Inferência Estatística com Abordagem Bayesiana

Rosangela Helena Loschi [1]

[1]Departamento de Estatística
Universidade Federal de Minas Gerais

16 de dezembro de 2021

INFERÊNCIA E DECISÃO

- Problema de Decisão
- Inferência Pontual
- Regiões de credibilidadee
- Testes de Hipóteses

Metodologia Inferencial

Em Estatística Bayesiana

* A melhor e mais completa inferência sobre o parâmetro de interesse é a distribuição *a posteriori*.
* A distribuição *a posteriori* é TUDO o que se necessita saber para se obter
  + Estimativas pontuais
  + Intervalos ou regiões de credibilidade
  + Testar hipóteses.
* As estimativas (pontuais ou por regiões) e a evidência sobre as hipóteses são resumos da distribuição *a posteriori*.
* Por tanto, a única amostra que se considera para se fazer inferência e tomar decisões é a amostra observada.

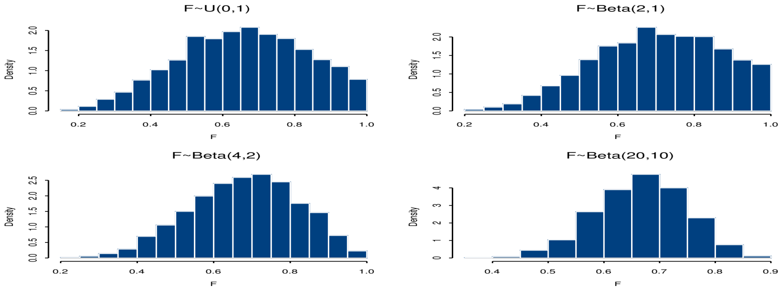
Metodologia Inferencial

- Na prática queremos resumir a informação atualizada para facilitar nossa comunicação.
- Esta análise *a posteriori* passa por
  - uma análise gráfica da distribuição *a posteriori*.
  - obter resumos de localização da distribuição *a posteriori* (como a média, a moda e a mediana que seriam nossas estimativas pontuais); resumos de dispersão (como a variância *a posteriori* e quantis o que nos dá uma idéia do grau de conhecimento adquirido sobre o parâmetro) etc.
- a análise gráfica é muito útil pois nos permite ver
  - as regiões do espaço paramétrico Θ com mais massa probabilística, revelando os valores mais prováveis de θ.
  - multimodalidade: os valores mais prováveis de θ pertencem a intervalos /regiões não conectadas
  - assimetria, revelando que valores de θ de um lado da distribuição *a posteriori* são mais prováveis do que os valores θ do outro lado da curva.

Análise dos Datos de Síndrome de Down: Distribuições *a posteriori*

$F = \phi$ probabilidade de que a não disjunção ocorra na primeria fase da meiose.

Figure: Distribuições *a posteriori* (método SIR)



Análise dos Datos de Síndrome de Down: Distribuições *a posteriori*

- Qual resumo melhor resumiria a distribuição *a posteriori*?
  - A moda parece sempre resumir bem a informação. Seria este o caso se a distribuição fosse bimodal??
  - No caso de distribuições assimétricas a média pode nao ser um bom resumo pois pode estar longe da região com mais massa probabilística.
- Se o pesquisador afirma que $\phi < 0,5$ que decisão você tomaria sobre esta afirmação?
  - Se *a priori* $\phi \sim Beta(20, 10)$, então $P(\phi < 0,5 \mid \boldsymbol{x}) \approx 0$. Evidência contra a afirmação do pesquisador.

Table: Análise dos Datos de Síndrome de Down Estimadores de Bayes

Especificação a Priori					*Resultados a Posteriori -SIR*		
α	β	Média	Var	Moda	Média	Var	Moda
1.0	1.0	0.500	0.080	–	0.6571	0.0311	0.6690
2.0	1.0	0.667	0.060	1.000	0.7043	0.0268	0.6842
4.0	2.0	0.667	0.030	0.750	0.6498	0.0194	0.7147
20.0	10.0	0.667	0.007	0.677	0.6671	0.0065	0.6765

- Quando *a priori* assumimos que $\phi \sim Beta(20, 10)$ obtivemos uma distribuição *a posteriori* mais simétrica.
- Consequência: as estimativas pontuais usando a média e a moda ficaram bem próximas.
- O que a variância nos diz? O qual certos estamos sobre o valor real de ϕ *a posteriori*.

Como escolher o estimador de Bayes?

