Inferência bayesiana sob a ótica da teoria da decisão

Prof. Caio Azevedo

Motivação

- De uma forma geral, todas as questões de Inferência Estatística (frequentista ou bayesiana), podem ser estudadas sob a ótica da Teoria da Decisão (ref 1, ref 2).
- De um modo rudimentar, a Teoria da Decisão consiste em um conjunto de metodologias que nos levam à escolher uma (eventualmente mais de uma), entre várias ações (decisões) possíveis, de modo a minimizar alguma perda ou a maximizar algum ganho.
- Por exemplo, qual estimador (frequentista) escolher (ação, decisão), entre aqueles que são não viciados? Podemos escolher aquele, por exemplo, que apresenta a menor variância (perda).

Elementos básicos da Teoria da Decisão (frequentista & bayesiana)

- Medidas de qualidade básica são:
  - Função de perda.
  - Perda esperada (risco frequentista).
  - Risco de Bayes.
  - Risco de Bayes integrado ou à posteriori.
- A inferência frequentista utiliza os dois primeiros elementos acima.
- A inferência bayesiana utiliza os quatro elementos acima.
- Seja $\boldsymbol{X}|\boldsymbol{\theta} = (X_1|\boldsymbol{\theta}, ..., X_n|\boldsymbol{\theta})'$ um vetor aleatório com fdp $f_{\boldsymbol{X}|\boldsymbol{\theta}}(.|\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}^k$.

Introdução à Teoria da Decisão

- Os elementos básicos da Teoria da Decisão são:

  1. Θ : espaço paramétrico (estado da natureza).
  2. $\mathcal{A}$: é o conjunto das possíveis ações que podem ser tomadas (exemplos):
     - $\mathcal{A}$: {é o conjunto dos estimadores pontuais}.
     - $\mathcal{A}$: {é o conjunto das funções de teste}.
  3. Uma função de decisão definida por $d : \mathcal{X} \rightarrow \mathcal{A}$, $(\boldsymbol{x} \rightarrow d(\boldsymbol{x}))$ que é formulada para se obter informações sobre $\boldsymbol{\theta}$. Notação: $\mathcal{D}$: {classe de todas as funções de decisão}.
  4. Para avaliar a escolha de uma função de decisão ($d \in \mathcal{D}$), vamos considerar o conceito de função perda.

Função de perda

- Seja $\delta(\boldsymbol{X})$ um estimador (frequentista ou bayesiano) e $\delta(\boldsymbol{x})$ a respectiva estimativa.
- Seja θ o parâmetro de interesse.
- Seja $\mathcal{D}$ o conjunto de todos os estimadores "possíveis" para θ.

Função de perda

- A função de perda (as vezes chamada de função perda) é definida por:

$$
\begin{aligned}
L(\delta, \theta) &= L(\delta(\boldsymbol{x}), \theta) \\
(\mathcal{D}, \Theta) &\rightarrow \mathcal{R}^{+}
\end{aligned}
$$

  tal que $L(.,.)$ satisfaz as seguintes propriedades (próximo slide):

  a) $L(\delta, \theta) \geq 0, \forall d \in \mathcal{D}, \theta \in \Theta$.
  b) $L(\delta, \theta) = 0$, se $d = \theta$.

- Note que considerarmos $d = d(\boldsymbol{X})$ então $L(d(\boldsymbol{X}), \theta)$ é uma va.

Exemplos de função de perda

- $L_1 = (\delta - \theta)^2$ (perda quadrática).
- $L_2 = |\delta - \theta|$ (perda modular).
- $L_3 = \begin{cases} A, & \text{se } |\delta - \theta| > \epsilon, \epsilon > 0 \\ 0, & \text{se } |\delta - \theta| \leq \epsilon \end{cases}$ (perda (0,A)).
- $L_4 = \rho(\theta)|\delta - \theta|^r, \rho(\theta) \geq 0, \forall\theta, r > 0$ (perda generalizada).

Perda esperada (risco frequentista)

- Para uma dada função de perda, o risco frequentista é definido como sendo o seu valor esperado em relação à distribuição de $\boldsymbol{X}|\theta$, ou seja

$$R(\delta,\theta) = \mathcal{E}_{\boldsymbol{X}|\theta}[L(\delta,\theta)] = \begin{cases} \sum_{\boldsymbol{x}\in\boldsymbol{X}(\Omega)} L(\delta(\boldsymbol{X}),\theta)p(\boldsymbol{x}|\theta), \text{ se } \boldsymbol{X} \text{ for discreto} \\ \int_{\boldsymbol{X}(\Omega)} L(\delta(\boldsymbol{X}),\theta)p(\boldsymbol{x}|\theta)d\boldsymbol{x}, \text{se } \boldsymbol{X} \text{ for contínuo} \end{cases}$$

- Exemplo [Erro quadrático médio (frequentista)]: $\mathcal{E}_{\boldsymbol{X}|\theta}[(\delta-\theta)^2]$.
- Obs: $\mathcal{X} = \boldsymbol{X}(\Omega)$.

