Inferência Estatística com Abordagem Bayesiana

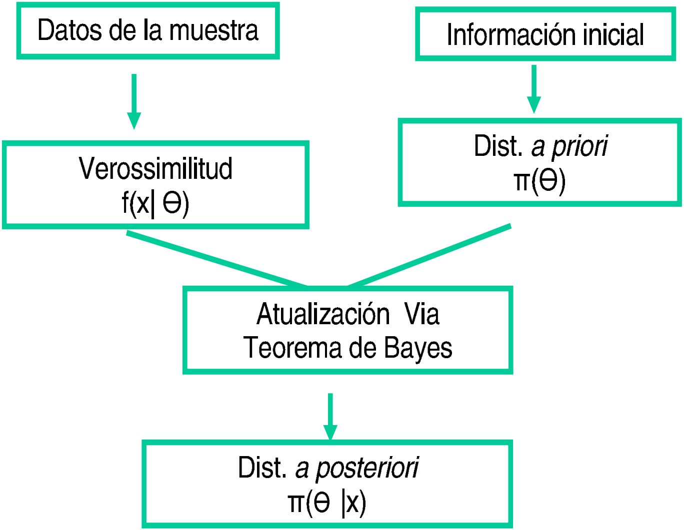
Rosangela Helena Loschi [1]

[1]Departamento de Estatística
Universidade Federal de Minas Gerais

29 de outubro de 2021

MODELAGEM BAYESIANA

Modelagem Bayesiana



Modelagem

A modelagem Bayesiana é composta de duas fases:

- A construção da função de verossimilhança;
- A construção da distribuição *a priori*

Denotemos por

- θ é o parâmetro de interesse e Θ é o espaço paramétrico;
- $\boldsymbol{x} = (x_1, \ldots, x_n)$ é a amostra observada da população alvo e $\mathcal{X}$ é o espaço amostral.
- $\pi(\theta)$ é a distribuição *a priori* sobre θ.
- $f(\boldsymbol{x} \mid \theta)$ é a função de verossimilhança
- $\pi(\theta \mid \boldsymbol{x})$ é a distribuição *a posteriori* sobre θ.

Modelagem

A função de Verossimilhança: Resume a informação trazida pelos dados sobre θ. É construída considerando-se:

- considerações teóricas sobre a natureza do fenômeno estudo.
- considerações teóricas sobre as técnicas experimentais, etc.

Definição: A função de verossimilhança $f(\boldsymbol{x} \mid \theta)$ é uma função definida de $\Theta \times \mathcal{X}$ em $\mathbb{R}_+$ tal que:

- fixado θ, $f(\boldsymbol{x} \mid \theta)$ é uma medida de probabilidade, condicional em θ, representando a incerteza associada à observação de cada valor $\boldsymbol{x}$ de $\boldsymbol{X}$, se θ é conhecido.
- Fixado $\boldsymbol{X} = \boldsymbol{x}$, $f(\boldsymbol{x} \mid \theta)$ é uma função de θ (sem interpretação probabilística.)

Modelagem

Exemplo: Seja $\theta > 0$ a taxa de ocorrência de um evento e seja X_i o tempo até a ocorrência deste evento na unidade i. Assuma que dado θ, a amostra $X_1, \ldots, X_n$ seja i.i.d. com distribuição exponencial com parâmetro θ. Construa a função de verossimilhança.

Solucao: Como a amostra é iid, temos que

$$f(\boldsymbol{x} \mid \theta) = \prod_{i=1}^{n} \theta \exp\{-\theta x_i\} = \theta^n \exp\{-\theta \sum_{i=1}^{n} x_i\}$$

Note que se $n = 2$ e fixarmos $\theta = 1$ a função de verossimilhança fornece a probabilidade de ocorrência de cada amostra retirada da população em que $\theta = 1$.