Estimação Pontual

- A escolha dos estimadores Bayesianos de $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$ depende da forma da distribuição *a posteriori*.
- Ou da forma como escolhemos para sermos penalizados por nossas ações equivocadas (ver adiante).
- Os estimadores pontuais mais utilizados são média, moda e mediana *a posteriori*.
- Estes estimadores fornecem estimativas não muito similares se a distribuição *a posteriori* é assimétrica ou multimodal.
- Os resumos *a posteriori* no caso multiparamétrico são generalizações dos resumos no caso uniparamétrico.

Estimação Pontual

Seja $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ e denote por $\pi(\boldsymbol{\theta} \mid \boldsymbol{x})$ a distribuição *a posteriori* de $\boldsymbol{\theta}$.

- Moda *a posteriori*: é um vetor $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ tal que

$$\pi(\hat{\boldsymbol{\theta}} \mid \boldsymbol{x}) = \max_{\boldsymbol{\theta} \in \Theta} \pi(\boldsymbol{\theta} \mid \boldsymbol{x}) = \max_{\boldsymbol{\theta} \in \Theta} \pi(\boldsymbol{\theta}) f(\boldsymbol{x} \mid \boldsymbol{\theta})$$

- Média *a posteriori*: é um vetor $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ tal que

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta} \mid \boldsymbol{x}) = (E(\theta_1 \mid \boldsymbol{x}), \ldots, E(\theta_k \mid \boldsymbol{x}))$$

  onde $E(\theta_i \mid \boldsymbol{x}) = \hat{\theta}_i = \int_\Theta \theta_i \pi(\theta_i \mid \boldsymbol{x}) d\theta_i$, $i = 1, \ldots, k$.

- Mediana *a posteriori*: é um vetor $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ tal que

$$P(\theta_i \geq \hat{\theta}_i \mid \boldsymbol{x}) = 1/2 \ \text{ e } \ P(\theta_i \leq \hat{\theta}_i \mid \boldsymbol{x}) = 1/2, \quad i = 1, \ldots, k.$$

Estimação Pontual: Características importantes

a) Apenas a moda *a posteriori* pode ser obtida conhecendo-se apenas o núcleo da distribuição *a posteriori*.

b) para obtermos a média e a mediana *a posteriori* temos que conhecer a distribuição *a posteriori* completamente. No caso multiparamétrico temos que conhecer as distribuições marginais *a posteriori* completamente.

c) Se a distribuição *a priori* é uma constante K então a moda *a posteriori* é igual ao estimador de máxima verossimilhança (EMV) do parâmetro.

$$\max_{\boldsymbol{\theta}\in\Theta} \pi(\boldsymbol{\theta} \mid \boldsymbol{x}) = \max_{\boldsymbol{\theta}\in\Theta} \pi(\boldsymbol{\theta}) f(\boldsymbol{x} \mid \boldsymbol{\theta}) = \max_{\boldsymbol{\theta}\in\Theta} K f(\boldsymbol{x} \mid \boldsymbol{\theta}) = \max_{\boldsymbol{\theta}\in\Theta} f(\boldsymbol{x} \mid \boldsymbol{\theta})$$

d) o valor $\boldsymbol{\theta}$ onde a distribuição *a posteriori* assume valor máximo **não** é, necessariamente, o vetor formado pelas modas *a posteriori* das distribuições *a posteriori* marginais de cada componente de θ_i.

Exemplo (Ilustrando (c) e (d)): Se $X_1, \ldots, X_n$, dado μ e σ^2, são variáveis aleatórias i.i.d com distribuição Normal(μ, σ^2), em que μ e σ^2 são desconhecidas. Se, *a priori*, a distribuição para (μ, σ^2) é $\mu \mid \sigma^2 \sim Normal(m, V\sigma^2)$ e $\sigma^2 \sim Gama - Inversa(a, d)$. Então, A distribuição *a posteriori* é

$$\mu \mid \sigma^2, \boldsymbol{x} \sim N\left(\frac{nV\bar{x} + m}{nV + 1}, \frac{V}{nV + 1}\sigma^2\right) = N(m^*, V^*\sigma^2)$$

$$\begin{aligned}\sigma^2 \mid \boldsymbol{x} &\sim Gama - Inversa\left(a + \sum_{i=1}^{n} x_i^2 + \frac{m^2}{V} - \frac{(nV\bar{x} + m)^2}{v(nV + 1)}, d + n\right)\\ &= Gama - Inversa(a^*, d^*)\end{aligned}$$

Encontre os estimadores de Bayes para (μ, σ^2).

