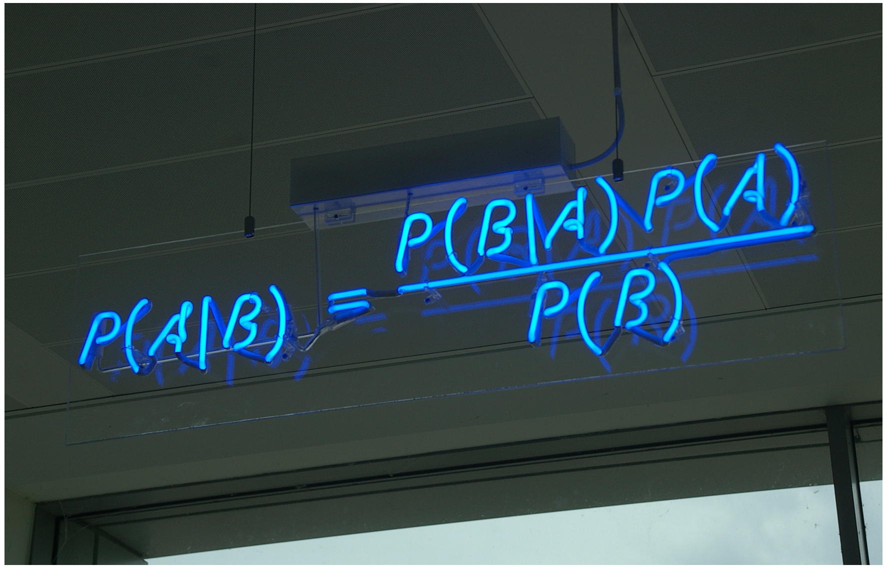
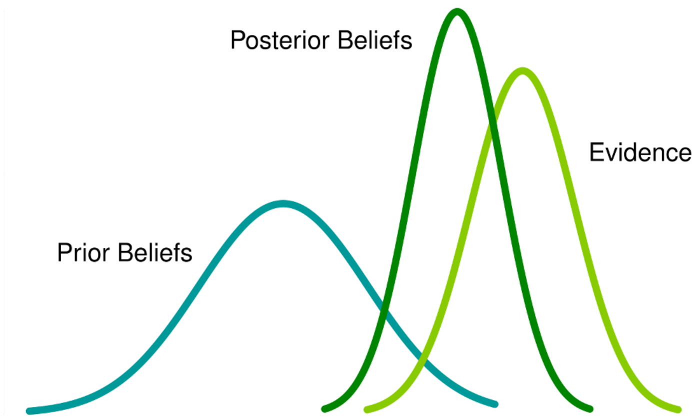
Estatística bayesiana

- Os paradigmas bayesiano e frequentista;
- Distribuição *a priori* e *a posteriori*;
- Função de verossimilhança;





Definição 9

***Permutabilidade**. Uma coleção finita de variáveis aleatórias $X_1, X_2, \ldots, X_n$ com densidade conjunta f é dita **permutável** se*

$$f(x_1, x_2, \ldots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}),$$

para qualquer permutação $\boldsymbol{\pi} = \{\pi(1), \pi(2), \ldots, \pi(n)\}$ dos seus elementos. Uma coleção infinita é permutável se qualquer subconjunto finito é permutável.

- Note que uma amostra permutável não precisa ser independente;
- Note também que IID $\implies$ permutável;
- A intuição é simples: simetria.

Parâmetros como limites de variáveis aleatórias.