Table: Exemplos-Verossimilhança fixado $\theta = 1$

Amostra	verossimilhança	
(0,1)	$\exp\{-1\}$	$= f_{X_1, X_2}((0,1) \mid \theta = 1)$
(1,1)	$\exp\{-2\}$	
(1,2)	$\exp\{-3\}$	
(0.5,0.5)	$\exp\{-1\}$	

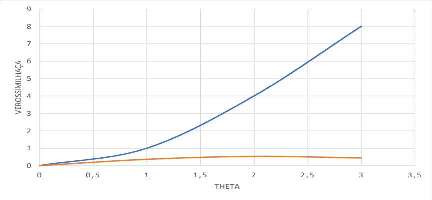
Modelagem

Se a função de verossimilhança é avaliada na amostra observada, obteremos uma função de θ.

Table: Exemplos-Verossimilhança fixada a amostra observada

Amostra	verossimilhança	Cor-gráfico
(0,0)	θ^2	(azul)
(1,0)	$\theta^2 \exp\{-\theta\}$	(laranja)

Note que a verossimilhança, quando vista como função de θ, pode assumir valores superiores a 1.



Modelagem

Distribuição *a priori*: Deve representar e modelar todo o conhecimento inicial, nao proveniente da amostra observada, que temos sobre θ. Existem vários métodos para construírmos a distribuição *a priori*. Discutiremos alguns.

- Método subjetivista (veremos o próximo exemplo.)
- Conjugação
- Distribuições de referência

A Inferência

Distribuição *a posteriori*: É uma atualização da distribuição *a priori* sobre θ. Usamos o Teorema de Bayes para misturar a informação dos dados e a informação *a priori*.

- Se $\Theta = \{\theta_1, \ldots, \theta_n\}$ é um conjunto finito ou é um conjunto infinito enumerável $\Theta = \{\theta_1, \theta_2, \ldots\}$ e, *a priori*, $\pi(\theta_i) = P(\theta = \theta_i)$ então a distribuição *a posteriori* é

$$\pi(\theta_i \mid \boldsymbol{x}) = P(\theta = \theta_i \mid \boldsymbol{x}) = \frac{f(\boldsymbol{x} \mid \theta = \theta_i)P(\theta = \theta_i)}{\sum_{j=1} f(\boldsymbol{x} \mid \theta = \theta_j)P(\theta = \theta_j)}$$

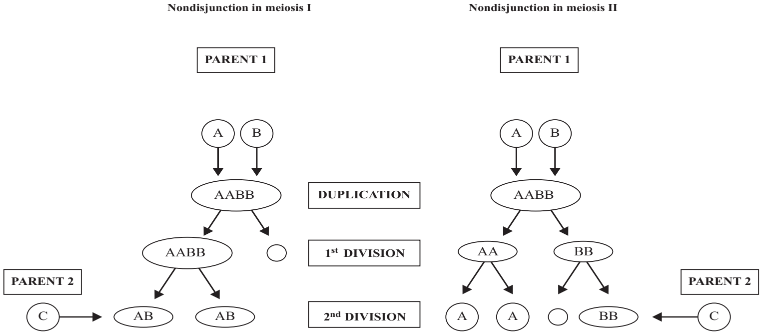
- Se $\Theta \subset \mathbb{R}$ é um conjunto denso e, *a priori*, $\pi(\theta)$ então a distribuição *a posteriori* é

$$\pi(\theta \mid \boldsymbol{x}) = \frac{f(\boldsymbol{x} \mid \theta)\pi(\theta)}{\int_\Theta f(\boldsymbol{x} \mid \theta)\pi(\theta)d\theta}$$

Modelagem:Datos de Síndrome de Down

- **Amostra:** Pacientes com Síndrome de Down (Trisomia no cromossomo 21)
- Esta síndrome é consequência da não-disjunção meiótica.
- **Meta:** Determinar a probabilidade ϕ de que a não-disjunção ocorra na primeira fase da Meiosis.
- **Importância:**Identificar fatores que poderiam gerar tal doença.

