Inferência Estatística com Abordagem Bayesiana

Rosangela Helena Loschi [1]

[1]Departamento de Estatística
Universidade Federal de Minas Gerais

25 de outubro de 2021

MÉTODOS DE CONSTRUÇÃO DA DISTRIBUIÇÃO *A PRIORI*

- **Distribuições conjugadas naturais**
- Distribuições de referências: Métodos Bayes-Laplace e Jeffreys

Distribuições Conjugadas

Quando decidimos fazer uma **análise conjugada** dos dados estamos buscando por

(1) Simplicidade na derivação da distribuição *a posteriori*.
(2) Parâmetros interpretáveis *a posteriori*

sem perdermos a versatilidade em acomodar opiniões *a priori* de diferentes pessoas que, eventualmente, têm informações iniciais muito distintas.

Conjugação é uma propriedade de famílias de distribuições.

Se dois individuos decidem fazer uma análise conjugada do mesmo problema,

- suas distribuições *a priori* pertencerão à mesma família;
- no entanto, estas distribuições podem ser diferentes, dependendo do conhecimento inicial que têm.

Distribuições Conjugadas

Definição de Conjugação natural: Seja $\mathcal{F}$ uma família de distribuições amostrais indexadas pelo parâmetro θ, isto é,

$$\mathcal{F} = \{f(\mathrm{x} \mid \theta) : \theta \in \Theta, \mathrm{x} \in \mathcal{X}\}.$$

Seja $\mathcal{P}$ uma família de distribuções de probabilidade definida no espaço paramétrico Θ,

$$\mathcal{P} = \{\pi(\theta \mid a) : a \in \mathcal{A}\},$$

onde $\mathcal{A}$ é um conjunto de hiperparâmetros. Dizemos que as famílias $\mathcal{F}$ e $\mathcal{P}$ são **famílias conjugadas naturais** se

(*i*) $\mathcal{P}$ é fechada sob a amostragem de $\mathcal{F}$, ou seja, se $f(\mathrm{x} \mid \theta) \in \mathcal{F}$ é proporcional a um membro de $\mathcal{P}$, para cada $\mathrm{x} \in \mathcal{X}$.

(*ii*) $\mathcal{P}$ é fechada com respeito ao produto, ou seja, para todo $a_0, a_1 \in \mathcal{A}$, existe a_2 tal que

$$\pi(\theta \mid a_0)\pi(\theta \mid a_1) = \pi(\theta \mid a_2) \in \mathcal{P}.$$

Distribuições Conjugadas

Como consequência, se estamos fazendo uma análise conjugada

- a verosimilhança, se a vemos como uma função de θ, é proporcional à distribuição *a priori*
- as distribuições *a priori* e *a posteriori* pertencem à mesma família de distribuições.

A subjetividade na especificação da distribuição *a priori* ainda existe. Perceba que para um mesmo problema

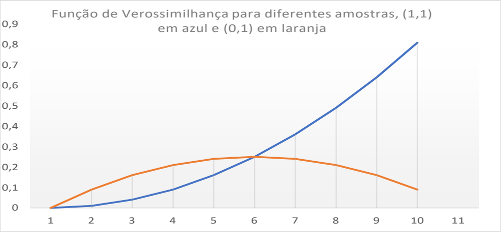
- A família de distribuição *a priori* é a mesma para todos aqueles que fazem a análise conjugada.
- A subjetividade está na especificação dos hiperparâmetros *a*, que dependerá do conhecimento prévio cada um tem sobre o parâmetro.

Distribuições Conjugadas: Caso Beta-Bernoulli

Se $X_1, \ldots, X_n$, dado θ, são i.i.d com distribuição Bernoulli(θ) e se, *a priori*, **$\theta \sim \mathrm{Beta}(\alpha, \beta)$, então** *a posteriori*,
$\theta \mid \boldsymbol{x} \sim \mathrm{Beta}(\alpha + \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x_i}, \beta + \mathrm{n} - \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x_i})$.

