Introdução à Inferência Bayesiana

Prof. Caio Azevedo

Introdução e motivação

- Considere que desejamos estimar a proporção de eleitores que votarão no candidato A nas próximas eleições para governador do estado de São Paulo.
- Assuma que existam dois candidatos A e B e que cada eleitor votará apenas em um ou no outro (não é possível votos nulos, brancos etc).
- Para alcançar tal objetivo, entrevista-se eleitores (selecionados à partir de algum plano amostral) e a partir das respostas calcula-se estimativas pontuais e intervalares.
- Uma possibilidade: utilizar estimadores de máxima verossimilhança sob o modelo *Bernoulli*(θ) (considerando um plano amostral AAS_s sob certas condições).

Contexto estatístico

- Deseja-mos estimar $\theta \in \Theta = [0,1]$, Θ (espaço paramétrico).
- Dentro do contexto proposto (amostragem aleatória simples, sem reposição) e sob a inferência frequentista, considera-se $X_1, ..., X_n$ uma amostra aleatória de $X \sim Bernoulli(\theta)$, ou seja $X_i \overset{i.i.d.}{\sim} Bernoulli(\theta)$.
- Verossimilhança ($\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$)

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} P(X_i = x_i|\theta) \\
&= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}\mathbb{1}_{\{0,1\}}(x_i) = \theta^{n\overline{x}}(1-\theta)^{n(1-\overline{x})}
\end{aligned}
$$

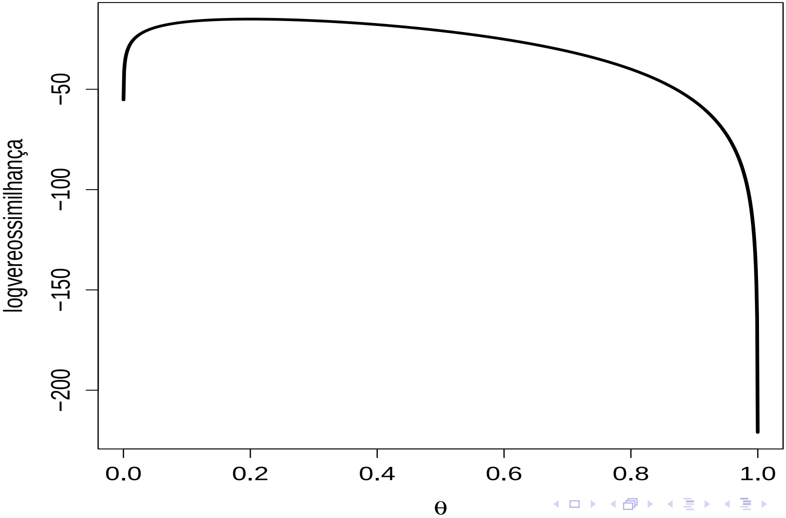
Contexto estatístico

- Logverossimilhança

$$l(\theta) = \ln L(\theta) \quad = \quad n\overline{x}\ln\theta + n(1-\overline{x})\ln(1-\theta)$$

- Os estimadores de máxima verossimilhança são obtidos maximizando-se a verossimilhança, ou, de modo equivalente (e mais simples) a log-verossimilhança.

Exemplo de logverossimilhança: $n = 30$ e $\overline{x} = 0,20$



Estimador de máxima verossimilhança

- Função (vetor) escore

$$S(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \frac{n\overline{x}}{\theta} - \frac{n(1-\overline{x})}{1-\theta}$$

- Matriz Hessiana

$$H(\theta) = \frac{\partial^2 l(\theta)}{\partial \theta^2} = -\frac{n\overline{x}}{\theta^2} - \frac{n(1-\overline{x})}{(1-\theta)^2}$$

- Informação de Fisher

$$I(\theta) = -\mathcal{E}(H(\theta)) = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)}$$

Estimador de máxima verossimilhança

- Estimador de máxima verossimilhança (exercício: provar que é ponto de máximo)

$$S(\widehat{\theta}) = 0 \rightarrow \widehat{\theta} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Notação :
  - $\widehat{\theta} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (estimador). $\mathcal{V}(\widehat{\theta}) = \dfrac{\theta(1-\theta)}{n}$.
  - $\widetilde{\theta} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (estimativa). $\mathcal{V}(\widetilde{\theta}) = 0$.
- Temos que $\widehat{\theta} \in \Theta = [0,1]$ e $\mathcal{E}(\widehat{\theta}) = \theta$.
- Note que $\mathcal{V}(\widehat{\theta}) = \dfrac{\theta(1-\theta)}{n}$ é um parâmetro.

