (under H_0 as well as under the alternative). What is the implication of the Likelihood Principle when $x = 2$?

Tópico: Princípios da Verossimilhança e Suficiência

Considere o procedimento: rejeito se $x \neq 1$. Logo a região crítica é $S = \{2, 3\}$.

$$\underbrace{\mathbb{P}(X \in S \mid \theta = 1)}^{\text{Rejeitar sob } H_1} = f(2|1) + f(3|1) = 0.9$$

$$\underbrace{\mathbb{P}(X \notin S \mid \theta = 0)}_{\text{Não rejeitar sob } H_0} = f(1|0) = 0.9$$

Quando $x = 2$ é observado, $f(x|0) = f(x|1)$, isto é, $\ell(\theta|x=2)$ é constante em $\{0, 1\}$. Segundo o Princípio da Verossimilhança, qualquer verossimilhança constante leva às mesmas inferências. Em particular, não conseguimos distinguir $\theta = 0$ de $\theta = 1$.

1.26 Show that, if the likelihood function $\ell(\theta|x)$ is used as a density on θ, the resulting inference does not obey the Likelihood Principle (*Hint:* Show that the posterior distribution of $h(\theta)$, when h is a one-to-one transform, is not the transform of $\ell(\theta|x)$ by the Jacobian rule.)

Tópico: Princípio da Verossimilhança

Primeiro temos que supor que $\ell(\theta|x)$ é integrável, pois do contrário não poderia ser densidade.

Pelo Teorema da Mudança de Variáveis,

$$p(h(\theta)|x) = p(\theta|x) \cdot \left| \det \left[\left. \frac{dh^{-1}(z)}{dz} \right|_{z=h(\theta)} \right] \right|$$

$$\propto \ell(\theta|x) \left| \det \left[\left. \frac{dh^{-1}(z)}{dz} \right|_{z=h(\theta)} \right] \right|$$

Note que tanto $p(h(\theta)|x)$, quanto $p(\theta|x)$ obtém a informação de x apenas através de $\ell(\theta|x)$, mas $p(h(\theta)|x)$ recebe mais informação mesmo sendo uma transformação bijetiva de θ.

↳ Ver seção 3.5.1 e página 20.

1.32 (Olkin et al. (1981)) Consider n observations $x_1, \ldots, x_n$ from $\mathcal{B}(k, p)$ where both k and p are unknown.

a. Show that the maximum likelihood estimator of k, $\hat{k}$, is such that

$$(\hat{k}(1-\hat{p}))^n \geq \prod_{i=1}^{n}(\hat{k}-x_i) \quad \text{and} \quad ((\hat{k}+1)(1-\hat{p}))^n < \prod_{i=1}^{n}(\hat{k}+1-x_i),$$

where $\hat{p}$ is the maximum likelihood estimator of p.

b. If the sample is 16, 18, 22, 25, 27, show that $\hat{k} = 99$.

c. If the sample is 16, 18, 22, 25, 28, show that $\hat{k} = 190$ and conclude on the stability of the maximum likelihood estimator.

a) $f(x|K,p) = \binom{K}{x} p^x (1-p)^{K-x}$

$$L(K,p|\vec{x}) = \prod_{i=1}^{n} \binom{K}{x_i} p^{x_i}(1-p)^{K-x_i}$$

$$\ell(K,p) = \sum_{i=1}^{n} \ln \binom{K}{x_i} + \ln p \sum_{i=1}^{n} x_i + \left(nK - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

$$\frac{\partial}{\partial p}\ell(\hat{K},\hat{p}) = \frac{\sum_{i=1}^{n} x_i}{\hat{p}} - \frac{n\hat{K} - \sum_{i=1}^{n} x_i}{1-\hat{p}} = 0 \Rightarrow \hat{p} = \frac{\sum_{i=1}^{n} x_i}{n\hat{K}}$$

Note que

$$\frac{L(K,\hat{p})}{L(K-1,\hat{p})} = \prod_{i=1}^{n} \frac{K}{K-x_i}(1-\hat{p}) = \frac{[K(1-\hat{p})]^n}{\prod_{i=1}^{n}(K-x_i)}$$

Quando $[K(1-\hat{p})]^n \geq \prod_{i=1}^{n}(K-x_i)$, L cresce em K
Como $L(\hat{K},\hat{p})$ maximiza a verossimilhança,

$$[\hat{K}(1-\hat{p})]^n \geq \prod_{i=1}^{n}(\hat{K}-x_i) \;\; e \;\; [(\hat{K}+1)(1-\hat{p})]^n < \prod_{i=1}^{n}(\hat{K}+1-x_i)$$

b) e c) Basta verificar que $\hat{K}$ satisfaz as relações acima.