Exemplo 1

Ensaio Clínico (De Groot, exemplo 7.1.3). *Suponha que estamos interessados na taxa de recrudescência ("recaída") de uma determinada doença entre pacientes tratados com uma droga. Seja X_i a variável aleatória que indica se o i-ésimo paciente recrudesceu ($X_i = 1$) ou não ($X_i = 0$). Seja P a proporção de indivíduos que recrudescem num grupo grande de pacientes. Se P é desconhecida, podemos modelar $X_1, X_2, \ldots$ como variáveis aleatórias Bernoulli IID com parâmetro p* ***condicional*** *a $P = p$. Em notação estatística:*

$$X_1, X_2, \ldots \mid P = p \sim \text{Bernoulli}(p).$$

Assuma *que $X_1, X_2, \ldots$ é uma sequência permutável infinita. Agora chamemos de P_n a proporção de pacientes que recrudescem nos n primeiros pacientes. Podemos mostrar que o limite $\lim_{n\to\infty} P_n = \lim_{n\to\infty} \sum_{i=1}^{n} X_i/n$ existe com probabilidade 1 e que pode ser visto como a proporção P.*

No Exemplo 1 podemos encarar o problema de duas maneiras:

A) P é uma variável aleatória e $X_1, X_2, \ldots$ são Bernoulli(p) **condicional** ao evento $P = p$, $p \in (0, 1)$.

B) Para uma constante fixa (e inobservável) p, $X_1, X_2, \ldots$ tem distribuição Bernoulli com parâmetro p – isto é, indexada por $p \in (0, 1)$.

Uma diferença *sutil*, não é? A tradição estatística que entende parâmetros como variáveis aleatórias como em A) é chamada de **Estatística bayesiana**[7]. Já os que aderem à abordagem B) são chamados **frequentistas** – ou ortodoxos, como Jaynes gosta de chamá-los. Neste curso veremos conceitos e exemplos destas duas escolas de pensamento.

[7] Em homenagem ao reverendo inglês Thomas Bayes (1701 – 1761).

