Lista de exercícios 2

Estatística Bayesiana

Obs.: esse é apenas um rascunho!

2.13 Show that, for a loss function $L(\theta, d)$ strictly increasing in $|d-\theta|$ such that $L(\theta,\theta) = 0$, there is no uniformly optimal statistical procedure. Give a counterexample when

$$L(\theta, \varphi) = \theta(\mathbb{I}_{\mathbb{R}^*}(\theta) - \varphi)^2.$$

Suponha que exista $d^* \in D$ tal que $\forall \theta \in \Omega$, temos que

$$L(\theta, d^*) \leq L(\theta, d), \ \forall d \in D.$$

Como consequência, devemos ter $L(\theta, d^*) = 0, \ \forall \theta \in \Omega$.
Todavia, é fácil verificar que existem θ_1 e θ_2 tal que

$$|d^* - \theta_1| < |d^* - \theta_2|,$$

e, portanto,

$$0 = L(\theta_1, d^*) < L(\theta_2, d^*) = 0,$$

e entramos em contradição. Concluo que não existe d^*.

Quando $L(\theta, \varphi) = \theta(\mathbb{I}_{\mathbb{R}^*}(\theta) - \varphi)^2$ (Note que concluímos que $\theta \geq 0$), ponha $\varphi = 1$. Nesse caso

$$L(\theta, \varphi) = \begin{cases} 0, & \theta = 0 \\ \theta(\mathbb{I}_{\mathbb{R}^*}(\theta) - 1)^2 = 0, & \theta \neq 0. \end{cases}$$

Em particular $L(\theta, 1) \leq L(\theta, \varphi), \ \forall \varphi \in D$.

2.25 Consider the case when $\Theta = \{\theta_1, \theta_2\}$ and $\mathcal{D} = \{d_1, d_2, d_3\}$, for the following loss structure

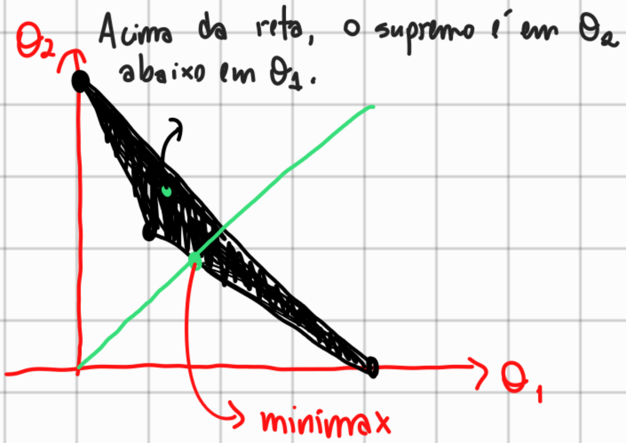
	d_1	d_2	d_3
θ_1	2	0	0.5
θ_2	0	2	1

a. Determine the minimax procedures.

b. Identify the least favorable prior distribution. (*Hint:* Represent the risk space associated with the three actions as in Example 2.4.12.)

Note que $\forall d \in D^*$, $d = \alpha_1 d_1 + \alpha_2 d_2 + \alpha_3 d_3$ com $\sum \alpha_i = 1$. Em particular, $L(\theta, d) = \sum_i L(\theta, d_i)\alpha_i$.

Considere $R(\theta, d) = L(\theta, d)$, pois não existe modelo para os dados. Assim o conjunto de risco é formado pelas combinações convexas de $L(\theta, d_i)$, cujos vetores são $(2,0)$, $(0,2)$, $(0.5, 1)$, representado por:



$$0.5\alpha + 2(1-\alpha) = \alpha + 0(1-\alpha) \Rightarrow \alpha = \frac{4}{5}.$$

Logo o minimax é $\frac{4}{5} d_3 + \frac{1}{5} d_1$

b) Note que em estimador de Bayes, queremos minimizar $R(\theta_1, \delta)\pi(\theta_1) + R(\theta_2, \delta)\pi(\theta_2)$ em δ. Geometricamente, dado $\pi_1 = \pi(\theta_1)$, estamos olhando a reta

$$(\pi_1 x, (1-\pi_1)y) = (\pi_1 x, 0) + (0, (1-\pi_1)y),$$

em particular queremos minimizar a norma soma desse vetor, sujeito a algumas restrições. Note que $y = \pi_1/_{1-\pi_1} x$ é a equação da reta.