Solução: Iniciemos encontrando a função de verossimilhança. Como, dado θ, $X_1, \ldots, X_n$ são i.i.d temos que, para todo $x_i \in \{0, 1\}, i = 1, \ldots, n$,

$$
\begin{aligned}
f(x_1, \ldots, x_n \mid \theta) &= \prod_{i=1}^{n} f(x_i \mid \theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} \\
&= \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}
\end{aligned}
$$



Distribuições Conjugadas: Caso Beta-Bernoulli

A distribuição preditiva *a priori*

$$
\begin{aligned}
f(\boldsymbol{x}) &= f(x_1, \ldots, x_n) = \int_0^1 f(\boldsymbol{x} \mid \theta)\pi(\theta)d\theta \\
&= \int_0^1 \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \theta^{\alpha+\sum_{i=1}^n x_i-1}(1-\theta)^{\beta+n-\sum_{i=1}^n x_i-1}d\theta \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+\sum_{i=1}^n x_i)\Gamma(\beta+n-\sum_{i=1}^n x_i)}{\Gamma(\alpha+\beta+n)} \qquad (1)
\end{aligned}
$$

para todo $x_i \in \{0, 1\}, i = 1, \ldots, n$.

Distribuições Conjugadas: Caso Beta-Bernoulli

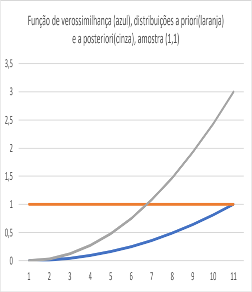
A distribuição *a posteriori* para $\theta \in (0,1)$

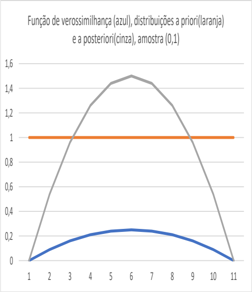
$$
\begin{aligned}
f(\theta \mid x_1, \ldots, x_n) &= \frac{f(\boldsymbol{x} \mid \theta)\pi(\theta)}{f(\boldsymbol{x})} \\
&= \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha+\sum_{i=1}^n x_i - 1}(1-\theta)^{\beta+n-\sum_{i=1}^n x_i - 1} d\theta}{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+\sum_{i=1}^n x_i)\Gamma(\beta+n-\sum_{i=1}^n x_i)}{\Gamma(\alpha+\beta+n)}} \\
&= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+\sum_{i=1}^n x_i)\Gamma(\beta+n-\sum_{i=1}^n x_i)} \\
&\times \quad \theta^{\alpha+\sum_{i=1}^n x_i - 1}(1-\theta)^{\beta+n-\sum_{i=1}^n x_i - 1}
\end{aligned}
$$

Logo, $\theta \mid \boldsymbol{x} \sim \mathrm{Beta}(\alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i)$.

Distribuições Conjugadas: Caso Beta-Bernoulli

Assumindo que $\theta \sim Unif(0,1) \equiv Beta(1,1)$ e observando as amostras $(0,1)$ e $(1,1)$ temos





Distribuições Conjugadas: Caso Beta-Bernoulli

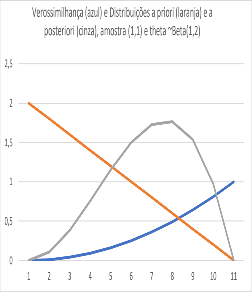
Neste caso,

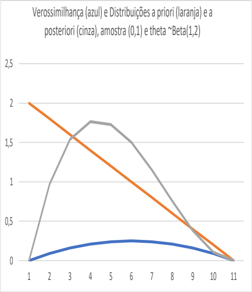
- a distribuição *a priori* é não informativa: não discrimina nenhum valor que θ.
- a verossimilhança discrimina valores de θ próximos de 1 se a amostra observada é $(1,1)$ e discrimina valores próximos de $0,5$ se a amostra observada é $(0,1)$.
- a informação dada pela distribuição *a posteriori* concorda com a informação dada pela verossimilhança: Os valores de θ mais prováveis *a posteriori* são valores acima de 0.7 se a amostra observada é $(1,1)$ e estao entre 0.3 e 0.7 se a amostra observada é $(0,1)$
- *A priori*, $E(\theta) = 1/2$ e $Var(\theta) = 1/12$. Se observamos a amostra $(0,1)$ então $\theta \mid x_1 = 0, x_2 = 1 \sim Beta(2,2)$. Assim, temos

$$E(\theta \mid x_1 = 0, x_2 = 1) = 1/2$$
$$Var(\theta \mid x_1 = 0, x_2 = 1) = 1/20$$

Distribuições Conjugadas: Caso Beta-Bernoulli

Assumindo que $\theta \sim Beta(1,2)$ e observando as amostras $(0,1)$ e $(1,1)$ temos