Perda esperada (risco frequentista)

- Em geral, na inferência frequentista, busca-se o estimador que minimiza o risco acima definido (restrito à uma determinada subclasse de estimadores). Por exemplo, o ENVVUM.
- Obs: Se $L(\delta,\theta) = (\theta - \delta)^2$, então

$$
\begin{aligned}
R(d,\theta) &= \mathcal{E}_{\boldsymbol{X}|\theta}\left[(\theta-\delta)^2\right] = EQM_{\theta}(\delta) \\
&= \mathcal{V}_{\theta}(\delta) + B_{\theta}^{2}(\delta)
\end{aligned}
$$

- Mais detalhes sobre a inferência frequentista e teoria da decisão veja aqui.

Perda esperada (risco frequentista)

- Def: Dizemos que um procedimento d_1 é melhor do que um procedimento d_2 se

$$R(d_1, \theta) \leq R(d_2, \theta), \forall \theta \in \Theta \text{ e}$$

$$R(d_1, \theta) < R(d_2, \theta), \text{ para pelo menos um } \theta.$$

- Se d_1 e d_2 são tais que: $R(d_1, \theta) = R(d_2, \theta), \forall \theta \in \Theta$ dizemos que d_1 e d_2 são equivalentes.

Risco de Bayes

- Seja $p \equiv p(\theta)$ uma priori definida para θ.
- O risco de Bayes é definido como a perda média do risco frequentista, à priori, em relação à θ, ou seja

$$r(\delta, p) = \int_{\Theta} R(\delta, \theta) p(\theta) d\theta$$

- Contudo, em princípio, tal medida considera (explicitamente) apenas a priori.
- Como introduzir a posteriori no risco de Bayes?

Risco de Bayes (cont.)

- Note que

$$
\begin{aligned}
r(\delta, p) &= \int_{\Theta} R(\delta, \theta) p(\theta) d\theta \\
&= \int_{\Theta} \left[\int_{\boldsymbol{X}(\Omega)} L(\delta(\boldsymbol{X}), \theta) p(\boldsymbol{x}|\theta) d\boldsymbol{x} \right] p(\theta) d\theta \\
&= \int_{\Theta} \left[\int_{\boldsymbol{X}(\Omega)} L(\delta(\boldsymbol{X}), \theta) \frac{p(\boldsymbol{x}|\theta) p(\theta)}{p(\boldsymbol{x})} p(\boldsymbol{x}) d\boldsymbol{x} \right] d\theta \\
&= \int_{\boldsymbol{X}(\Omega)} \underbrace{\left[\int_{\Theta} L(\delta(\boldsymbol{X}), \theta) p(\theta|\boldsymbol{x}) d\theta \right]}_{R_B(\delta, p)} p(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int_{\boldsymbol{X}(\Omega)} R_B(\delta, p) p(\boldsymbol{x}) d\boldsymbol{x}
\end{aligned}
$$

Risco à posterioiri (ou integrado)

- Em geral, $R_B(\delta, p) = \int_{\Theta} L(\theta, d(\boldsymbol{x}))\pi(\theta|\boldsymbol{x})d\theta = \mathcal{E}_{\theta|\boldsymbol{x}}\left(L(\theta, d(\boldsymbol{x}))\right)$ é chamado de risco integrado ou à posteriori de Bayes.
- O estimador Bayesiano (δ^p) que minimiza $R_B(\delta, p)$ é chamado de estimador de Bayes. Ou seja,

$$r(\delta^p, p) = \inf_{\delta \in \mathcal{D}} r(\delta, p) \tag{1}$$

- Encontrar o estimador de Bayes significa minimizar $\mathcal{E}_{\theta|\boldsymbol{X}}(L(\delta, \theta))$.

Risco à posteriori

- Obs: Segundo a definição acima, pode ocorrer que:
  - Não exista δ^p em $\mathcal{D}$ que satisfaça (1).
  - Pode existir mais de um estimador de Bayes.
- A priori não precisa ser uma distribuição de probabilidade, tampouco ser própria $\left(\int_{\Theta} p(\theta)d\theta = \infty\right)$.

Estimadores Bayesianos associados à funções de perda específicas

- O estimador EAP é o estimador de Bayes para a função de perda $L_1 = (\delta - \theta)^2$.
- O estimador MeAP (mediana a posteriori) é o estimador de Bayes para a função de perda $L_4 = a|\delta - \theta|$, em que a é uma constante.
- O estimador MAP (moda a posteriori) é o estimador de Bayes para a função de perda $L_3 = \begin{cases} A, & \text{se } |\delta - \theta| > \epsilon, \epsilon > 0 \\ 0, & \text{se } |\delta - \theta| \leq \epsilon \end{cases}$

Demonstração para a função de perda $L_1 = (\delta - \theta)^2$

- Queremos minimizar $g(\delta) = \int_{\Theta} (\delta - \theta)^2 p(\theta|\boldsymbol{x}) d\theta$ em relação à δ. Com efeito, note que:

$$
\begin{aligned}
g(\delta) &= \delta^2 \int_{\Theta} p(\theta|\boldsymbol{x}) d\theta - 2\delta \int_{\Theta} \theta p(\theta|\boldsymbol{x}) d\theta + \int_{\Theta} \theta^2 p(\theta|\boldsymbol{x}) d\theta \\
&= \delta^2 - 2\delta \mathcal{E}(\theta|\boldsymbol{X}) + \mathcal{E}(\theta^2|\boldsymbol{X})
\end{aligned}
$$

