Princípios fundamentais da Inferência bayesiana

Prof. Caio Azevedo

Relembrando

- Posteriori : $p(\theta|\boldsymbol{x}) = \dfrac{p(\boldsymbol{x}|\theta)p(\theta)}{p(\boldsymbol{x})}$, em que

$$p(\boldsymbol{x}) = \begin{cases} \displaystyle\int_{\Theta} p(\boldsymbol{x}|\theta)p(\theta)d\theta, & \text{se } \theta \text{ for uma v.a.c.,} \\ \displaystyle\sum_{\theta \in \Theta} p(\boldsymbol{x}|\theta)p(\theta), & \text{se } \theta \text{ for uma v.a.d.} \end{cases}$$

- Se $p(\theta) \propto I_{\Theta}(\theta)$, então $p(\theta|\boldsymbol{x}) \propto p(\boldsymbol{x}|\theta)$ (a posteriori é proporcional à verossimilhança) e o estimador bayesiano MAP coincide (numericamente) com o EMV.
- Na obtenção da posteriori, tudo que não depender de θ, tanto na verossimilhança quanto na priori, pode ser desconsiderado.

Princípios da IB

- Os três princípios básicos da Inferência bayesiana são:
  - Princípio da Verossimilhança.
  - Princípio da Suficiência.
  - Princípio da Condicionalidade.
- A inferência frequentista compactua com o princípio da suficiência, pode ou não compactuar com o princípio da condicionalidade, mas não compactua com o princípio da verossimilhança.

Princípio da Verossimilhança

- Sejam $\boldsymbol{X} = (X_1, ..., X_n)' \in \Omega_x$ e $\boldsymbol{Y} = (Y_1, .., Y_n)' \in \Omega_y$ ($\Omega_{(.)}$ sendo o respectivo suporte das distribuições) dois vetores aleatórios (pertencentes ao mesmo espaço estatístico), que dependem do mesmo parâmetro θ e que possuem verossimilhanças distintas, mas que obedecem a seguinte relação :

$$p(\boldsymbol{x}|\theta) = K(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}|\theta) \propto p(\boldsymbol{y}|\theta), \forall \theta \in \Theta$$

em que $K(\boldsymbol{x}, \boldsymbol{y})$ não depende de θ. Então, sob a mesma priori, as posteriores obtidas a partir de $\boldsymbol{x}$ e $\boldsymbol{y}$ são iguais, ou seja

$$p(\theta|\boldsymbol{x}) = p(\theta|\boldsymbol{y})$$

Princípio da Verossimilhança (cont.)

- Demonstração (caso contínuo):

$$
\begin{aligned}
p(\theta|\boldsymbol{x}) &= \frac{p(\boldsymbol{x}|\theta)p(\theta)}{\int_{\Theta} p(\boldsymbol{x}|\theta)p(\theta)d\theta} = \frac{p(\boldsymbol{y}|\theta)K(\boldsymbol{x},\boldsymbol{y})p(\theta)}{\int_{\Theta} p(\boldsymbol{y}|\theta)K(\boldsymbol{x},\boldsymbol{y})p(\theta)d\theta} \\
&= \frac{p(\boldsymbol{y}|\theta)p(\theta)}{\int_{\Theta} p(\boldsymbol{y}|\theta)p(\theta)d\theta} = p(\theta|\boldsymbol{y})
\end{aligned}
$$

- Exercício: repetir para o caso discreto.

Princípio da Verossimilhança (cont.)

- Exemplo: Suponha que desejamos estimar θ (probabilidade de observar cara (C) no lançamento de uma moeda) e que, para um determinado experimento, observou-se

$$\{C, \overline{C}, C, C, \overline{C}, \overline{C}, C, \overline{C}, \overline{C}, \overline{C}\}$$

Princípio da Verossimilhança (cont.)

- Entre outras possibilidades, os dados acima poderiam ter sido gerados a partir dos seguintes experimentos:
  - Lançar a moeda 10 vezes e contabilizar o número de caras ($X \sim$ binomial(10,θ)).
  - Lançar uma moeda, até obter um total de 4 caras, contabilizando o número de lançamentos ($Y \sim$ Binomial-Negativa(4,θ)).

Princípio da Verossimilhança (cont.)

- Em termos de verossimilhança dos modelos:
  - Primeiro caso:

    $p(x|\theta) = \frac{10!}{x!(10-x)!}\theta^x(1-\theta)^{10-x}\mathbb{1}_{\{0,1,\ldots,10\}}(x) \propto \theta^x(1-\theta)^{10-x}$
  - Segundo caso:

    $p(y|\theta) = \frac{(y-1)!}{(y-4)!3!}(1-\theta)^{y-r}\theta^r\mathbb{1}_{\{r,r+1,\ldots,\}}(y) \propto (1-\theta)^{y-4}\theta^4.$
- Utilizando os dados observados nas verossimilhanças:
  - No primeiro caso, para os dados em questão, temos que:

    $p(6|\theta) \propto \theta^4(1-\theta)^6.$
  - No segundo caso, para os dados em questão, temos que:

    $p(10|\theta) \propto \theta^4(1-\theta)^6.$

Princípio da Verossimilhança (cont.)