Distribuições Conjugadas: Caso Beta-Bernoulli

Neste caso,

- a distribuição *a priori* coloca mais massa probabilística para valores de θ próximos de zero.
- a verossimilhança discrimina valores de θ próximos de 1 se a amostra observada é $(1, 1)$ e discrimina valores proximos de $0, 5$ se a amostra observada é $(0, 1)$.
- a distribuição *a priori* e a função de verossimilhança fornecem informaçãos conflitantes sobre θ, principalmente, se a amostra observada é $(1, 1)$.
- Neste caso, quando misturamos as informações *a priori* e amostral sobre θ, obtemos uma informação *a posteriori* sobre θ bem modificada.
- a distribuição *a posteriori* sobre θ tende a estar mais próxima da informação trazida pelos dados pois a distribuição *a priori* é fraca (tem uma variância grande)

Distribuições Conjugadas: Caso Beta-Bernoulli

- Neste exemplo, se pode perceber o processo de atualização da informação sobre θ mais claramente.
- Como $\theta \mid \boldsymbol{x} \sim \mathrm{Beta}(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i)$, a média *a posteriori* é

$$
\begin{aligned}
E(\theta \mid \mathrm{x}) &= \frac{\alpha + \sum_{i=1}^{n} x_i}{(\alpha + \sum_{i=1}^{n} x_i) + (\beta + n - \sum_{i=1}^{n} x_i)} \\
&= \frac{\alpha + \sum_{i=1}^{n} x_i}{\alpha + \beta + n} \\
&= \frac{\alpha + \beta}{\alpha + \beta + n} E(\theta) + \frac{n}{\alpha + \beta + n} \bar{X}
\end{aligned}
$$

- $E(\theta \mid \mathrm{x})$ é uma combinação linear convexa da média *a priori* e do estimador de máxima verosimilhança para θ.

Distribuições Conjugadas: Caso Beta-Bernoulli

Perceba que

- Se $n \to \infty \Rightarrow E(\theta \mid \mathrm{x}) \to \bar{X}$: informação amostral é forte e domina a informação *a priori*.
- Se $\alpha \to 0$ e $\beta \to 0 \Rightarrow E(\theta \mid \mathrm{x}) \to \bar{X}$: informação *a priori* fraca e "qualquer"que seja a informação amostral ela dominará a distribuição *a priori*.

Denote por $\omega = \frac{n}{\alpha+\beta+n}$ o peso que se atribui à informação dos dados.

Distribuições Conjugadas: Caso Beta-Bernoulli

Efeito do tamanho amostral Suponha que o verdadeiro valor de θ é 0.6 e que *a priori* assimimos que $\theta \sim \mathrm{Beta}(5, 45)$. Note que nossa inferência *a priori* está muito longe do verdadeiro θ pois $E(\theta) = 0.1$. Suponha também que tenhamos gerado amostras de diferentes tamanhos ($n = 5, 100, 1000, 10000$) de uma Bernoulli(0.6) e que observamos 60% de sucessos em todas as amostras geradas

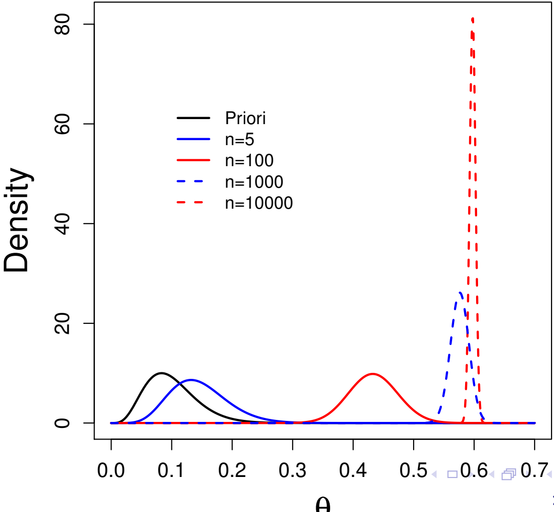
Table: Médias *a posteriori*

n	Distr. Post.	$E(\theta)$	EMV	ω	$E(\theta \mid x)$	$V(\theta \mid x)$
5	$\mathrm{Beta}(8, 47)$	0.1	0.6	0.0909	0.1454	2.22×10^{-3}
100	$\mathrm{Beta}(65, 85)$			0.6667	0.4334	1.63×10^{-3}
1000	$\mathrm{Beta}(605, 445)$			0.9524	0.5762	2.32×10^{-4}
10000	$\mathrm{Beta}(6005, 4045)$			0.9950	0.5975	2.39×10^{-5}

Quando maior o tamanho da amostra, menor é a influência da distribuição *a priori* na inferência *a posteriori*.