1.37 If $x \sim \mathcal{N}(\theta, \sigma^2)$, $y \sim \mathcal{N}(\varrho x, \sigma^2)$, as in an autoregressive model, with ϱ known, and $\pi(\theta, \sigma^2) = 1/\sigma^2$, give the predictive distribution of y given x.

Vou supor que temos n amostras. Depois basta substituir por $n = 1$.

① Cálculo de posteriori

$$p(\theta, \sigma^2 | x) \propto \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \overbrace{(x_i - \theta)^2}^{x_i^2 - 2x_i\theta + \theta^2}\right\} \cdot \frac{1}{\sigma^2}$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+1} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i - \underset{\text{média amostral}}{\bar{x}})^2 + n(\bar{x} - \theta)^2\right]\right\}$$

$$= \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+1} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)\underset{\text{variância amostral}}{s^2} + n(\bar{x} - \theta)^2\right]\right\}$$

É integrável? A priori é imprópria!

$$\int_0^{+\infty} \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+1} \exp\left\{-\frac{1}{2\sigma^2}(n-1)s^2\right\} \int_{-\infty}^{\infty} \overbrace{\exp\left\{\frac{-n(\bar{x} - \theta)^2}{2\sigma^2}\right\}}^{\text{Núcleo de } \mathcal{N}(\bar{x}, \sigma^2/n)} d\theta \, d\sigma^2$$

$$= \int_0^{+\infty} \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+1} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\} \cdot \sqrt{2\pi\sigma^2/n} \, d\sigma^2$$

$$= \sqrt{2\pi/n} \int_0^{+\infty} \underbrace{\left(\frac{1}{\sigma^2}\right)^{\frac{n+1}{2}+1} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\}}_{\text{núcleo da Inv Gamma}\left(\frac{n+1}{2}, \frac{(n-1)s^2}{2}\right)} d\sigma^2$$

$< +\infty$ se, e somente se, $n > 1$. Portanto, a preditiva $g(y|x)$ talvez não esteja bem definida. Vamos verificar.

② Cálculo da preditiva

$$g(y|x) \propto \int_0^{+\infty}\int_{-\infty}^{+\infty} \left(\frac{1}{\sigma^2}\right)^{1/2} \exp\left\{-\frac{1}{2\sigma^2}(y-\theta x)^2\right\} \underbrace{\left(\frac{1}{\sigma^2}\right)^{3/2} \exp\left\{-\frac{1}{2\sigma^2}(x-\theta)^2\right\}}_{N(x,\sigma^2)} d\theta\, d\sigma^2$$

$$\propto \int_0^{+\infty} \left(\frac{1}{\sigma^2}\right)^2 \exp\left\{-\frac{1}{2\sigma^2}(y-\theta x)^2\right\} (\sigma^2)^{1/2}\, d\sigma^2$$

$$= \int_0^{+\infty} \left(\frac{1}{\sigma^2}\right)^{\frac{3}{2}+1} \exp\left\{-\frac{(y-\theta x)^2}{2\sigma^2}\right\} d\sigma^2 \rightarrow \text{InvGamma}$$

$$= \frac{\Gamma(3/2)}{\left(\frac{(y-\theta x)^2}{2}\right)^{1/2}} \propto \frac{1}{|y-\theta x|}$$

Como esperado $g(y|x)$ não é bem definido.

1.41 *Given a couple (x, y) of random variables, the marginal distributions $f(x)$ and $f(y)$ are not sufficient to characterize the joint distribution of (x, y).

a. Give an example of two different bivariate distributions with the same marginals. (*Hint:* Take these marginals to be uniform $\mathcal{U}([0,1])$ and find a function from $[0,1]^2$ to $[0,1]^2$ which is increasing in both its coefficients).

b. Show that, on the contrary, if the two conditional distributions $f(x|y)$ and $f(y|x)$ are known, the distribution of the couple (x, y) is also uniquely defined.

c. Extend b. to a vector $(x_1, \ldots, x_n)$ such that the full conditionals $f_i(x_i|x_j, j \neq i)$ are known. [*Note:* This result is called the *Hammersley–Clifford* Theorem, see Robert and Casella (2004).]

d. Show that property b. does not necessarily hold if $f(x|y)$ and $f(x)$ are known, i.e., that several distributions $f(y)$ can relate $f(x)$ and $f(x|y)$. (*Hint:* Exhibit a counter-example.)

e. Give some sufficient conditions on $f(x|y)$ for the above property to be true. (*Hint:* Relate this problem to the theory of complete statistics.)