- Logo,

$$p(x|\theta) \propto p(y|\theta)$$

  e, portanto, sob uma mesma priori para θ, a posteriori obtida a partir de x seria igual à posteriori obtida a partir de y. Ou seja a "inferência bayesiana" seria a mesma.

- Contudo, por exemplo, os estimadores de máxima verossimilhança (emv) de θ sob cada um dos modelos, seriam diferentes.

Princípio da suficiência

- Seja $T = g(\boldsymbol{X})$ uma estatística suficiente para um parâmetro θ. associado à uma verossimilhança $p(.|\theta)$. Então, a distribuição à posteriori dependerá apenas de T (ou, equivalentemente, de sua distribuição).
- Prova (caso contínuo): Se T é uma estatística suficiente, então pelo teorema da fatoração de Neyman, temos que:

$$p(\boldsymbol{x}|\theta) = g(t;\theta)h(\boldsymbol{x}),$$

assim, vem que

$$p(\theta|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\theta)p(\theta)}{\int_{\Theta} p(\boldsymbol{x}|\theta)p(\theta)d\theta} = \frac{g(t;\theta)h(\boldsymbol{x})p(\theta)}{\int_{\Theta} g(t;\theta)h(\boldsymbol{x})p(\theta)d\theta} = \frac{g(t;\theta)p(\theta)}{\int_{\Theta} g(t;\theta)p(\theta)d\theta}$$

Princípio da suficiência (cont.)

- Em geral, se $g(;\theta)$ não for a própria distribuição de T, está associada à ela.
- Com efeito, considere $X_1|\mu, ..., X_n|\mu \overset{iid}{\sim} N(\mu, \sigma^2)$, com σ^2 conhecido (exercício).
- Ao invés de usarmos a verossimilhança associada à amostra, podemos usar a distribuição da estatística suficiente $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Princípio da condicionalidade

- Suponha que dispomos de m experimentos possíveis de serem realizados, denotados por $E_1, ...E_m$, a fim de levantar informações (dados) para se fazer inferência à respeito de um parâmetro θ. Considere que tais experimentos, embora diferentes, sejam semelhantes entre si.
- Suponha que sorteamos um experimento, entre os m, ao acaso, e que o utilizamos para levantar informações. O princípio da condicionalidade diz que os outros experimentos que não foram sorteadas são irrelevantes para se estimar θ. Ou seja, apenas o experimento realizado é relevante.

Princípio da condicionalidade (cont.)

- Exemplo: (extraído de Paulino et al (2018)). Pretende-se testar as hipóteses $H_0 : \theta = 0$ vs $H_1 : \theta = 10$. Considere-se que o experimento E (a ser utilizado) é uma mistura de dois experimentos, E_1, escolhido com probabilidade p e consistindo em observar uma variável aleatória $X \sim N(\theta, 10)$ (ou seja, usar um instrumento menos preciso) e E_2, escolhido com probabilidade $1 - p$ e consistindo em observar uma variável aleatória $X' \sim N(\theta, 1)$ (ou seja, usar um instrumento mais preciso).

Princípio da condicionalidade (cont.)

- Suponha que um primeiro Estatístico adote o seguinte procedimento:
  - Se for escolhido o experimento E_1, rejeita H_0 quando $X > K_0$, com $\Phi(K_0/10) = 1 - \xi$.
  - Se for escolhido o experimento E_2, rejeita H_0 quando $X' > K_1$, com $\Phi(K_1) = 1 - \zeta$.

  e declara que seu teste tem nível de significância $\alpha = p\xi + (1 - p)\zeta$.
- Tal procedimento viola o princípio da condicionalidade.

Princípio da condicionalidade (cont.)

- Por outro lado, suponha que um segundo Estatístico adote o seguinte procedimento:
  - Se for escolhido o experimento E_1, ignora o que poderia acontecer se E_2 tivesse sido sorteado, rejeita H_0 quando $X > K_0$ com $\Phi(K_0/10) = 1 - \xi$ e declara que seu teste tem nível de significância ξ.
  - Se for escolhido o experimento E_2, ignora o que poderia acontecer se E_1 tivesse sido sorteado, rejeita H_0 quando $X' > K_0$ com $\Phi(K_1) = 1 - \zeta$ e declara que seu teste tem nível de significância ζ.
- Tal procedimento respeita o princípio da condicionalidade.

Princípio da condicionalidade (cont.)

- Suponha, ainda, um terceiro Estatístico, que adota o seguinte procedimento:
  - Se for escolhido o experimento E_1, ignora o que poderia acontecer se E_2 tivesse sido sorteado, observa $X = x$ e afirma interessar-se somente na evidência provida por x sobre H_0, ou seja $P(H_0|x)$.
  - Se for escolhido o experimento E_2, ignora o que poderia acontecer se E_1 tivesse sido sorteado, observa $X = x'$ e afirma interessar-se somente na evidência provida por x' sobre H_0, ou seja $P(H_0|x')$.
- Tal procedimento respeita o princípio da condicionalidade. Essencialmente, trata-se de usar a posteriori como estimador (bayesiano) de interesse.