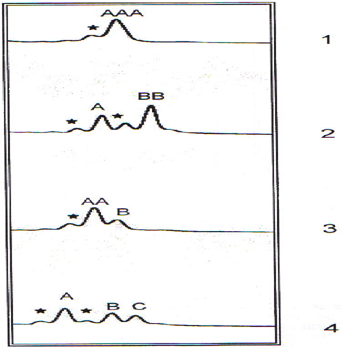
A não-disjunção meiótica pode ser representada da seguinte forma:



Modelagem:Datos de Síndrome de Down

Coleta de los Datos

- Uma amostra de sangue da pessoa con síndrome de Down é submetida ao procedimento PCR (Polimerase chain reaction) seguido de uma análise via densitometria laser.
- Obtemos o número de alelos distintos que o indivíduo tem no loco de interesse.



Modelagem:Datos de Síndrome de Down

Construção da função de verossimilhança

- Supõem-se que a hipótese do equilíbrio de Hardy-Weinberg é verificada para a população.
- Além disto, em pacientes trisômicos, é possivel identificar no loco de interesse o seguinte:
  - **Três picos:** (i) se a não-disjunção ocorre na Meiosis I, (ii) se a mãe é heterozigota e (iii) se o alelo do pai neste loco é distinto dos alelos maternos.
  - **Dois picos:**(i) se a não-disjunção ocorre ou na Meiosis I ou na Meiosis II e depende da combinação de cromossomos transmitido pelos pais.
  - **Um pico:** se a não-disjunção ocorre ou na Meiosis I ou na Meiosis II e os pais são homozigotos.
- Seleciona-se na população uma amostra i.i.d de n indivíduos com trissomia (síndrome de Down).

Modelagem:Datos de Síndrome de Down

Sob tais supostos e denotanto por

- $\boldsymbol{Y} = (Y_1, Y_2, Y_3)$ e Y_i o número de indivíduos na amostra que têm i, $i = 1, \ldots, 3$, picos no loco de interesse;
- ϕ probabilidade de que a não-disjunção ocorra na meiose I.
- p_i frequência do alelo i na população (conhecida);

A função de verossimilhança é

$$P(\boldsymbol{Y} = \boldsymbol{y} \mid \phi) = \frac{n!}{y_1!y_2!y_3!}[\theta_1(\phi)]^{y_1}[\theta_2(\phi)]^{y_2}[\theta_3(\phi)]^{y_3},$$

onde $\sum_{l=1}^{3} y_l = n$ e

$$\begin{aligned}
\theta_1(\phi) &= \phi\sum_{i=1}^{m} p_i^3 + (1-\phi)\sum_{i=1}^{m} p_i^2; \\
\theta_2(\phi) &= 3\phi\sum_{i=1}^{m}\sum_{j=1}^{m} p_i^2 p_j + (1-\phi)\sum_{i=1}^{m}\sum_{j=1}^{m} p_i p_j, \quad \text{para} \quad i \neq j; \\
\theta_3(\phi) &= \phi\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} p_i p_j p_k, \quad \text{para} \quad i \neq j \neq k.
\end{aligned}$$

Modelagem:Datos de Síndrome de Down

Construção da distribuição *a priori* :

- O parâmetro de interesse é $\phi \in (0,1)$
- Para construírmos a distribuição *a priori*, utilizados duas abordagens.
  - Assumimos que não tínhamos qualquer informação sobre ϕ

$$\phi \sim Uniforme(0,1) \equiv Beta(1,1)$$

  - Como não tínhamos contato com um especialista, buscamos outras fontes de informação, obtendo:

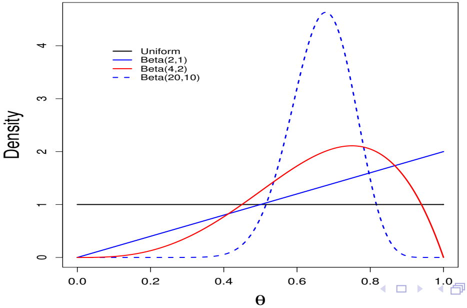
Table: Algumas estimativas de ϕ para outras populações

Referencia	n	$\hat{\phi}$	Referencia	n	$\hat{\phi}$
Lorber et al. (1992)	52	0.5192	Koehler *et al.*(1996)	776	0.7384
Petersen *et al.*(1992)	60	0.6833	Yoon *et al.*(1996)	103	0.6893
Zaragosa *et al.*(1994)	249	0.6867	Savage *et al.*(1998)	606	0.6930
Griffin (1996)	436	0.7133	Nicolaidis y Petersen (1998)	797	0.7189