Distribuições exata/aproximada de θ

- Qual a distribuição exata do estimador de MV? Sabemos que $S = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, \theta)$. Mas qual a distribuição exata de $\widehat{\theta} = \dfrac{S}{n}$?
- Distribuição assintótica do estimador de MV. Para n suficientemente grande:

$$\widehat{\theta} \approx N(\theta, I(\theta)^{-1}),$$

em que $I(\theta)^{-1} = \frac{\theta(1-\theta)}{n}$. Ou $\sqrt{n}\left(\dfrac{\widehat{\theta} - \theta}{\sqrt{\widehat{\theta}}}\right) \xrightarrow[n\to\infty]{D} N(0,1)$.
- A partir de que valor podemos considerar n suficientemente grande? Ou seja, a partir de que valor de n a aproximação acima pode ser considerada satisfatória?

Intervalo de confiança para θ

- Como desconhecemos a distribuição exata do EMV (estimador de máxima verossimilhança), uma alternativa é recorrer à distribuição assintótica de θ para construir intervalos de confiança e testar hipóteses.
- Intervalos de confiança:
  - IC($\theta;\gamma$): $\left[\widehat{\theta} - z_{\frac{1-\gamma}{2}}\sqrt{\frac{\widehat{\theta}(1-\widehat{\theta})}{n}};\widehat{\theta} + z_{\frac{1-\gamma}{2}}\sqrt{\frac{\widehat{\theta}(1-\widehat{\theta})}{n}}\right]$ (aleatório).
  - IC($\theta;\gamma$): $\left[\widetilde{\theta} - z_{\frac{1-\gamma}{2}}\sqrt{\frac{\widetilde{\theta}(1-\widetilde{\theta})}{n}};\widetilde{\theta} + z_{\frac{1-\gamma}{2}}\sqrt{\frac{\widetilde{\theta}(1-\widetilde{\theta})}{n}}\right]$ (numérico).
- $P(Z \geq z_{\frac{1-\gamma}{2}}) = \frac{1-\gamma}{2}, Z \sim N(0,1)$.

Voltando ao exemplo

- Suponha que numa amostra específica de tamanho $n = 30$ obtivemos $\overline{x} = 0,05$ e fixamos $\gamma = 0,95$. Neste caso

$$IC(\theta; 0,95) = [-0,03; 0,13]$$

- Como contornar o problema acima?
- Truncar o intervalo de confiança à esquerda do zero, ou seja

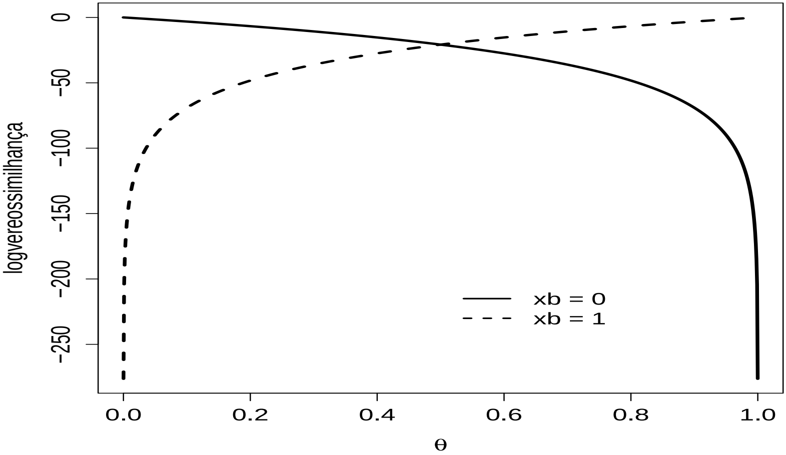
$$IC(\theta; 0,95) = [0,00; 0,13]$$

- Como interpretar os intervalos (aleatório e numérico) de confiança?