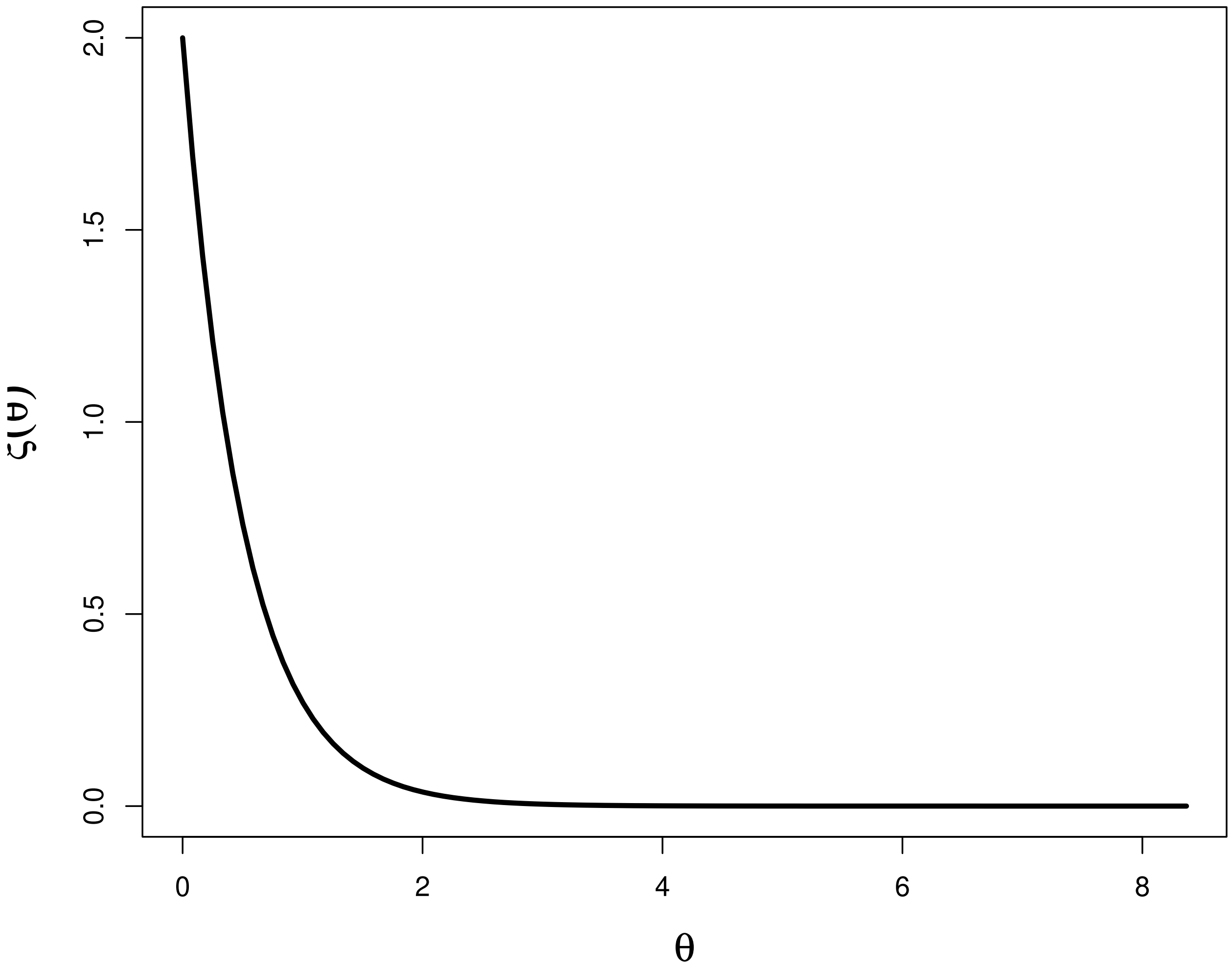
Exemplo 2

Duração de componentes eletrônicos (De Groot, exemplo 7.2.1). *Suponha que uma empresa esteja interessadas em saber o quanto duram os produtos que ela produz. Se representamos os tempos de duração de n objetos como n variáveis aleatórias* $X_1, X_2, \ldots, X_n$ *IID com distribuição exponencial com parâmetro* θ *de modo que*

$$f(x_i \mid \theta) = \theta \exp(-\theta x_i), x_i > 0.$$

Observação: $n / \sum_{i=1}^{n} \xrightarrow{p} \theta.$
Aqui, θ *é a taxa de falha dos componentes, e é um parâmetro de interesse. Suponha que uma pessoa experiente na empresa diga que a taxa de falha é mais ou menos* 0.5*/ano. Como representamos esta informação?*

Uma análise bayesiana (cont.)



Definição 10

Distribuição a priori. *Se tratamos o parâmetro θ como uma variável aleatória, então a distribuição a priori, que também chamaremos simplesmente de priori, é a distribuição que damos a θ* ***antes*** *de observarmos as outras variáveis aleatórias de interesse. Em geral, vamos denotar a função de densidade/massa de probabilidade da priori por $\xi(\theta)$.*

Exemplos:

- Podemos dizer que a probabilidade de uma moeda cair cara, p, tem distribuição uniforme entre 0 e 1;
  - Ou que tem distribuição Beta$(2, 2)$;
- A altura média dos jogadores de basquete do CR Flamengo tem distribuição normal com média $\mu_0 = 200cm$ e variância $\sigma_0^2 = 25cm^2$;
- A posição de Júpter em relação ao Sol hoje tem coordenadas X, Y, Z, de modo que $X \sim$ Normal$(\mu_x, 1)$, $Y \sim$ Normal$(\mu_y, 1)$, $Z \sim$ Normal$(\mu_z, 1)$.