a) Seguindo a dica, sejam $x, y \sim \mathcal{U}([0,$

* $f_{X,Y}(x,y) = 1, \ \forall x,y \in [0,1]^2.$

$0 \leq x \leq 1, \ f_X(x) = \int_0^1 f_{X,Y}(x \qquad 1$

$0 \leq y \leq 1, \ f_Y(y) = \int_0^1 f_{X,Y}(x,y)\,dx = 1$

* $\mathcal{U} \sim \text{Dirichlet}(1/3, 2/3, 2/3, 1/3)$

$X = U_1 + U_2 \sim \text{Beta}(1,1)$, mas $\text{cor}(X,Y) \neq 0.$
$Y = U_1 + U_3 \sim \text{Beta}(1,1)$

b) Primeiro supomos que $f_{X,Y}(x,y)$ existe. Nesse caso, pelo Teorema de Bayes,

$$\frac{f_{X|Y}(x|y)}{f_{Y|X}(y|x)} = \frac{f_X(x)}{f_Y(y)}, \quad \forall x,y \in \mathcal{X} \times \mathcal{Y}$$

e, então,

$$\frac{1}{f_Y(y)} = \int_X \frac{f_X(x)}{f_Y(y)}\,dx = \int_X \frac{f_{X|Y}(x|y)}{f_{Y|X}(y|x)}\,dx,$$

desde que $f_Y(y) > 0$. Portanto,

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)\, f_Y(y)$$
$$= \frac{f_{X|Y}(x|y)}{\int_X \frac{f_{X|Y}(x|y)}{f_{Y|X}(y|x)}\,dx}. \qquad (1)$$

Concluo que se $f_{X,Y}(x,y)$ existe, ela deve satisfazer equação (1)

c) Agora assumimos que $f(x_1,\dots,x_n)$ existe. Vamos provar por indução que ela é unicamente definida a partir das condicionais completas conhecidas. Para $n=2$, acabamos de provar. Suponha para n. Assim:

$$f(x_1,\dots,x_n,x_{n+1}) = f(x_{n+1}|x_1,\dots,x_n)\cdot f(x_1,\dots,x_n)$$

e $f(x_1,\dots,x_n)$ é unicamente determinada a partir de

$$f(x_i | x_j,\ 1 \le j \le n,\ j \neq i)$$

Note que $\dfrac{f(x_{n+1}|x_1,\dots,x_n)}{f(x_i|x_j|_{j=1,j\neq i}^{n}, x_{n+1})} = \dfrac{f(x_{n+1}|x_j|_{j=1,j\neq i}^{n})}{f(x_i|x_j|_{j=1,j\neq i}^{n})}$ para $i=1,\dots,n$ pelo Teorema de Bayes. Portanto

$$f(x_i|x_j|_{j=1,j\neq i}^{n}) = \frac{1}{\int_X \frac{f(x_{n+1}|x_j|_{j=1,j\neq i}^{n})}{f(x_i|x_j|_{j=1,j\neq i}^{n})}\,dx_{n+1}} = \frac{1}{\int_X \frac{f(x_{n+1}|x_1,\dots,x_n)}{f(x_i|x_j|_{j=1,j\neq i}^{n}, x_{n+1})}\,dx_i}$$

que é unicamente determinado. Portanto vale para $n+1$.

d) Vou montar um exemplo discreto. Considere as tabelas:

	$y=1$	$y=2$	$y=3$	$f(x)$
$x=1$	$3/8$	$1/8$	$1/4$	$3/4$
$x=2$	$1/8$	$1/8$	0	$1/4$
$f(y)$	$1/2$	$1/4$	$1/4$	1

	$y=1$	$y=2$	$y=3$	$f(x)$
$x=1$	$9/16$	$1/16$	$1/8$	$3/4$
$x=2$	$3/16$	$1/16$	0	$1/4$
	$3/4$	$1/8$	$1/8$	1

$$f(x|y=1) = \begin{cases} 3/4 \\ 1/4 \end{cases} \qquad f(x|y=1) = \begin{cases} 3/4 \\ 1/4 \end{cases}$$

$$f(x|y=2) = \begin{cases} 1/2 \\ 1/2 \end{cases} \qquad f(x|y=2) = \begin{cases} 1/2 \\ 1/2 \end{cases}$$

$$f(x|y=3) = \begin{cases} 1 \\ 0 \end{cases} \qquad f(x|y=3) = \begin{cases} 1 \\ 0 \end{cases}$$

Assim, mesmo $f(x|y)$ e $f(x)$ sendo conhecidos, a conjunta $f(x,y)$ não é única.

e) Dada $f(x)$, suponha que

$$\int_{\mathcal{Y}} f(x|y)\, dy = f(x)$$

e que $f(x|y) > 0$, $\forall x,y \in \mathcal{X} \times \mathcal{Y}$. Assim,

$$0 = \int_{\mathcal{Y}} \left(f(x,y) - f(x|y)\right) dy$$

$$= \int_{\mathcal{Y}} f(x|y)\left(f(y) - 1\right) dy \Rightarrow f(y) = 1,\ \forall y \in \mathcal{Y}$$