Mais sobre a verossimilhança

- Suponha que, para um dado n, $\overline{x} = 0$ ou $\overline{x} = 1$.
- Respectivamente, temos que $l(\theta) = n \ln(1-\theta)$ e $l(\theta) = n \ln(\theta)$.
- Quem seria o emv, em cada um desses casos?
- Entretanto, por convenção, adota-se, respectivamente, $\widehat{\theta} = \widetilde{\theta} = 0$ e $\widehat{\theta} = \widetilde{\theta} = 1$.
- Contudo, não quer dizer que, de fato, a proporção de eleitores seja 0 e 1, respectivamente!

logverossimilhanças para $\overline{x} \in \{0, 1\}$



Testes de hipótese

- Suponha que queiramos testar: $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$, θ_0 conhecido.
- Estatística do teste: $Z = \dfrac{\widehat{\theta} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}$
- Sob H_0, $Z \approx N(0,1)$, para n suficientemente grande.
- Procedimento, rejeita-se H_0 se $z_c \geq z_{\text{crítico}}$ em que $z_c = \dfrac{\widetilde{\theta} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}$, $P(Z \geq z_{\text{crítico}}|H_0) = \alpha$ e α é o nível de significância do teste.
- p-valor (nível descritivo): $P(Z \geq z_c|H_0)$. Rejeita-se H_0 se $p - valor \leq \alpha$.

Testes de hipótese

- Poder do teste $\psi(\theta) = P(\text{Rejeitar } H_0|H_1) = P(Z \geq z_{\text{crítico}}|\theta > \theta_0)$.
- Note que, sob H_1, $Z \approx N\left(\frac{\theta-\theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}, 1\right)$, para n suficientemente grande.
- Assim,

$$\begin{aligned}\psi(\theta) &= P\left(\underbrace{Z - \frac{\theta-\theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}}_{Z^*} \geq \underbrace{z_{\text{crítico}} - \frac{\theta-\theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}}_{z^*} \middle| \theta > \theta_0\right) \\ &= P(Z^* \geq z^*|\theta > \theta_0)\end{aligned}$$

em que $Z^* \sim N(0,1)$ e $z^* = z_{\text{crítico}} - \frac{\theta-\theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}$.

Voltando ao exemplo

- Suponha que no conjunto de hipóteses anterior, fixemos $\theta_0 = 0,05$.
- Para uma amostra $n = 30$, obtivemos $\overline{x} = 0,10$.
- Neste caso, o poder estimado é $\psi(\widetilde{\theta}) = \psi(\overline{x}) = 0,3489$.
- Mesmo com um teste com boas propriedades (UMP), o poder associado pode ser baixo.

Resumindo (limitações da abordagem frequentista no exemplo)

- Dependência de resultados assintóticos (tamanho da amostra suficientemente grande).
- Intervalos de confiança parcialmente fora do espaço paramétrico.
- Interpretação do IC numérico não é intuitiva (baseada na possibilidade de repetição do processo de inferência).
- Mesmo testes com boas propriedades podem apresentar baixo poder.
- Dificuldade de se implementar inferência fequentista em modelos complexos.

Outro ponto de interesse

- Quase sempre, antes de se realizar experimentos, levantamento de dados (pesquisa eleitoral), ou algum tipo de estudo, o pesquisador interessado possui alguma idéia à respeito do resultado esperado (conhecimento prévio).
- Na análise estatística de dados, isso, em geral, se traduz em termos do possível valor "verdadeiro" do parâmetro.
- Na inferência estatística frequentista (paramétrica), essencialmente, a única ligação entre os dados e o parâmetro é a verossimilhança.
- Em geral, o conhecimento prévio é incorporado apenas de modo parcial, através de verossimilhança, ou não o é.

Questão

- Questão: como incorporar de modo pleno e/ou para além da verossimilhança, algum conhecimento prévio existente?
- Uma solução: atribuir (à priori, ou seja, antes do experimento ser realizado) probabilidades à cada valor do espaço paramétrico.
- Voltando ao exemplo: é mais provável que o candidato A ganhe ou perca a eleição?