Distribuições Conjugadas: Caso Beta-Bernoulli

Figure: Distribuciones *a posteriori*



Distribuições Conjugadas: Caso Beta-Bernoulli

Efeito da distribuição *a priori* **1:** Assuma que o verdadeiro valor de θ é 0.6 e que eliciamos distintas distribuições *a priori* para θ. Suponha que geramos uma amostra de tamanho $n = 5$ de uma Bernoulli(0.6) e que observamos 60% de sucessos na amostra gerada.

Table: Médias *a posteriori*

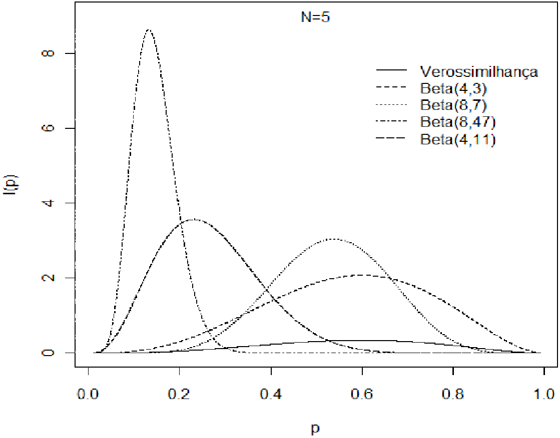
Dist. Priori	Dist. Post.	$E(\theta)$	$V(\theta)$	EMV	$1-\omega$	$E(\theta\|x)$	$V(\theta\|\mathbf{x})$
$Beta(1,1) - vaga$	$Beta(4,3)$	0.5	$0,0833$	0.6	0.2857	0.5714	$0,0306$
$Beta(5,5)$	$Beta(8,7)$	0.5	$0,0227$		0.6667	0.5333	$0,0156$
$Beta(5,45)$	$Beta(8,47)$	0.1	$0,0015$		0.9091	0.1454	$0,0022$
$Beta(1,9) - vaga$	$Beta(4,11)$	0.1	$0,0082$		0.6667	0.2666	$0,0122$

Neste caso,

- a informação dos dados estimam θ perfeitamente.
- Todas as distribuições *a priori* subestimam $\theta = 0.6$.
- melhor inferência *a posteriori* é obtida se *a priori* $\theta \sim Unif(0,1)$.
- Quando as informações amostral e *a priori* são **discordantes** ocorre um aumento na variância *a posteriori* se comparado com a variância *a priori*. Se são informações concordantes, a variância *a posteriori* diminui.

Distribuições Conjugadas: Caso Beta-Bernoulli

Figure: Distribuciones *a posteriori*



Sempre devemos utilizar distribuições a priori vagas ou não informativas?

Distribuições Conjugadas: Caso Beta-Bernoulli

Efeito da distribuição *a priori* **2:** Se o verdadeiro valor de θ é 0.5 e se eliciamos distintas distribuições *a priori* para θ. Suponha que geramos uma amostra de tamanho $n = 10$ de uma Bernoulli(0.5) e que observamos 60% de sucessos na amostra gerada.

Table: Médias *a posteriori*

Distr. Priori	Distr. Post.	$E(\theta)$	EMV	$1-\omega$	$E(\theta \mid x)$
$\mathrm{Beta}(1,1)$-vaga	$\mathrm{Beta}(7,5)$	0.5	0.6	0.1667	0.5833
$\mathrm{Beta}(50,50)$	$\mathrm{Beta}(56,54)$			0.9091	0.5090
$\mathrm{Beta}(1,9)$-vaga	$\mathrm{Beta}(6.1,6.4)$	0.1		0.0909	0.5545
$\mathrm{Beta}(5,45)$	$\mathrm{Beta}(11,49)$			0.8334	0.1833

Note que

- a informação dos dados levam a superestimação de θ
- as distribuições *a priori* $Beta(1,1)$ (vaga) e $Beta(50,50)$ (muito informativa) levam a estimativas (Médias *a priori*) perfeitas de θ.
- Sob tais distribuições *a priori*, a distribuição mais informativa leva a melhor estimativa *a posteriori*
- No caso de distribuições *a priori* que subestimam θ, a melhor estimativa *a posteriori* é obtida sob a distribuição mais vaga.