Concluímos que os estimadores de Bayes ficam nas bordas inferiores.

Agora queremos o máximo de $r(\pi)$ que está nas bordas. Pelo Lema 2.4.13, o minimax, que é Bayes nessa reta com respeito a alguma π, faz com que essa π seja a distribuição no mínimo favorável.

2.28 Consider $x \sim \mathcal{B}(n, \theta)$, with n known.

a. If $\pi(\theta)$ is the beta distribution $\mathcal{B}e(\sqrt{n}/2, \sqrt{n}/2)$, give the associated posterior distribution $\pi(\theta|x)$ and the posterior expectation, $\delta^\pi(x)$.

b. Show that, when $\mathrm{L}(\delta, \theta) = (\theta - \delta)^2$, the risk of δ^π is constant. Conclude that δ^π is minimax.

c. Compare the risk for δ^π with the risk function of $\delta_0(x) = x/n$ for $n = 10, 50$, and 100. Conclude about the appeal of δ^π.

a) $\pi(\theta|x)$ é $Be(\sqrt{n}/2 + x, \sqrt{n}/2 + n - x)$. Para ver isso, faça

$$\pi(\theta|x) \propto \theta^x (1-\theta)^{n-x}\, \theta^{\sqrt{n}/2-1} (1-\theta)^{\sqrt{n}/2-1}$$
$$= \theta^{\sqrt{n}/2+x-1} (1-\theta)^{\sqrt{n}/2+n-x-1},$$

que é o núcleo da distribuição beta.

$$\delta^\pi(x) = \mathbb{E}^\pi[\theta|x] = \frac{\sqrt{n}/2 + x}{\sqrt{n} + n} = \frac{1/2 + x/\sqrt{n}}{1 + \sqrt{n}}$$

b) $$R(\theta, \delta^\pi) = \int_{\mathcal{X}} (\theta - \delta^\pi(x))^2 f(x|\theta)\, dx$$

Var + E²

$$= \theta^2 - 2\theta E[\delta^\pi(x)] + E[\delta^\pi(x)^2]$$
$$= \theta^2 - 2\theta\left(\frac{1/2 + n\theta}{1+\sqrt{n}}\right) + \frac{\theta(1-\theta)}{(1+\sqrt{n})^2} + \left(\frac{1/2+n\theta}{1+\sqrt{n}}\right)^2$$

$\propto \theta^2(1+\sqrt{n})^2 - 2\theta(1+\sqrt{n})(1/2+n\theta) + \theta(1-\theta) + (1/2+n\theta)^2$
$= 1/4$. Logo R é constante em θ. Em particular,

$$r(\pi) = R(\theta, \delta^\pi),$$

$\forall \theta \in \Omega$ e δ^π é estimador de Bayes para essa perda. Logo δ^π é minimax.

c) $$R(\theta, \delta_0) = \theta^2 - 2\theta E[\delta_0(x)] + E[\delta_0(x)^2]$$
$$= \theta^2 - 2\theta^2 + \frac{\theta(1-\theta)}{n} + \theta^2$$
$$= \theta(1-\theta)/n,$$

contra $R(\theta, \delta^\pi) = 1/4(1+\sqrt{n})^2$, que decresce mais rapidamente.

2.36 Show that, under squared error loss, if two real estimators δ_1 and δ_2 are distinct and satisfy

$$R(\theta,\delta_1) = (\theta - \delta_1(x))^2 = R(\theta,\delta_2) = (\theta - \delta_2(x))^2,$$

the estimator δ_1 is not admissible. (*Hint:* Consider $\delta_3 = (\delta_1+\delta_2)/2$ or $\delta_4 = \delta_1^\alpha \delta_2^{1-\alpha}$.) Extend this result to all strictly convex losses and construct a counterexample when the loss function is not convex.