Observações

- Na abordagem frequentista, os problemas citados podem ser total ou parcialmente tratados:
  - Aproximações assintóticas: podem ser melhoradas através de métodos de correções assintóticas e/ou podemos trabalhar com aproximações empíricas (jacknife, bootstrap).
  - Incorporação de informações (à priori): por exemplo, trabalhar com log-verossimilhanças penalizadas.
  - Resultados numéricos que não pertençam ao espaço paramétrico: trabalhar com aproximações empíricas (jacknife, bootstrap) ou inferência restrita.

Inferência Frequentista (IF) $\times$ Inferência Bayesiana (IB)

- Inferência
  - IF: Utiliza estimadores, quantidades pivotais, estatísticas do teste. Ou seja utiliza-se o suporte das respectivas distribuições. Apesar desses conjuntos (suportes), eventualmente, coincidirem com o espaço paramétrico (parâmetro), eles nunca serão os mesmos.
  - IB: utiliza-se diretamente o espaço paramétrico (distribuição a posteriori).
- Interpretações
  - IF: baseada na possibilidade de se repetir o experimento e vista em função do espaço amostral.
  - IB: baseada diretamente no espaço paramétrico.

Inferência Frequentista (IF) $\times$ Inferência Bayesiana (IB)

- Probabilidade
  - IF: Vista como o limite das frequências relativas de eventos de interesse quando o número de repetições do experimento tende à infinito.
  - IB: utiliza-se diretamente o espaço paramétrico (distribuição a posteriori).
- Parâmetro
  - IF: quantidade não aleatória, geralmente fixa e não observável.
  - IB: "quantidade aleatória", geralmente fixa e não observável.

Inferência Frequentista (IF) × Inferência Bayesiana (IB)

- Observações
  - Todos os conceitos (suficiência, ancilaridade, completitude etc) vistos no contexto frequentista, podems er aplicados (adaptados) no contexto bayesiano.
  - Podemos comparar resultados frequentistas e bayesianos em ambos os contextos (frequentrista: variância, eqm, comprimento do IC, poder do teste etc)
  - Matematicamente, os estimadores de máxima verossimilhança são um caso particular dos estimadores bayesianos.

Teorema de Bayes

- Em 1763, o Reverendo Thomas Bayes, em seu artigo "An Essay towards Solving a Problem in the Doctrine of Chances", postulou o famoso Teorema de Bayes.
- Suponha que existam n urnas e que, em cada uma existam um total de $B_i, i = 1, ..., n$ bolas brancas e $A_i, i = 1, ..., n$ bolas azuis e que, dentro de cada urna, cada bola tenha a mesma probabilidade de ser sorteada. Além disso, considere que cada urna tenha uma probabilidade $P(U_i) = p_i, i = 1, ..., n$ de ser selecionada.
- Experimento: sorteia-se, ao acaso, uma urna e, dela, sorteia-se uma bola ao acaso.

Teorema de Bayes

- A probabilidade da bola sorteada ser azul é $P(A) = \sum_{i=1}^{n} P(U_i)P(A|U_i) = \sum_{i=1}^{n} p_i P(A|U_i)$ (teorema da probabilidade total).
- Dado que a bola sorteada é azul, qual a probabilidade de ter sido selecionada a urna $k, k \in \{1, ..., n\}$?

$$P(U_k|A) = \frac{P(U_k)P(A|U_k)}{\sum_{i=1}^{n} P(U_i)P(A|U_i)} = \frac{p_k P(A|U_k)}{\sum_{i=1}^{n} p_i P(A|U_i)}$$

Teorema de Bayes

- Atualizamos nosso conhecimento sobre a probabilidade de selecionarmos a urna k depois de termos observado uma bola azul.
- Implicitamente, na inferência frequentista assume que $p_i = \frac{1}{n}$.
- Paralelo: urnas - espaço paramétrico, bola(s) - amostra.

Voltando ao exemplo inicial

- Suponha que antes de realizarmos o experimento, tenhamos quatro "conjecturas" sobre as eleições.
- Cientista político 1: não faz idéia do resultado.
- Cientista político 2: acha que é mais provável que o candidato A perca.
- Cientista político 3: acha que é mais provável que o candidato A ganhe.
- Cientista político 4: acha que é mais provável que o candidato haja um empate.