Modelagem:Datos de Síndrome de Down

Baseando-se nestas informações construímos a distribuição de ϕ como segue:

* Consideramos a média 0.667 das estimativas de ϕ destas populações.
* Como $\phi \in (0, 1) \leftarrow$ Distribuições Beta são escolhas possíveis.
* Qual? Fizemos uma análise de sensibilidade escolhendo distribuições *a priori* com média 0.667 e diferentes graus de incerteza (controlado pela variância de ϕ)



A distribução *a posteriori*:

Se *a priori* $\phi \sim \mathcal{B}(\alpha, \beta)$, $\alpha, \beta \in \mathcal{R}_+$,

$$\pi(\phi) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\phi^{\alpha-1}(1-\phi)^{\beta-1} \quad \phi \in (0,1).$$

a distribuição *a posteriori* para ϕ é

$$\pi(\phi|y) = \frac{[\theta_1(\phi)]^{y_1}[\theta_2(\phi)]^{y_2}[\theta_3(\phi)]^{y_3}\phi^{\alpha-1}(1-\phi)^{\beta-1}}{\int_0^1 [\theta_1(\phi)]^{y_1}[\theta_2(\phi)]^{y_2}[\theta_3(\phi)]^{y_3}\phi^{\alpha-1}(1-\phi)^{\beta-1}d\phi}, \quad \phi \in (0,1).$$

* Não é completamente conhecida.
* Utilizamos métodos numéricos e de reamostragem para aproximá-la.

Modelagem:Datos de Síndrome de Down

* Datos: 34 indivíduos brasileiros com Síndrome de Down dos quais 6, 22 e 6, respectivamente, têm 1,2 e 3 alelos distintos no loco de interesse.
* frequências alélicas: temos somente 6 alelos na população com frequências 0.12, 0.45, 0.09, 0.31, 0.01 e 0.02.
* Estimador de máxima verosimilhança e variância assintótica são 0.6552 e 0.0481, respectivamente.

Figure: Distribuições *a priori* e *a posteriori* (método SIR) para ϕ

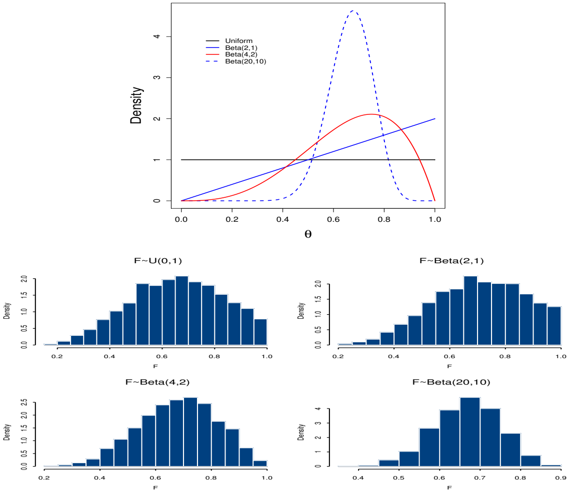


Table: Estimadores de Bayes

Especificação a Priori					*Resultados a Posteriori -SIR*			*Resultados a Posteriori-Simpson*		
α	β	Média	Var	Moda	Média	Var	Moda	Média	Var	Moda
1.0	1.0	0.500	0.080	–	0.6571	0.0311	0.6690	0.6549	0.0305	0.6553
2.0	1.0	0.667	0.060	1.000	0.7043	0.0268	0.6842	0.7015	0.0272	0.7244
4.0	2.0	0.667	0.030	0.750	0.6498	0.0194	0.7147	0.6814	0.0195	0.7013
20.0	10.0	0.667	0.007	0.677	0.6671	0.0065	0.6765	0.6670	0.0063	0.6753

A posteriori

- Temos mais certeza sobre o valor verdadeiro de ϕ, qualquer que seja a distribuição *a priori* escolhida.
- Se *a priori* $\phi \sim Uniforme(0, 1)$ e usando Simpson, mais provavelmente ϕ é $\approx 0,6553$
- **Curiosidade:** a Estimativa de máxima verosimilhança é 0.6553. Coincide com a moda *a posteriori* se a distribuição *a priori* assumida é a $Uniforme(0, 1)$.