Considere $\delta_3 = (\delta_1 + \delta_2)/2$. Assim, $\forall \theta \in \Omega$,

$$R(\theta,\delta_3) = \int_{\mathcal{X}} L(\theta,\delta_3)\, f(x|\theta)\, dx$$

convexidade estrita

$$< \int_{\mathcal{X}} \left[\frac{1}{2} L(\theta,\delta_1) + \frac{1}{2} L(\theta,\delta_2)\right] f(x|\theta)\, dx$$

$$= \frac{1}{2} R(\theta,\delta_1) + \frac{1}{2} R(\theta,\delta_2) = R(\theta,\delta_i),\ i=1,2,$$

logo δ_1 e δ_2 são não admissíveis.

Um exemplo bobo é o exercício 2.13

$$L(\theta,\varphi) = \theta\,(1_{\theta>0} - \varphi)^2$$

e X com distribuição contínua. Tome $x_0 \in \mathcal{X}$ e defina $\delta_1(x) = 1$ e $\delta_2(x) = \begin{cases} 1, & x \neq x_0 \\ 0, & x = x_0 \end{cases}$. logo

$$R(\theta,\delta_1) = \theta \int_{\mathcal{X}} (1_{\theta>0} - 1)^2 f(x|\theta)\, dx = 0$$

$$R(\theta,\delta_2) = \theta \int_{\mathcal{X}-\{x_0\}} (1_{\theta>0} - 1)^2 f(x|\theta)\, dx = 0,$$

pois $\int_{\{x_0\}} f(x|\theta)\, dx = 0.$

2.42 *(Zellner (1986a)) Consider the LINEX loss in $\mathbb{R}$, defined by

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

a. Show that $L(\theta, d) > 0$ and plot this loss as a function of $(\theta - d)$ when $c = 0.1, 0.5, 1, 2$.

b. Give the expression of a Bayes estimator under this loss.

c. If $x_1, \ldots, x_n \sim \mathcal{N}(\theta, 1)$ and $\pi(\theta) = 1$, give the associated Bayes estimator.

a) Vou provar que $e^x > x + 1$, $\forall x > 0$. Temos que

$$e^x = 1 + x + \frac{x^2}{2} + \cdots,$$

pela expansão de Taylor. Logo $x > 0 \Rightarrow e^x > x + 1$.
Em especial, $L(\theta, d) > 0$, se $\theta \neq d$. Quando $\theta = d$, note que $L(\theta, d) = 0$

b)
$$\rho(\pi, d \mid x) = \int_\Omega \left(e^{c(\theta-d)} - c(\theta-d) - 1\right) \pi(\theta \mid x)\, d\theta$$

$$= e^{-cd} \int_\Omega e^{c\theta} \pi(\theta \mid x)\, d\theta - c \int_\Omega \theta \pi(\theta \mid x)\, d\theta + cd - 1$$

$$= e^{-cd} E^\pi[e^{c\theta} \mid x] - c E^\pi[\theta \mid x] + cd - 1$$

$$= e^{-cd} M_{\theta|x}(c) - c E^\pi[\theta \mid x] + cd - 1$$

Função geradora de momentos

Logo $\min_d \rho(\pi, d \mid x)$ é equivalente a

$$-c e^{-cd} M_{\theta|x}(c) + c = 0 \Rightarrow e^{cd} = M_{\theta|x}(c)$$

$$\Rightarrow d = \frac{1}{c} \log M_{\theta|x}(c),$$

isto é, $\delta^\pi(x) = c^{-1} \log E^\pi[e^{c\theta} \mid x]$

c) $x_1, \ldots, x_n \sim \mathcal{N}(\theta, 1)$, $\theta \sim$ Lebesgue

$$\pi(\theta \mid x) \propto \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2\right\}$$

$$= \exp\left\{-\frac{1}{2}\left(n\theta^2 - 2\theta n\bar{x}_n + \sum_{i=1}^{n} x_i^2\right)\right\}$$

$$\propto \exp\left\{-\frac{n}{2}\left(\theta^2 - 2\theta\bar{x}_n + \bar{x}_n^2 - \bar{x}_n^2\right)\right\}$$

$$\propto \exp\left\{-\frac{n}{2}\left(\theta - \bar{x}_n\right)^2\right\},$$

isto é, $\theta \mid x \sim \text{Normal}(\bar{x}_n, 1/n)$.