Incorporando a incerteza sobre o problema

- Como incorporar cada um dos pontos de vista acima à análise dos dados (à verossimilhança)?
- Probabilidade (frequentista): limite de frequências relativas quando o número de repetições (do experimento) tende ao infinito. (inferência frequentista).
- Probabilidade (subjetivista): quantificação da incerteza acerca de algum fenômeno (experimento aleatório, pesquisa eleitoral etc...) (inferência bayesiana).

Incorporando a incerteza sobre o problema (cont.)

- Pode-se incorporar a incerteza ao problema (à verossimilhança) através de distribuições de probabilidade definidas para o parâmetro.
- No nosso caso $\theta \in (0, 1)$. Que distribuição de probabilidade pode ser associada ao parâmetros θ?
- Opções: beta, simplex, Kumaraswamy, Weibull unitária dentre outras.

Distribuição à priori

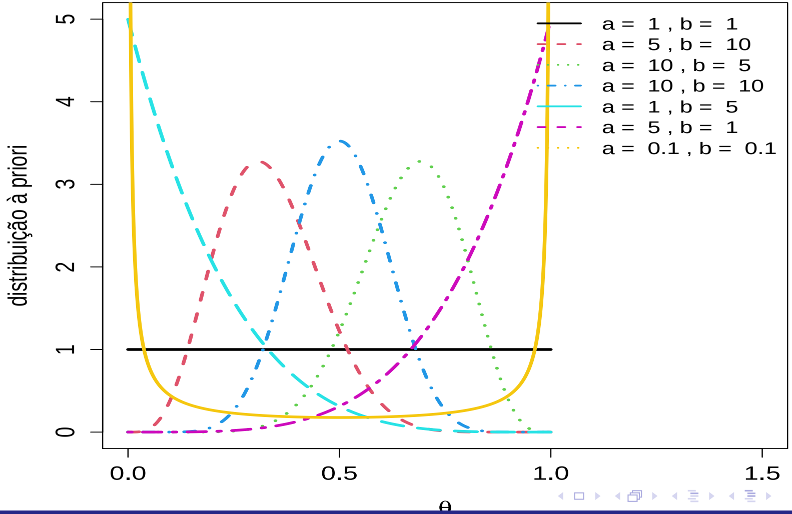
- Dizemos que $\theta \sim$ Beta(a,b) (beta), $(a, b)' \in R^+$, se sua fdp é da forma:

$$p(\theta) = \frac{1}{\beta(a, b)} \theta^{a-1} (1-\theta)^{b-1} \mathbb{1}_{(0,1)}(\theta)$$

em que $\beta(a, b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \displaystyle\int_0^1 x^{a-1}(1-x)^{b-1} dx,$

$$\Gamma(a) = \begin{cases} \int_0^{\infty} e^{-x} x^{a-1} dx, & \text{se a} \notin \mathcal{Z}^+ \\ (a-1)!, & \text{se a} \in \mathcal{Z}^+ \end{cases}$$

Exemplos de densidade da distribuição Beta



Voltando ao problema

- Notação:

$$p(x_i|\theta) = \begin{cases} P(X_i = x_i|\theta), & \text{caso discreto} \\ f(x_i|\theta), & \text{caso continuo} \end{cases}$$

e $\boldsymbol{x} = (x_1, ..., x_n)'$ (vetor coluna).

- Verossimilhança: $L(\theta) = p(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} p(x_i|\theta) = \theta^{n\overline{x}}(1-\theta)^{n(1-\overline{x})}$
- Priori: $p(\theta) = \frac{1}{\beta(a,b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0,1)}(\theta)$

Voltando ao problema (cont.)

- Posteriori :

$$p(\theta|\boldsymbol{x}) = \frac{p(\theta)p(\boldsymbol{x}|\theta)}{\int_0^1 p(\theta)p(\boldsymbol{x}|\theta)d\theta}$$

- Note que, se $p(\theta) \propto I_{(0,1)}(\theta)$, então

$$p(\theta|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\theta)}{\int_0^1 p(\boldsymbol{x}|\theta)d\theta} \propto p(\boldsymbol{x}|\theta) = L(\theta)$$