Distribuições Conjugadas: Caso Beta-Bernoulli

DISCUSSÕES IMPORTANTES: Obviamente,

- a distribuição *a priori* influencia a inferência *a posteriori*.
- Assim, ela deve traduzir adequadamente o seu conhecimento inicial sobre θ.
- as distribuições *a priori* pouco informativas nem sempre conduzem às melhores inferências.
- a distribuição *a priori* é dominada pelos dados se
  + a amostra é grande.
  + a distribuição *a priori* é pouco informativa.

DISCUSSÕES IMPORTANTES:

- Na construção da verossimilhança utilizamoos o conceito de **independência condicional**, isto é, assumimos que X_1 é inde pendente de X_2, dado θ, isto é,

$$f(X_1, X_2 \mid \theta) = f(X_1 \mid \theta) f(X_2 \mid \theta)$$

. **Podemos dizer que X_1 é independente de X_2?**

- Dado θ, X_1 e X_2 são independentes $\Rightarrow Cov(X_1, X_2 \mid \theta) = 0$. Além disto, $E(X_i \mid \theta) = \theta$, $i = 1, 2$. Daí, usando propriedades de covariância condicional, temos

$$\begin{aligned} Cov(X_1, X_2) &= E(Cov(X_1, X_2 \mid \theta)) + COV(E(X_1 \mid \theta), E(X_2 \mid \theta)) \\ &= E(0) + Cov(\theta, \theta) = Var(\theta). \end{aligned} \tag{2}$$

- X_1 e X_2 são correlacionadas $\Rightarrow X_1$ e X_2 não são independentes.

Distribuições Conjugadas: Caso Beta-Bernoulli

Moda da distribuição $Beta(\alpha, \beta)$: Consideremos o logaritmo da densidade de θ

$$\ln \pi(\theta) = \ln \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} + (\alpha - 1)\ln\theta + (\beta - 1)\ln(1-\theta)$$

Se θ é um ponto no interior do intervalo $(0, 1)$, derivando com respeito a θ e igualando a zero, temos

$$\begin{aligned} \frac{d \ln \pi(\theta)}{d\theta} &= \frac{(\alpha-1)}{\theta} + \frac{(\beta-1)}{1-\theta} = 0 \\ \Rightarrow \quad (\alpha - 1) &= \theta(\alpha+\beta-2) \end{aligned}$$

Se $\alpha + \beta \neq 2 \;\; \Rightarrow \hat{\theta} = \frac{(\alpha-1)}{(\alpha+\beta-2)} \leftarrow$ candidato a Máximo.
Considerando a segunda derivada com respeito a θ, temos

$$\frac{d^2 \ln \pi(\theta)}{d\theta^2} = -\frac{(\alpha-1)}{\theta^2} - \frac{(\beta-1)}{(1-\theta)^2} \tag{3}$$

Distribuições Conjugadas: Caso Beta-Bernoulli

Avaliando a segunda derivada em $\hat{\theta}$ temos

$$\frac{d^2 \ln \pi(\theta)}{d\theta^2} = -\left[\frac{(\alpha+\beta-2)^2}{\alpha-1} + \frac{(\alpha+\beta-2)^2}{\beta-1}\right]$$

- se $\alpha \leq 1$ e $\beta \leq 1 \Rightarrow \frac{d^2 \ln \pi(\theta)}{d\theta^2} > 0 \Rightarrow$ não há ponto de máximo (moda)
- se $\alpha > 1$ e $\beta > 1 \Rightarrow \frac{d^2 \ln \pi(\theta)}{d\theta^2} < 0 \Rightarrow$ há moda e ela é

$$\hat{\theta} = \frac{(\alpha-1)}{(\alpha+\beta-2)}$$

- Para os demais valores de α e β caso tem-se que fazer a análise. Pode ou não haver moda.
- A moda pode estar nos extremos. Por exemplo, se $\alpha = 2$ e $\beta = 1$ então $\pi(\theta) = 2\theta \leftarrow$ função estritamente crescente com moda (máximo em) $\hat{\theta} = 1$.