Predição *a priori* e *a posteriori*

Distribuição preditiva *a priori:* É a distribuição de $\boldsymbol{X}$. Usamos o Teorema da probabilidade total e obtemos

- Se $\Theta = \{\theta_1, \ldots, \theta_n\}$ é um conjunto finito ou é um conjunto infinito enumerável $\Theta = \{\theta_1, \theta_2, \ldots\}$ e, *a priori*, $\pi(\theta_i) = P(\theta = \theta_i)$ então a distribuição preditiva *a priori* é

$$f(\boldsymbol{x}) = \sum_{j=1}^{n} f(\boldsymbol{x} \mid \theta = \theta_j) P(\theta = \theta_j)$$

- Se $\Theta \subset \mathbb{R}$ é um conjunto denso e, *a priori*, $\pi(\theta)$ então a distribuição preditiva *a priori* é

$$f(\boldsymbol{x}) = \int_{\Theta} f(\boldsymbol{x} \mid \theta) \pi(\theta) d\theta$$

- Se $\boldsymbol{X}$ é um vetor/variável discreta esta distribuição fornece a probabilidade de ocorrência da amostra observada

Exemplo: Seja X_i uma variável binária, i.é, $X_i \in \{0, 1\}$, $i = 1, 2$. Denote por $\theta = P(X_i = 1)$. Se $\Theta = \{1/3, 1/2\}$ e, *a priori* temos que $P(\theta = 1/3) = P(\theta = 1/2) = 0, 5$. Qual é a probabilidade de ocorrência da amostras $(0, 0)$?

- que função fornece resposta a esta pergunta: a função de verossimilhança ou a distribuição distribuição preditiva *a priori*?

Predição *a priori* e *a posteriori*

- Preditiva *a priori*

$$\begin{aligned}
P(X_1 = x_1, X_2 = x_2) &= P(X_1 = x_1, X_2 = x_2 \mid \theta = 1/3)P(\theta = 1/3) \\
&+ P(X_1 = x_1, X_2 = x_2 \mid \theta = 1/2)P(\theta = 1/2) \\
&= (1/3)^{x_1+x_2}(2/3)^{2-x_1-x_2}0,5 \\
&+ (1/2)^{x_1+x_2}(1/2)^{2-x_1-x_2}0,5
\end{aligned}$$

- verossimilhança fixado, por exemplo, $\theta = 1/3$

$$P(X_1 = x_1, X_2 = x_2 \mid \theta = 1/3) = (1/3)^{x_1+x_2}(2/3)^{2-x_1-x_2}$$

Predição *a priori* e *a posteriori*

Table: Distribuição Preditiva *a priori* e função de verossimilhança avaliada em $\theta = 1/3$

Amostra	Preditiva a priori	Verossimilhança fixado $\theta = 1/3$
$(0,0)$	$1/2((2/3)^2 + (1/2)^2) = 25/72$	$(2/3)^2 = 4/9$
$(1,0)$	$1/2((2/3) * (1/3) + (1/2)(1/2)) = 17/72$	$1/3 * 2/3 = 2/9$
$(0,1)$	$1/2((2/3) * (1/3) + (1/2)(1/2)) = 17/72$	$1/3 * 2/3 = 2/9$
$(1,1)$	$1/2((1/3)^2 + (1/2)^2) = 13/72$	$(1/3)^2 = 1/9$

- A função de verossimilhança fornece a probabilidade de ocorrência de cada amostra se soubessemos de que população a extraímos. No caso, extraímos da população em que $\theta = 1/3$
- A distribuição preditiva *a priori* fornece a probabilidade de ocorrência de cada amostra. Note que ela agrega a incerteza que temos sobre a população da qual extraímos a amostra.