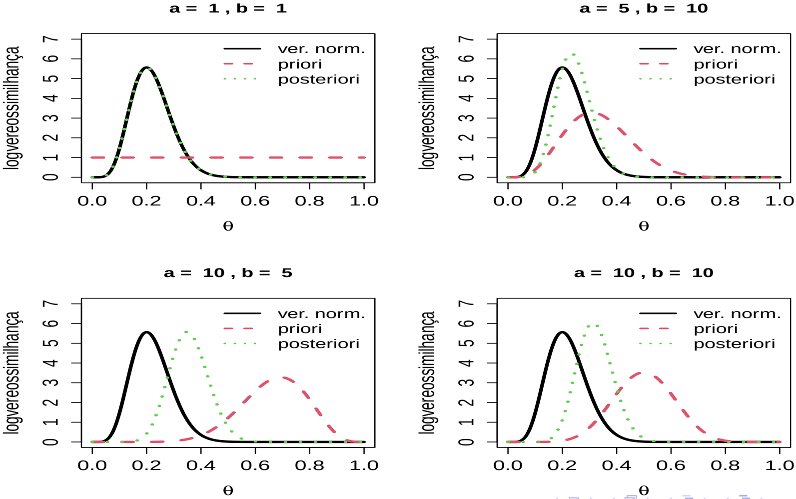
(inferência frequentista)

Obtendo a posteriori

$$
\begin{aligned}
p(\theta|\boldsymbol{x}) &= \frac{\frac{1}{\beta(a,b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0,1)}(\theta)\theta^{n\overline{x}}(1-\theta)^{n(1-\overline{x})}}{\int_0^1 \frac{1}{\beta(a,b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0,1)}(\theta)\theta^{n\overline{x}}(1-\theta)^{n(1-\overline{x})}d\theta} \\
&= \frac{\theta^{n\overline{x}+a-1}(1-\theta)^{n(1-\overline{x})+b-1}\mathbb{1}_{(0,1)}(\theta)}{\underbrace{\int_0^1 \theta^{n\overline{x}+a-1}(1-\theta)^{n(1-\overline{x})+b-1}d\theta}_{\beta(n\overline{x}+a,n(1-\overline{x})+b)}} \\
&= \frac{1}{\beta(n\overline{x}+a,n(1-\overline{x})+b)}\theta^{n\overline{x}+a-1}(1-\theta)^{n(1-\overline{x})+b-1}\mathbb{1}_{(0,1)}(\theta)
\end{aligned}
$$

Ou seja, $\theta|\boldsymbol{x} \sim \text{Beta}(n\overline{x}+a, n(1-\overline{x})+b)$.

Voltando ao exemplo n=30 e $\overline{x} = 0,20$



Introdução

- Inferência Bayesiana: aprendizado, combinação entre o conhecimento prévio e resultados de alguma pesquisa (no sentido amplo da palavra).
- Parâmetro: tratado como variável aleatória (pode ou não ser considerado "fixo").
- Incerteza: introduzida através da priori e atualizada combinando-a com a verossimilhança.

Definições e Notações

- Consideramos um modelo estatístico formado por $(\Omega, \mathcal{A}, \mathcal{P})$, (espaço amostral, sigma álgebra de eventos de Ω, família de medidas de probabilidade).
- $\mathcal{P} = \{\mathcal{P}_\theta : \theta \in \Theta\}$, $\Theta \subset \mathcal{R}^p$ (em geral infinito não-enumerável). Basicamente: $\mathcal{P}_\theta = F_{X|\theta}(.)$ (fda, postulado).
- Realizar inferência Bayesiana (estimação pontual, intervalar e testar hipóteses estatísticas) em um dado espaço estatístico.
- $p^* = \{p_\theta : \theta \in \Theta\}$ (distribuição à priori). Os elementos de p^* não precisam ser fdp's (função de probabilidade ou função densidade de probabilidade).

Continuação

- Suporte da fdp:
  - $\Omega = \{x : f_{X|\theta}(x) \geq 0\}$, $f_{X|\theta}(.)$ é a fdp associada à fda $F_{X|\theta}(.)$. Eventualmente os subíndices podem ser suprimidos.
- Assim, (Ω, Θ) passa a ser nosso conjunto de interesse.
- Amostra aleatória (aa): $X_1|\theta, ..., X_n|\theta$ são condicionalmente independentes (dado θ) e identicamente distrbuídas segundo $X|\theta \sim F_{X|\theta}$. Eventualmente, as suposições de independência e mesma distribuição podem ser relaxadas.
- Variável aleatória: $X_i|\theta$. Valor observado: x_i.