Definição 11

*Considere o problema estatístico com parâmetro θ e variáveis aleatórias observáveis $X_1, X_2, \ldots, X_n$. A distribuição condicional de θ dados os valores observados das variáveis aleatórias, $\boldsymbol{x} := \{x_1, x_2, \ldots, x_n\}$ é a **distribuição a posteriori** de θ. Denotamos por $\xi(\theta \mid \boldsymbol{x})$ a f.d.p/f.m.p. condicional a $X_1 = x_1, X_2 = x_2, \ldots X_n = x_n$.*

Teorema 5

Considere a amostra aleatória $X_1, X_2, \ldots, X_n$ de uma distribuição com f.d.p./f.m.p. $f(x \mid \theta)$. Se a distribuição a priori é $\xi(\theta)$, temos

$$\xi(\theta \mid \boldsymbol{x}) = \frac{\xi(\theta) \prod_{i=1}^{n} f(x_i \mid \theta)}{g_n(\boldsymbol{x})}, \ \theta \in \Omega. \tag{4}$$

Chamamos $g_n(\boldsymbol{x})$ de distribuição marginal de $X_1, X_2, \ldots, X_n$.

Prova: Usar a premissa de amostra aleatória para escrever $f(x_1, x_2, \ldots, x_n \mid \theta)$, escrever a distribuição conjunta de θ e $\boldsymbol{x}$ e computar $g_n(\boldsymbol{x})$ usando a lei da probabilidade total.

Distribuição *a posteriori*: exemplo

Continuando com o Exemplo 2, fica claro que

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

onde $S = \sum_{i=1}^{n} x_n$. Desta forma, temos

$$f(\boldsymbol{x} \mid \theta)\xi(\theta) = \theta^{n+1} \exp(-(S+2)\theta).$$

Para obter $g_n(\boldsymbol{x})$, computamos

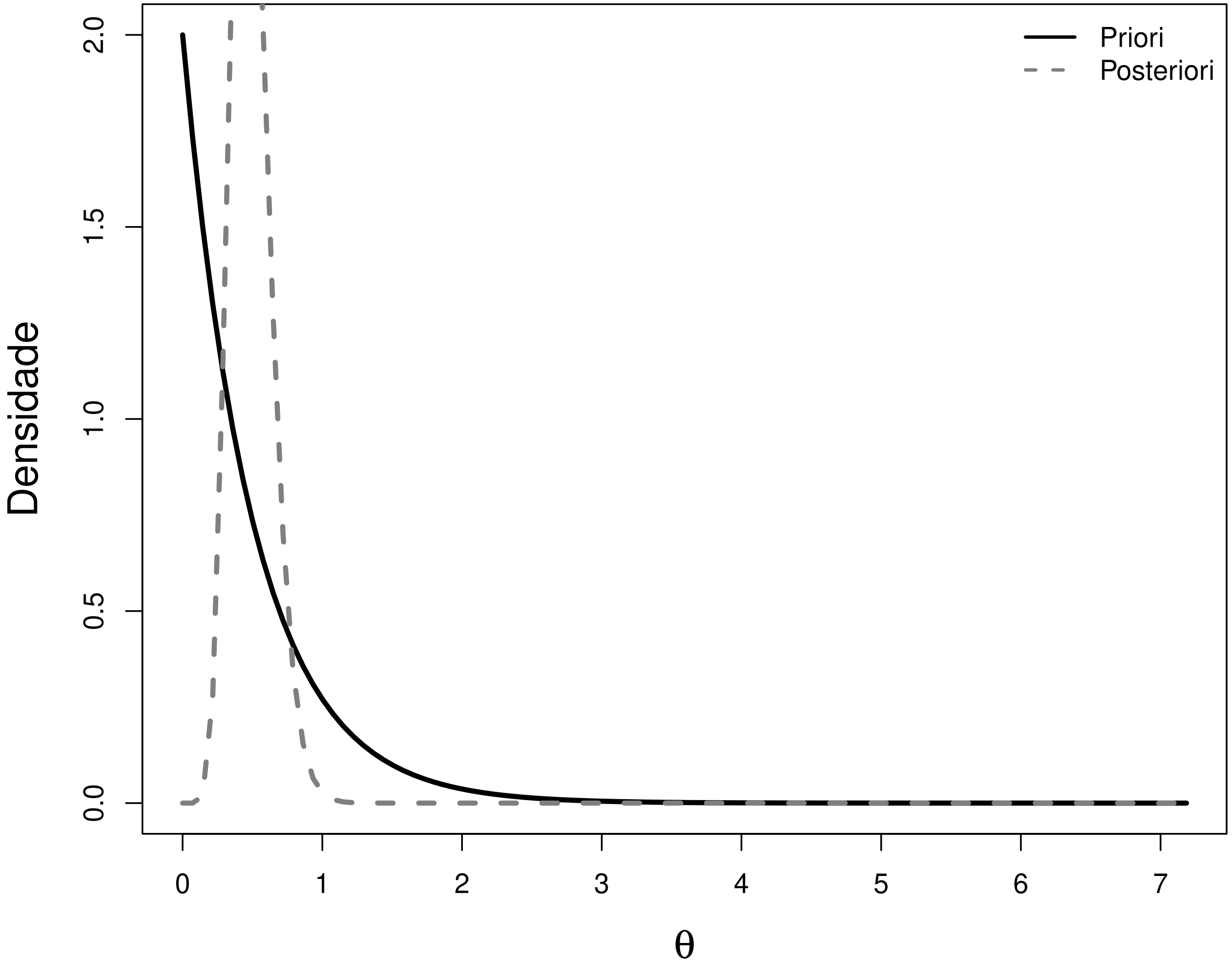
$$g_n(\boldsymbol{x}) = \int_0^\infty t^{n+1} \exp(-(S+2)t)\, dt = \frac{\Gamma(n+2)}{(S+2)^{n+2}}.$$