Predição *a priori* e *a posteriori*

Distribuição preditiva *a posteriori:* Seja $\boldsymbol{X}$ a amostra observada e seja $\boldsymbol{Y}$ valores futuros da variável aleatória de interesse ainda não observados. Seja $\pi(\theta)$ a distribuição *a priori* descrevendo a incerteza inicial sobre o parâmetro populacional de interesse.

- Se $\Theta \subset \mathbb{R}$ é um conjunto denso então a distribuição preditiva *a posteriori* é

$$
\begin{aligned}
f(\boldsymbol{Y} \mid \boldsymbol{X}) &= \frac{f(\boldsymbol{Y}, \boldsymbol{X})}{f(\boldsymbol{X})} = \frac{\int_\Theta f(\boldsymbol{Y}, \boldsymbol{X} \mid \theta)\pi(\theta)d\theta}{f(\boldsymbol{X})} \\
&= \int_\Theta f(\boldsymbol{Y} \mid \boldsymbol{X}, \theta)\frac{f(\boldsymbol{X} \mid \theta)\pi(\theta)}{f(\boldsymbol{X})}d\theta \\
&= \int_\Theta f(\boldsymbol{Y} \mid \boldsymbol{X}, \theta)\pi(\theta \mid \boldsymbol{X})d\theta.
\end{aligned}
$$

- Se, dado θ, $\boldsymbol{X}$ e $\boldsymbol{Y}$ são independentes, temos

$$
f(\boldsymbol{Y} \mid \boldsymbol{X}) = \int_\Theta f(\boldsymbol{Y} \mid \theta)\pi(\theta \mid \boldsymbol{X})d\theta.
$$

Predição *a priori* e *a posteriori*

- Se Θ é um conjunto finito ou infinito enumerável temos *a priori* que $\pi(\theta_i) = P(\theta = \theta_i)$ então a distribuição preditiva *a posteriori* é

$$f(\boldsymbol{Y} \mid \boldsymbol{X}) = \sum_i f(\boldsymbol{Y} \mid \boldsymbol{X}, \theta_i) P(\theta = \theta_i \mid \boldsymbol{X})$$

- Se, dado θ, $\boldsymbol{X}$ e $\boldsymbol{Y}$ são independentes, temos

$$f(\boldsymbol{Y} \mid \boldsymbol{X}) = \sum_i f(\boldsymbol{Y} \mid \theta_i) P(\theta = \theta_i \mid \boldsymbol{X})$$

Predição *a priori* e *a posteriori*

Exemplo: Seja $\Theta = \{1/3, 1/2\}$ e *a priori* assuma que $P(\theta = 1/3) = P(\theta = 1/2) = 0,5$. Se, ao realizarmos o experimento, a amostra $(X_1, X_2) = (1, 1)$ foi observada, qual é a predição que se faz para uma próxima observação Y deste mesmo experimento?

Solução: A primeira coisa a fazer é encontrar a distribuição *a posteriori* de θ se a amostra $(x_1, x_2) = (1, 1)$ foi observada.

$$\begin{aligned} P(\theta = 1/3 \mid X_1 = 1 X_2 = 1) &= \frac{P(X_1 = 1 X_2 = 1 \mid \theta = 1/3) P(\theta = 1/3)}{P(X_1 = 1 X_2 = 1)} \\ &= \frac{1/9 * 1/2}{13/72} = 4/13 \\ P(\theta = 1/2 \mid X_1 = 1 X_2 = 1) &= 9/13 \end{aligned} \tag{1}$$

Predição *a priori* e *a posteriori*

Note que $Y \in \{0,1\}$. A distribuição preditiva *a posteriori* é

$$
\begin{aligned}
P(Y=0 \mid X_1=1 X_2=1) &= P(Y=0 \mid \theta=1/3)P(\theta=1/3 \mid X_1=1X_2=1) \\
&+ P(Y=0 \mid \theta=1/2)P(\theta=1/2 \mid X_1=1X_2=1) \\
&= 2/3 * 4/25 + 1/2 * 21/25 = 0,55128 \\
&= 1 - P(Y=1 \mid X_1=1X_2=1) \qquad (2)
\end{aligned}
$$

Ou seja, se tivermos que apostar, temos uma chance um pouco maior de que o resultado de um experimento futuro seja zero.