Solução:

- As distribuições *a posteriori* marginais de μ e σ^2, são:

$$\mu \mid \boldsymbol{x} \sim t\left(m^*, V^*, a^*, d^*\right)$$
$$\sigma^2 \mid \boldsymbol{x} \sim Gama - Inversa\left(a^*, d^*\right)$$

Estimação Pontual: Características importantes

- Média *a posteriori*:

$$(\hat{\mu}, \hat{\sigma^2}) = (E(\mu \mid \boldsymbol{x}), E(\sigma^2 \mid \boldsymbol{x})) = \left(m^*, \frac{a^*}{d^* - 2}\right)$$

- Mediana *a posteriori*:

$$(\hat{\mu}, \hat{\sigma^2}) = (Md(\mu \mid \boldsymbol{x}), Md(\sigma^2 \mid \boldsymbol{x})) = \left(m^*, F_{GI}^{-1}(1/2, a^*, d^*)\right)$$

onde $F_{GI}^{-1}(1/2, a^*, d^*)$ é a inversa da FDA da distribuição Gama-Inversa com parâmetros a^* e d^* avaliada no ponto $1/2$.

- Como fazemos. Seja M a mediana que procuramos. Entao

$$\int_0^M \pi(\sigma^2 \mid \boldsymbol{x}) d\sigma^2 = 1/2 \Rightarrow F_{GI}(M, a^*, d^*) = 1/2$$

$$\Rightarrow M = F_{GI}^{-1}(1/2, a^*, d^*).$$

Estimação Pontual: Características importantes

Solução:

- Moda *a posteriori*: Temos que encontrar $(\hat{\mu}, \hat{\sigma^2})$ que maximizam a distribuição a posteriori.
- maximizaremos $\ln \pi(\mu, \sigma^2 \mid \boldsymbol{x})$

$$\ln \pi(\mu, \sigma^2 \mid \boldsymbol{x}) = \ln K - \frac{d^* + 3}{2} \ln \sigma^2 - \frac{1}{2\sigma^2}\left[a^* + \frac{(\mu - m^*)^2}{V^*}\right]$$

$$\frac{\partial}{\partial \mu} \ln \pi(\mu, \sigma^2 \mid \boldsymbol{x}) = \frac{(\mu - m^*)}{V^* \sigma^2} = 0 \Rightarrow \hat{\mu} = m^*$$

$$\frac{\partial}{\partial \sigma^2} \ln \pi(\mu, \sigma^2 \mid \boldsymbol{x}) = -\frac{d^* + 3}{2\sigma^2} + \frac{(\mu - m^*)^2/V^* + a^*}{(\sigma^2)^2} = 0$$

$$\Rightarrow \hat{\sigma^2} = \frac{a^*}{d^* + 3}$$

- O ponto $(\hat{\mu}, \hat{\sigma^2}) = (m^*, a^*/(d^* + 3))$ é candidato a ponto de máximo.

Estimação Pontual: Características importantes

- a matriz Hessiana com as segundas derivadas parciais é:

$$H(\mu, \sigma^2) = \begin{bmatrix} -\frac{1}{V^*\sigma^2} & \frac{(\mu - m^*)V^*}{(\sigma^2)^2} \\ \frac{(\mu - m^*)V^*}{(\sigma^2)^2} & -\frac{2a^*}{(\sigma^2)^3} + \frac{(d^*+3)}{(\sigma^2)^2} \end{bmatrix}$$

- o determinante de $H(\mu, \sigma^2)$ avaliado em $(\hat{\mu}; \hat{\sigma^2}) = (m^*; a^*/(d^* + 3))$ é

$$\frac{1}{(a^*)^2} \frac{(d^* + 3)^4}{a^* V^*} > 0$$

- Como $\det H(\mu, \sigma^2) > 0$ e $\frac{\partial^2}{\partial \mu^2} < 0$ concluimos que a moda *a posteriori* é

$$(\hat{\mu}; \hat{\sigma^2}) = (m^*; a^*/(d^* + 3)).$$

Estimação Pontual: Características importantes

Observação importante relacionado ao Exemplo:

- Note que, neste caso, as distribuições marginais de μ e σ^2, são:

$$\mu \mid \boldsymbol{x} \sim t\,(m^*, V^*, a^*, d^*)$$
$$\sigma^2 \mid \boldsymbol{x} \sim Gama - Inversa\,(a^*, d^*)$$

- a moda das distribuições marginais são

$$\begin{aligned} Moda(\mu \mid \boldsymbol{x}) &= m^* \\ Moda(\sigma^2 \mid \boldsymbol{x}) &= \frac{a^*}{d^*+2} \end{aligned}$$

- Note que é diferente da moda encontrada anteriormente. Logo, a moda da distribuição conjunta não é formada pela moda das distribuições marginais.

Estimação Pontual: Características importantes

e) Os Estimadores de Bayes não possuem a propriedade da invariância observada para os estimadores de máxima verossimilhança.
   - Se $\hat{\theta}$ é o EMV para θ então o EMV para $g(\theta)$ é $g(\hat{\theta})$.
   - No caso Bayesiano, para encontrarmos a estimativa de $g(\theta)$, precisamos encontrar a distribuição *a posteriori* de $g(\theta)$.