Continuação

- Se $X_1|\theta, ..., X_n|\theta$ é uma aa de $X|\theta$, então $X_1, .., X_n$ serão permutáveis (veremos mais adiante). $\underset{\sim}{X} = \boldsymbol{X} = (X_1, ..., X_n)$.
- Média à posteriori: Estimador $\widehat{\theta}_{EAP} = \mathcal{E}(\theta|\boldsymbol{X})$; Estimativa $\widetilde{\theta}_{EAP} = \mathcal{E}(\theta|\boldsymbol{x})$.
- Moda à posteriori: Estimador $\widehat{\theta}_{Mo} = \text{Moda}(\theta|\boldsymbol{X})$; Estimativa $\widetilde{\theta}_{Mo} = \text{Moda}(\theta|\boldsymbol{x})$.
- Mediana à posteriori: Estimador $\widehat{\theta}_{Md} = \text{Med}(\theta|\boldsymbol{X})$; Estimativa $\widetilde{\theta}_{Md} = \text{Med}(\theta|\boldsymbol{x})$.
- Medida de precisão (bayesiana) do estimador (variância a posteriori): $\mathcal{V}(\theta|\boldsymbol{x})$.

Revisão de Cálculo de probabilidades

- Usaremos, em geral, p(.) para denotarmos fdp's.
- Sejam X, Y e Z va's definidas em um mesmo espaço de probabilidade.

$$p(x|y) = \begin{cases} \dfrac{p(x,y)}{p(y)}, & \text{se } p(y) > 0, \\ 0, & \text{se } p(y) \end{cases}$$

$$p(x) = \begin{cases} \sum_y p(x,y), & \text{se y for discreto,} \\ \int_{\Omega_y} p(x,y)dy, & \text{se y for contínuo} \end{cases}$$

Revisão de Cálculo de probabilidades

- $E_X(X) = E_Y(E_{X|Y}(X|Y))$ e $\mathcal{V}_X(X) = \mathcal{E}_Y[\mathcal{V}_{X|Y}(X|Y)] + \mathcal{V}_Y[\mathcal{E}_{X|Y}[X|Y]]$.
- $p(x, y) = p(x|y)p(y) = p(y|x)p(x)$
- X e Y são independentes se e somente se $p(x, y) = p(x)p(y)$ ($\rightarrow p(x|y) = p(x)$ e $p(y|x) = p(y)$).
- X e Y são condicionalmente independentes dado Z se e somente se $p(x, y|z) = p(x|z)p(y|z)$ ($\rightarrow p(x|y, z) = p(x|z)$ e $p(y|x, z) = p(y|z)$).

Breve histórico

- Teorema de Bayes $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$, publicado em 1763, *An Essay towards solving a Problem in the Doctrine of Chances*.
- Fundamentos Bayesianos: Bruno de Finetti, Jim Berger entre outros.
- Existems algumas correntes de pensamento dentro da inferência Bayesiana:
  - Priori tem de ser completamente especificada usando conhecimento prévio.
  - Priori pode ser total ou parcialmente obtida através dos dados.

IF x IB

Aspecto	Inferência Frequentista	Inferência Bayesiana
Conhecimento básico:	verossimilhança	verossimilhança e priori
Parâmetros:	quantidades fixas	quantidades fixas mas de caráter aleatório ou totalmente aleatório
Teoria Assintótica:	Bastante utilizada	Pouco utilizada
Métodos computacionais	Extremamente dependente	Imprescindível na grande maioria das situações
Probabilidade:	limite de frequências relativas	medida de credibilidade

IF x IB

Aspecto	Inferência Frequentista	Inferência Bayesiana
Estimação pontual:	máxima verossimilhança, momentos, mínimos quadrados	momentos, quantis da posteriori e risco de Bayes
Estimação intervalar:	limites são aleatórios	parâmetro é aleatório
Teste de hipótese:	estatística, região crítica, poder, TUMP	probabilidade de cada hipótese ser verdadeira