Concluímos que

$$\xi(\theta \mid \boldsymbol{x}) = \frac{(S+2)^{n+2}}{\Gamma(n+2)} \theta^{n+1} \exp(-(S+2)\theta),$$

ou seja, $\theta \mid \boldsymbol{x} \sim \text{Gama}(n+2, \sum_{i=1}^{n} x_n + 2)$.

Distribuição *a posteriori*: exemplo (cont.)



A função de verossimilhança

Note que o denominador em (4) não depende do parâmetro, θ. Deste modo, podemos escrever

$$\xi(\theta \mid \boldsymbol{x}) \propto f(\boldsymbol{x} \mid \theta)\xi(\theta),$$

querendo dizer que os dois lados de $\propto$ são iguais a não ser talvez por uma constante que independe de θ. Por vezes podemos escrever também $\xi(\theta \mid \boldsymbol{x}) \propto_\theta f(\boldsymbol{x} \mid \theta)\xi(\theta)$.

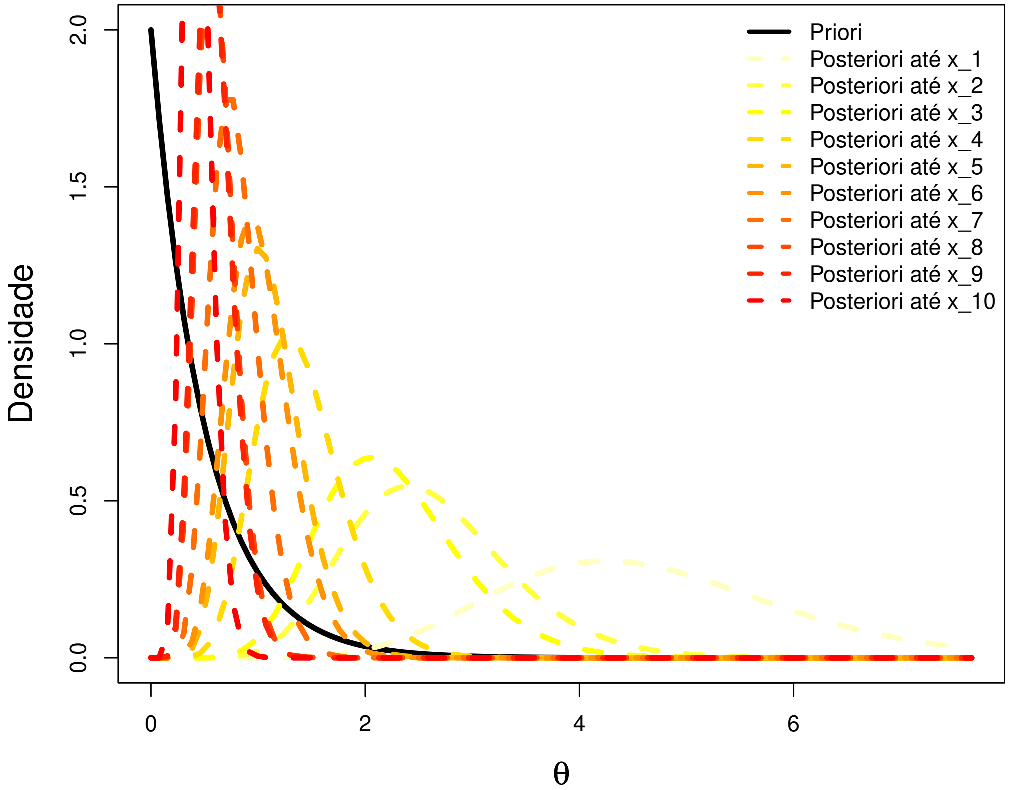
Definição 12

Função de verossimilhança. *Quando encaramos a f.d.p./f.m.p.* $f(x_1, x_2, \ldots, x_n \mid \theta)$ *como uma função do parâmetro* θ, *chamamos esta função de* ***função de verossimilhança***, *e podemos denotá-la como* $L(\theta; \boldsymbol{x})$ *ou, quando a notação não criar ambiguidade, simplesmente* $L(\theta)$.

Aprendizado bayesiano sequencial

Ainda sobre o Exemplo 2, considere a primeira observação x_1 e a distribuição *a posteriori* baseada apenas nesta observação: $\xi_1(\theta \mid x_1) \propto f(x_1 \mid \theta)\xi(\theta)$. Se assumirmos que $X_1, X_2, \ldots, X_n$ são condicionalmente independentes dado θ, podemos escrever