Exemplo: Sejam $X_1, \ldots, X_n$ v.a. que, dado θ, sao i.i.d com distribuição $Bernoulli(\theta)$. Assuma a distribuição *a priori* de Bayes-Laplace para θ. (a) Estime θ^2 usando o método da Máxima Verossimilhança.

Solução: (a) Os EMV gozam da propriedade de invariância, isto é podemos utilizar o método *plug-in* para encontrarmos o EMV de θ^2. Pelo método da máxima verossimilhança, o estimador para θ é $\hat{\theta}_{MV} = \bar{X} \Rightarrow$ o EMV para θ^2 é $\hat{\theta^2}_{MV} = \bar{X}^2$.

Estimação Pontual: Características importantes

(b) Estime θ^2 utilizando a média *a posteriori.*
Solução: (b) Como a distribuição de Bayes-Laplace para θ é a $Uniforme(0,1) = Beta(1,1)$, por causa da conjugação temos que a distribuição *a posteriori* para θ é

$$\theta \mid \boldsymbol{x} \sim Beta\left(1 + \sum_{i=1}^{n} x_i; 1 + n - \sum_{i=1}^{n} x_i\right) = Beta(\alpha^*, \beta^*)$$

Podemos considerar duas estratégias

- Encontramos a distribuição *a posteriori* de θ^2 e, a partir dela, sua média.
- Utilizamos alguma propriedade que nos permita encontrar $E(\theta^2 \mid \boldsymbol{x})$.

Estimação Pontual: Características importantes

Usando propriedades de esperança temos

$$
\begin{aligned}
\hat{\theta^2} &= E(\theta^2 \mid \boldsymbol{x}) = V(\theta \mid \boldsymbol{x}) + [E(\theta \mid \boldsymbol{x})]^2 \\
&= \frac{\alpha^*\beta^*}{(\alpha^*+\beta^*)^2(\alpha^*+\beta^*+1)} + \left(\frac{\alpha^*}{(\alpha^*+\beta^*)}\right)^2 \\
&= \left(\frac{\alpha^*}{\alpha^*+\beta^*}\right)^2 \left[\frac{\beta^*}{\alpha^*(\alpha^*+\beta^*+1)} + 1\right] \qquad (1)
\end{aligned}
$$

- Se a propriedade da invariância fosse válida deveria valer o método *Plug-in*. Como a média *a posteriori* para θ é

$$\hat{\theta} = E(\theta \mid \boldsymbol{x}) = \frac{\alpha^*}{\alpha^*+\beta^*},$$

usando o método *plug-in*, a média *a posteriori* ou estimativa *a posteriori* de θ^2 deveria ser

$$\hat{\theta^2} = (E(\theta \mid \boldsymbol{x}))^2 = \left(\frac{\alpha^*}{\alpha^*+\beta^*}\right)^2$$

- Esta estimativa difere da estimativa em (1). Ela é errada pois estamos utilizando o método *plug-in* que não é aplicável no contexto Bayesiano.

Estimação Pontual: Características importantes

(b) Estime θ^2 utilizando a moda *a posteriori*.

- Neste caso precisamos encontra a distribuição *a posteriori* de $\phi = \theta^2$.
- usando resultados de probabilidade, para todo $\phi > 0$, temos

$$\pi(\phi \mid \boldsymbol{x}) = \frac{1}{2\sqrt{\phi}}[\pi_\theta(\sqrt{\phi} \mid \boldsymbol{x}) + \pi_\theta(-\sqrt{\phi} \mid \boldsymbol{x})]$$

- como $\theta \sim Beta(\alpha^*, \beta^*)$, então $\pi_\theta(-\sqrt{\phi} \mid \boldsymbol{x}) = 0$.
- A distribuição *a posteriori* para ϕ é

$$\pi(\phi \mid \boldsymbol{x}) = \frac{1}{2\sqrt{\phi}} \frac{\Gamma(\alpha^* + \beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} (\sqrt{\phi})^{\alpha^*-1} (1-\sqrt{\phi})^{\beta^*-1}, \quad 0 < \phi < 1.$$

$$\pi(\phi \mid \boldsymbol{x}) = \frac{\Gamma(\alpha^* + \beta^*)}{2\Gamma(\alpha^*)\Gamma(\beta^*)} (\sqrt{\phi})^{\alpha^*-2} (1 - \sqrt{\phi})^{\beta^*-1}, \quad 0 < \phi < 1. \tag{2}$$

Estimação Pontual: Características importantes

Para achar a moda *a posteriori* de $\phi = \theta^2$ encontramos o máximo da distribuição dada na equação (2).