$M_{\theta|x}(c) = e^{c\bar{x}_n + \frac{c^2}{2n}}$, logo

$$\begin{aligned}\delta^{\pi}(x) &= c^{-1}(c\bar{x}_n + c^2/2n) \\ &= \bar{x}_n + c/2n.\end{aligned}$$

2.48 (Robert (1996b)) Show that both the entropic and the Hellinger losses are locally equivalent to the quadratic loss associated with the Fisher information,

$$I(\theta) = \mathbb{E}_\theta \left[\frac{\partial \log f(x|\theta)}{\partial \log} \left(\frac{\partial \log f(x|\theta)}{\partial \log} \right)^t \right],$$

that is,

$$L_e(\theta, \delta) = L_e(\theta - \delta)^t I(\theta)^{-1} (\theta - \delta) + O(\|\theta - \delta\|^2)$$

and

$$L_H(\theta, \delta) = c_H(\theta - \delta)^t I(\theta)^{-1} (\theta - \delta) + O(\|\theta - \delta\|^2),$$

where c_e and c_H are constants.

Por simplicidade, considere o caso unitário inicialmente.
Vamos calcular a expansão de Taylor de L_e e L_H:

Teorema convergência dominante

$$\frac{\partial}{\partial \delta} L_e(\theta,\delta) = \frac{d}{d\delta} E_\theta\left[\log\left(\frac{f(x|\theta)}{f(x|\delta)}\right)\right]$$

$g(\theta,\cdot) \in C^1$ e $\|\frac{\partial}{\partial\delta} g(\theta,\delta)\| \leq Z$ integrável

$$= E_\theta\left[\frac{-\frac{f(x|\theta)}{f(x|\delta)^2}\cdot\frac{d}{d\delta}f(x|\delta)}{f(x|\theta)/f(x|\delta)}\right]$$

$$= -E_\theta\left[\frac{\frac{d}{d\delta}f(x|\delta)}{f(x|\delta)}\right]$$

$$= -\int_{\mathcal{X}} \frac{d}{d\delta} f(x|\delta)\,\frac{f(x|\theta)}{f(x|\delta)}\,dx$$

Logo $\displaystyle \frac{\partial}{\partial\delta} L_e(\theta,\delta)\Big|_{\delta=\theta} = -\int \frac{\partial}{\partial\delta} f(x|\theta)\,dx = \frac{\partial}{\partial s}(-1) = 0.$

Agora

$$\frac{\partial^2}{\partial\delta^2} L_e(\theta,\delta) = -E_\theta\left[\frac{\partial}{\partial\delta}\Big(\underbrace{\frac{\partial}{\partial\delta}\log f(x|\delta)}_{\frac{\partial}{\partial s}f(x|\delta)\,/\,f(x|\delta)}\Big)\right]$$

$$= I(\delta)\big|_{\delta=\theta}$$

sob certas condições de regularidade. A expansão de Taylor afirma que

$$L_e(\theta,\delta) \approx L_e(\theta,\theta) + \frac{\partial}{\partial\delta}L(\theta,\delta)\Big|_{\delta=\theta}(\delta-\theta) + \frac{\partial^2}{\partial\delta^2}L(\theta,\delta)\Big|_{\delta=\theta}(\delta-\theta)^2/2$$

$$= \frac{1}{2}(\delta-\theta)^2 I(\theta).$$

A extensão para $\mathbb{R}^n$ é natural. Além disso, a perda L_H é muito similar.