$$\xi(\theta \mid x_1, x_2) \propto f(x_1, x_2 \mid \theta)\xi(\theta) = f(x_1 \mid \theta)f(x_2 \mid \theta)\xi(\theta) = f(x_2 \mid \theta)\xi_1(\theta \mid x_1).$$



Predição

Dentro do paradigma bayesiano, a predição de novos valores da(s) variável(is) aleatória(s) é feita a partir da distribuição *a posteriori*,

$$p(x_{n+1} \mid x_1, x_2, \ldots, x_n) = \int_\Omega f(x_{n+1} \mid \theta)\xi(\theta \mid x_1, x_2, \ldots, x_n)\, d\theta. \tag{5}$$

Chamamos a distribuição condicional em (5) de **distribuição preditiva *a posteriori***. Em contraste, temos a **distribuição preditiva *a priori***:

$$p(x_{n+1}) = \int_\Omega f(x_{n+1} \mid \theta)\xi(\theta)\, d\theta, \tag{6}$$

que é útil na aplicação de modelos bayesianos na prática, mas não será explorada aqui.

O que aprendemos?

- Bayesianismo X frequentismo;
  "Parâmetros como variáveis aleatórias ou constantes fixas e não-observáveis."
- Distribuição *a priori*, $\xi(\theta)$;
  "Nosso grau de crença antes de observamos dados."
- Função de verossimilhança, $L(\theta) \propto f(\boldsymbol{x} \mid \theta)$;
  "Codifica (toda) a informação sobre o modelo contida nos dados."
- Distribuição *a posteriori*, $\xi(\theta \mid \boldsymbol{x}) \propto L(\theta)\xi(\theta)$;
  "Nossa crença atualizada a partir da informação contida em $L(\theta)$."

Leitura recomendada

- De Groot seção 7.2;
- * Capítulo 1 de Schervish, M. J. (2012). Theory of statistics. Springer Science & Business Media.