$$\frac{d}{d\phi} \ln \pi(\phi \mid \boldsymbol{x}) = \frac{\alpha^* - 2}{2\phi} - \frac{(\beta^* - 1)\phi^{-1/2}}{2(1 - \sqrt{\phi})}$$

Igualando a zero temos

$$\frac{\alpha^* - 2}{2\phi} = \frac{(\beta^* - 1)\phi^{-1/2}}{2(1 - \sqrt{\phi})} \Rightarrow (\alpha^* - 2)(1 - \phi^{1/2}) = (\beta^* - 1)\phi^{1/2}$$

O candidato a máximo é

$$\hat{\phi} = \left(\frac{\alpha^* - 2}{\alpha^* + \beta^* - 3}\right)^2$$

e existe se $\alpha^* + \beta^* \neq 3$. (Deixo como exercício o cálculo e estudo da segunda para avaliar sob que condições $\hat{\phi}$ é ponto de máximo.)

Problema de decisão

Em Estatística Bayesiana, os problemas de estimação e teste de hipóteses são tratados como um problema de decisão.

- Elementos de um problema de decisão no contexto de inferência são:
  1. (1) Espaço de ações $\mathcal{A}$: conjunto de possíveis ações ou decisões ou respostas para o problema de inferência;
     - Exemplo: a) Aceitar ou rejeitar uma hipótese. b) fornecer uma estimativa para o parâmetro
  2. (2) Espaço Θ de todos os estados da natureza ou espaço paramétrico.
     - Exemplo: conjunto dos valores do parâmetro de interesse.
  3. (3) Espaço amostral $\mathcal{X}$: conjunto de todos os possíveis resultados de um experimento.
  4. (4) Regra de Decisão $a(x)$ é uma função definida de $\mathcal{X}$ em $\mathcal{A}$.
     - Exemplo: (a)Critério para a tomada de decisao sobre as hipóteses. (b) escolha de um estimador
  5. (5) Função perda $L(a, \theta) : \Theta \times \mathcal{A} \to \mathbb{R}^+$

Problema de decisão

A função perda $L(a,\theta)$ é uma escolha subjetiva que representa a perda que se sofre por escolher a ação $a(x) = a$ quando θ é o verdadeiro estado da natureza.

* Alguns exemplos
  (1) perda quadrática: $L(a,\theta) = (a-\theta)^2$
  (2) perda absoluta: $L(a,\theta) = |a-\theta|$
  (3) perda zero-umo: $L(a,\theta) = \lim_{\epsilon\to 0} 1(|a-\theta| > \epsilon)$.

- OBS: Se Θ é enumerável então a perda zero-um pode ser escrita da seguinte forma:

$$L(a,\theta) = 1 \text{ se } a \neq \theta$$

$$L(a,\theta) = 0 \text{ se } a = \theta$$

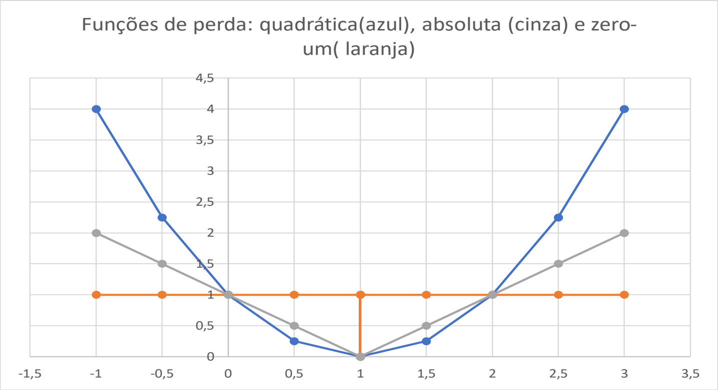
.

- No contexto Bayesiano, θ é aleatório pois é desconhecido.
- $L(a,\theta)$ é aleatória pois depende de θ.

Problema de decisão

Exemplo:

Suponha que se escolha a ação $a(x) = a = \bar{x}$. Se observamos que $\bar{x} = 1$ temos que, para todo $\theta \in \mathbb{R}$, a perda quadrática é $L(a, \theta) = (1 - \theta)^2$, a perda absoluta é $L(a, \theta) = |1 - \theta|$ e a perda zero-um é $L(a, \theta) = 0$ se $\theta = \bar{x} = 1$ e $L(a, \theta) = 1$ se $\theta = \bar{x} \neq 1$.



Problema de decisão

- A perda zero-um penaliza da mesma forma se você errou o valor de θ por pouco ou por muito.
- Se o valor real de θ dista pouco de 1 (está entre 0 e 2) a perda quadrática impõe a menor penalização de nossa decisão.
- Se o valor real de θ dista muito de 1 (é menor que 0 ou maior que 2) a perda quadrática impõe a maior penalização de nossa decisão.