- Derivando a expressão acima com relação à δ, igualando a expressão resultante à 0 e resolvendo-a em relação à δ, temos que

$$\widehat{\delta} = \widehat{\theta}_{Bayes} = \mathcal{E}(\theta|\boldsymbol{X})$$

Voltando ao Exemplo da $U(0,\theta)$

- $X_i|\theta \overset{i.i.d.}{\sim} U(0,\theta), i = 1, ..., n$. Considere a função de perda $L(\delta,\theta) = \frac{(\delta-\theta)^2}{\theta^2}$.
- Vamos considerar $p(\theta) = I_{(0,1)}(\theta)$.
- (Exercício) Pode-se provar que

$$p(\theta|\boldsymbol{x}) = \frac{\left(\frac{1}{\theta}\right)^n I_{(y_n,1)}(\theta)}{[1/(n-1)](y_n^{-n+1} - 1)}$$

- Queremos encontrar o estimador Bayesiano que minimiza

$$\int_{y_n}^{1} \frac{(\delta-\theta)^2}{\theta^2} p(\theta|\boldsymbol{x}) d\theta$$

Voltando ao Exemplo 9 (cont.)

- Ou seja, buscamos o estimador que minimiza

$$\begin{aligned}
&\int_{y_n}^{1} (\delta - \theta)^2 \frac{1}{\theta^{n+2}} d\theta \\
= \quad & \delta^2 \int_{y_n}^{1} \frac{1}{\theta^{n+2}} d\theta - 2\delta \int_{y_n}^{1} \frac{1}{\theta^{n+1}} d\theta + \int_{y_n}^{1} \frac{1}{\theta^{n}} d\theta
\end{aligned}$$

- Derivando a expressão acima com relação à δ, igualando a expressão resultante à 0 e resolvendo-a em relação à δ, temos que

$$\widehat{\theta}_{Bayes} = \widehat{\delta} = \frac{n+1}{n} \frac{1 - y_n^n}{1 - y_n^{n+1}} y_n$$

Estimador Minimax

- Voltando ao risco frequentista, o objetivo é encontrar o estimador (frequentista), digamos δ^*, que minimize o risco uniformemente em θ, ou seja:

$$R(\delta^*, \theta) \leq R(\delta, \theta), \forall \theta \in \theta, \text{ e}, \ \forall \delta \in \mathcal{D}$$

em que $\mathcal{D}$ a classe de todos os estimadores possíveis (sabemos que isso é praticamente impossível).

Estimador Minimax

- Uma outra abordagem, consiste em considerar o estimador Bayesiano que minimize o máximo (supremo) do risco frequentista, ou seja, encontrar o estimador (Bayesiano), digamos δ', tal que

$$sup_{\theta \in \Theta} R(\delta', \theta) \leq sup_{\theta \in \Theta} R(\delta, \theta), \forall \delta \in \mathcal{D}.$$

- Pode-se provar que se $\widehat{\theta}_{Bayes}$ é o estimador de Bayes de θ, tal que $R(\widehat{\theta}_{Bayes}, \theta)$ não dependa de θ, então $\widehat{\theta}_{Bayes}$ é o estimador minimax de θ.

Voltando ao Exemplo 1

- $X_i|\theta \overset{i.i.d.}{\sim} \text{Bernoulli}(\theta), i = 1, ..., n$.
- Considere a família conjugada de prioris e a função de perda quadrática $L_1 = (\delta - \theta)^2$.
- Vamos encontrar o estimador Minimax para θ sob a perda e a priori acima definidas.
- Sabemos que o estimador de Bayes, neste caso, é a esperança à posteriori, ou seja

$$\widehat{\theta}_{Bayes} = \frac{n\overline{X} + a}{n + a + b}$$

Voltando ao Exemplo 1 (cont.)

- Neste caso, temos que (Exercício):

$$\begin{aligned} R(\delta,\theta) &= \mathcal{E}_{\boldsymbol{X}|\theta}\left[\left(\frac{n\overline{X}+a}{n+a+b}-\theta\right)^2\right] \\ &= \frac{1}{(n+a+b)^2}\left[A\theta^2 - B\theta + a^2\right] \end{aligned}$$

em que $A = ((a+b)^2 - n)$ e $B = (2a(a+b) - n)$. Assim, $\widehat{\theta}_{Bayes}$ é minimax $\Leftrightarrow$ A $=$ B $= 0$, ou seja, $\Leftrightarrow a = b = \frac{\sqrt{n}}{2}$. Assim,

$$R(\widehat{\theta}_{Bayes},\delta) = \frac{1}{4(1+\sqrt{n})^2} \text{ e } \widehat{\theta}_{Bayes} = \frac{n\overline{X}+\sqrt{n}/2}{n+\sqrt{n}}$$