Problema de decisão

- As decisões são feitas considerando o risco *a posteriori* das ações.
- Risco de uma ação ou decisão $a \in \mathcal{A}$ é a perda esperada *a posteriori* dada por

$$R(a, x) = \int_{\Theta} L(a, \theta)\pi(\theta|\mathrm{x})d\theta.$$

- A ação ótima ou regra de Bayes é a ação a^* tal que

$$a^* = min_a R(a, x).$$

- Risco de Bayes $\leftarrow$ é o risco associado à decisão ótima $R(a^*, x)$.

Problema de decisão

Exemplo (Migon & Gamerman (1993): Se você é um médico e tem que decidir se envia ($a = 1$) ou não ($a = 0$) um paciente à uma cirurgia. Se a probabilidade do paciente estar doente ($\theta = 1$) é π (desconhecido) e se sua função de perda é

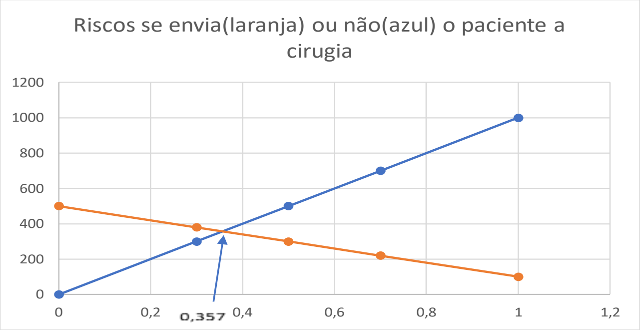
$$L(a,\theta) = \begin{cases} 0 & se \quad \theta = 0, a = 0 \\ 500 & se \quad \theta = 0, a = 1 \\ 1000 & se \quad \theta = 1, a = 0 \\ 100 & se \quad \theta = 1, a = 1 \end{cases} \tag{3}$$

Problema de decisão

Neste problema, os riscos associados a cada ação são:

$$
\begin{aligned}
R(0, x) &= 0(1-\pi) + 1000\pi = 1000\pi \\
R(1, x) &= 500(1-\pi) + 100\pi = 500 - 400\pi.
\end{aligned}
$$

Qual ação é melhor? Que decisão tomar?$\leftarrow$ depende da prevalência π da doença na população



Problema de decisão

- A decisão depende da incerteza que o médico tem sobre a prevalência da doença na população π.
- A decisão ótima é
  (1) Se $\pi > 0.357$, a regra de Bayes é $a = 1$ (indicar cirurgia)
  (2) Se $\pi < 0.357$, a regra de Bayes é $a = 0$ (não indicar cirurgia)
  (3) Se $\pi = 0.357$ não existe uma regra de Bayes.

Estimação Pontual como um problema de Decisão

- Em Estatística Bayesiana os problemas de Inferência podem ser vistos como problemas de decisão.

DEFINIÇÃO: Um estimador de Bayes é uma regra de Bayes segundo a função de perda que você escolhe, ou seja, se a função de perda é $L(a, \theta)$ então o estimador de Bayes é

$$\hat{\theta}_B = min_a E(L(a, \theta)|\mathrm{x})$$

Estimação Pontual como um problema de Decisão

Alguns Fatos Importantes:

(1) **Se assumimos a função de perda quadrática, então o estimador de Bayes é a média *a posteriori*.**

Prova:

$$
\begin{aligned}
E[L(a,\theta) \mid \mathrm{x}] &= E((\theta - a)^2 \mid \boldsymbol{x}) \\
&= E\left[[(\theta - E(\theta \mid \mathrm{x})) - (a - E(\theta \mid \mathrm{x}))]^2 \mid \mathrm{x}\right] \\
&= E\left[(\theta - E(\theta \mid \mathrm{x}))^2 \mid \boldsymbol{x}\right] - 2E\left[(\theta - E(\theta \mid \mathrm{x}))(a - E(\theta \mid \mathrm{x})) \mid \mathrm{x}\right] \\
&\quad + E\left[(a - E(\theta \mid \mathrm{x}))^2 \mid \boldsymbol{x}\right] \\
&= Var(\theta \mid \mathrm{x}) - 2(a - E(\theta \mid \mathrm{x}))E\left[(\theta - E(\theta \mid \mathrm{x})) \mid \mathrm{x}\right] \\
&\quad + E\left[(a - E(\theta \mid \mathrm{x}))^2 \mid \boldsymbol{x}\right] \\
&= Var(\theta \mid \mathrm{x}) + (a - E(\theta \mid \mathrm{x}))^2
\end{aligned}
$$

Como $(a - E(\theta \mid \mathrm{x}))^2 \geq 0$ para qualquer que seja a ação a, entâo $E[L(a,\theta) \mid \mathrm{x}]$ é mínima quando

$$a = E(\theta \mid \mathrm{x})$$

.

Estimação Pontual como um problema de Decisão

(2) **Se assumimos a função de perda zero-um, o estimador de Bayes é a moda *a posteriori*.**

Prova: Assuma Θ enumerável. Daí, a regra de Bayes é:

$$\begin{aligned}
\hat{\theta} &= min_{a\in\Theta} E(L(a,\theta) \mid \boldsymbol{x}) \\
&= min_{a\in\Theta} \{0 \times P(a=\theta \mid \boldsymbol{x}) + 1 \times P(a \neq \theta \mid \boldsymbol{x})\} \\
&= min_{a\in\Theta} \{P(a \neq \theta \mid \boldsymbol{x})\} \\
&= min_{a\in\Theta} \left\{ \sum_{\forall\theta\neq a} P(\theta \mid \boldsymbol{x}) \right\} \qquad (4)
\end{aligned}$$

Como a soma na expressão (4) envolve apenas valores maiores ou iguais a zero, $\sum_{\forall\theta\neq a} P(\theta \mid \boldsymbol{x})$ será mínima se excluírmos o maior valor de $P(\theta \mid \boldsymbol{x}) \Rightarrow a$ tem que ser o valor de θ tal que $P(\theta \mid \boldsymbol{x})$.

$\Rightarrow$ a ação ótima a tem que ser a moda a *a posteriori*.

Estimação Pontual como um problema de Decisão

Coloquemos em números: Considere $\Theta = \{1, 2, 3\}$ e que $P(\theta = 1 \mid \boldsymbol{x}) = 2/3$ e $P(\theta = 2 \mid \boldsymbol{x}) = P(\theta = 3 \mid \boldsymbol{x}) = 1/6$.

- Se $a = 1$ então
$$\sum_{\theta \neq 1} P(\theta \mid \boldsymbol{x}) = P(\theta = 2 \mid \boldsymbol{x}) + P(\theta = 3 \mid \boldsymbol{x}) = 2/6$$
- Se $a = 2$ então
$$\sum_{\theta \neq 2} P(\theta \mid \boldsymbol{x}) = P(\theta = 1 \mid \boldsymbol{x}) + P(\theta = 3 \mid \boldsymbol{x}) = 5/6$$
- Se $a = 3$ então
$$\sum_{\theta \neq 3} P(\theta \mid \boldsymbol{x}) = P(\theta = 1 \mid \boldsymbol{x}) + P(\theta = 2 \mid \boldsymbol{x}) = 5/6$$

Se escolhemos $a = 1$ minimizamos a soma $\sum_{\theta \neq a} P(\theta \mid \boldsymbol{x})$. Isto implica que $a = 1$ é a decisão ótima e $\theta = 1$ é a moda.

Estimação Pontual como um problema de Decisão

Prova para Θ geral: Seja $P_{\theta|\mathrm{x}}$ a medida *a posteriori* para θ.

$$
\begin{aligned}
R(a^*, x) &= min_{a\in\Theta}\{lim_{\epsilon\to 0}[1 - P(a-\epsilon < \theta < a+\epsilon \mid \mathrm{x})]\} \\
&= min_{a\in\Theta}\{1 - lim_{\epsilon\to 0} P(a-\epsilon < \theta < a+\epsilon \mid \mathrm{x})\}. \\
&= 1 - min_{a\in\Theta}\{lim_{\epsilon\to 0} P(a-\epsilon < \theta < a+\epsilon \mid \mathrm{x})\}.
\end{aligned}
$$

Como $\epsilon \to 0$, a sequência $P(a-\epsilon < \theta < a+\epsilon \mid \mathrm{x})$ define uma sequência que decresce para $\int_a^a dP_{\theta|\mathrm{x}}(a) = P_{\theta|\mathrm{x}}(a)$. Assim, temos que

$$R(a^*, x) = 1 - min_{a\in\Theta}[P_{\theta|\mathrm{x}}(a)] \to a = Mo(\theta \mid \mathrm{x}).$$

Estimação Pontual como um problema de Decisão

(3) **Se assumimos a função de perda absoluta, então o estimador de Bayes é a mediana *a posteriori*.**

Prova: Assuma que $\theta \in \mathbb{R}$. Neste caso, o risco *a posteriori* é

$$
\begin{aligned}
R(a,x) &= E(L(a,\theta) \mid \mathbf{x}) = \int_{-\infty}^{\infty} |a-\theta|\pi(\theta|\mathbf{x})d\theta \\
&= \int_{-\infty}^{a} (a-\theta)\pi(\theta|\mathbf{x})d\theta + \int_{a}^{\infty} (\theta-a)\pi(\theta|\mathbf{x})d\theta \\
&= a\int_{-\infty}^{a} \pi(\theta|\mathbf{x})d\theta - \int_{-\infty}^{a} \theta\pi(\theta|\mathbf{x})d\theta \\
&\quad + \int_{a}^{\infty} \theta\pi(\theta|\mathbf{x})d\theta - a\int_{a}^{\infty} \pi(\theta|\mathbf{x})d\theta \\
&= aF_{\theta|\mathbf{x}}(a) - a[1-F_{\theta|\mathbf{x}}(a)] - \int_{-\infty}^{a} \theta\pi(\theta|\mathbf{x})d\theta \\
&\quad + \left[\int_{-\infty}^{\infty} \theta\pi(\theta|\mathbf{x})d\theta - \int_{-\infty}^{a} \theta\pi(\theta|\mathbf{x})d\theta\right] \\
&= 2aF_{\theta|\mathbf{x}}(a) - a + E(\theta \mid \mathbf{x}) - 2\int_{-\infty}^{a} \theta\pi(\theta|\mathbf{x})d\theta \qquad (5)
\end{aligned}
$$

Estimação Pontual como um problema de Decisão

Para encontrarmos a regra de Bayes, precisamos encontrar a ação a que minimiza o risco

$$R(a,x) = 2aF_{\theta|\boldsymbol{x}}(a) - a + E(\theta \mid \boldsymbol{x}) - 2\int_{-\infty}^{a} \theta\pi(\theta|\boldsymbol{x})d\theta$$

Derivando $R(a,x)$ com respeito a a e igualando a zero temos:

$$\frac{d}{da}R(a,x) = 2F_{\theta|\boldsymbol{x}}(a) + 2a\pi(a \mid \boldsymbol{x}) - 1 - 2a\pi(a \mid \boldsymbol{x}) = 2F_{\theta|\boldsymbol{x}}(a) - 1 = 0$$

$$\Rightarrow 2F_{\theta|\boldsymbol{x}}(a) = 1$$

Logo a ação ótima é escolher a tal que

$$F_{\theta|\boldsymbol{x}}(a) = 1/2 \Rightarrow a = F^{-1}_{\theta|\boldsymbol{x}}(1/2) \Rightarrow a = Mediana(\theta \mid \boldsymbol{x})$$

Note que

$$\frac{d^2}{da^2}R(a,x) = 2\pi(a \mid \boldsymbol{x}) > 0 \tag{6}$$

Estimação Pontual: Aproximações Computacionais

Se tivermos uma amostra $\theta^{(1)}, \ldots, \theta^{(n)}$ da distribuição *a posteriori* $\pi(\theta \mid \mathrm{x})$, podemos obter aproximações para os resumos da distribuição *a posteriori* de forma simples:

- Média e Variância *a posteriori*

$$E(\theta \mid \mathrm{X}) \approx \sum_i \theta^{(i)}/n = \hat{\theta}$$

$$Var(\theta \mid \mathrm{X}) \approx \sum_i (\theta^{(i)} - \hat{\theta})^2/(n-1)$$

- Mediana e a moda *a posteriori* podem ser aproximadas pela mediana e moda da amostra.

Estimação Pontual: Aproximações Computacionais

- Para a moda *a posteriori*, uma melhor aproximação é obtida utilizando-se a seguinte estratégia

$$Mo(\theta \mid \mathrm{X}) \approx \max\{\mathrm{Ker}(\pi(\theta^{(1)} \mid \mathrm{x})), \ldots, \mathrm{Ker}(\pi(\theta^{(n)} \mid \mathrm{x}))\}.$$

- Para a mediana *a posteriori*
  - Agrupe os θs e Identifique a classe $[PI, PF]$ que contém a mediana;
  - Uma aproximação razoável da mediana é dada por:

$$Md(\theta \mid \boldsymbol{x}) \approx PI + \frac{(PF - PI)(0,5 - FA_{cam})}{(FA_{cm} - FA_{cam})}.$$

- Outra aproximação útil para a mediana: Obtenha as aproximações para a moda e média como descrito anteriomente e faça

$$Md(\theta \mid \boldsymbol{x}) \approx \frac{2E(\theta \mid \boldsymbol{x}) - Mo(\theta \mid \boldsymbol{x})}{3}$$

- Pode-se obter uma amostra da distribuição *a posteriori* de $g(\theta) \leftarrow g(\theta^{(1)}), \ldots, g(\theta^{(n)})$.