Dist. Conjugadas: Caso Normal com variância conhecida

Exemplo: Se $X_1, \ldots, X_n$, dado μ, são variáveis aleatórias i.i.d com distribuição Normal(μ, σ^2), em que a variância σ^2 é conhecida. Construa a famíla conjugada para o parâmetro μ. Temos que

- o espaço paramétrico é $\mathbb{R}$.
- a função de verossimilhança é

$$
\begin{aligned}
f(\boldsymbol{x} \mid \mu) &= (2\pi\sigma^2)^{-n/2} \exp\left\{ \frac{-1}{2\sigma^2} \left(\sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i + n\mu^2 \right) \right\} \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{ \frac{-1}{2\sigma^2} \sum_{i=1}^{n} x_i^2 \right\} \\
&\times \exp\left\{ \frac{-1}{2\sigma^2} \left(-2\mu \sum_{i=1}^{n} x_i + n\mu^2 \right) \right\} \\
&\propto \exp\left\{ -\frac{n}{2\sigma^2} \left(\mu^2 - 2\mu\bar{x} \right) \right\} \\
&\propto \exp\left\{ -\frac{n}{2\sigma^2} (\mu - \bar{x})^2 \right\} \qquad (4)
\end{aligned}
$$

Dist. Conjugadas: Caso Normal com variância conhecida

O fechamento por amostragem:

- Olhando para a verossimilhança como uma função de μ nos sugere que o núcleo da família de distribuições sobre μ que é fechada sob a amostragem da família normal com σ^2 conhecido deve ser

$$\pi(\mu) \propto \exp\left\{-\frac{1}{2V}(\mu - M)^2\right\}$$

- Por construção, temos que a família $\mu \sim \mathrm{Normal}(M, V)$ de distribuições sobre μ é fechada sob a amostragem da família amostral normal com variância conhecida.
- Esta família é fechada por produto?

Dist. Conjugadas: Caso Normal com variância conhecida

O fechamento por produto:

Assuma que, *a priori* $\mu \sim \mathrm{Normal}(M, V)$. Assim,

$$\pi(\mu) \propto \exp\left\{-\frac{1}{2V}(\mu - M)^2\right\} = \exp\left\{-\frac{1}{2V}(\mu^2 - 2\mu M)\right\}\exp\left\{-\frac{1}{2V}M^2\right\}$$

Misturando esta informação *a priori* com a verossimilhança em (4) temos que o núcleo da distribuição *a posteriori* para μ é

$$\begin{aligned}
\pi(\mu \mid \mathrm{x}) &\propto \exp\left\{-\frac{1}{2V}(\mu^2 - 2\mu M)\right\}\exp\left\{-\frac{n}{2\sigma^2}(\mu^2 - 2\mu\bar{x})\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\mu^2\left(\frac{1}{V} + \frac{n}{\sigma^2}\right) - 2\mu\left(\frac{M}{V} + \frac{n\bar{x}}{\sigma^2}\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\mu^2\left(\frac{\sigma^2 + nV}{V\sigma^2}\right) - 2\mu\left(\frac{M\sigma^2 + nV\bar{x}}{V\sigma^2}\right)\right]\right\} \\
&\propto \exp\left\{-\left(\frac{\sigma^2 + nV}{2V\sigma^2}\right)\left[\mu^2 - 2\mu\left(\frac{M\sigma^2 + nV\bar{x}}{\sigma^2 + nV}\right)\right]\right\}
\end{aligned} \tag{5}$$

Dist. Conjugadas: Caso Normal com variância conhecida

Ou seja,

- a distribuição *a posteriori*

$$\mu \mid \mathrm{x} \sim \text{Normal}\left(\frac{M\sigma^2 + nV\bar{x}}{\sigma^2 + nV}; \frac{V\sigma^2}{\sigma^2 + nV}\right) \tag{6}$$

- A família de distribuições para μ dada em (5) é fechada por produto.
- Temos **conjugação natural.**

Dist. Conjugadas: Caso Normal com variância conhecida

IMPORTANTE:

- A média *a posteriori* é uma combinação linear convexa da média *a priori* com o EMV de μ

$$E(\mu \mid \boldsymbol{x}) = M\frac{\sigma^2}{\sigma^2 + nV} + \bar{x}\frac{nV}{\sigma^2 + nV} \tag{7}$$

- Perceba que se $n \to \infty$
  - $E(\mu \mid \mathrm{x}) \to \bar{X}$: informação amostral é forte e domina a informação *a priori*.
  - $Var(\mu \mid \mathrm{x}) \to 0$
  - Ou seja, a distribuição *a posteriori* de μ converge para uma distribuição degenerada tal que $P(\mu = \bar{x} \mid \mathrm{x}) = 1$
- Perceba que Se $V \to \infty$
  - $E(\mu \mid \mathrm{x}) \to \bar{X}$: a informação *a priori* fraca e "qualquer"que seja a informação amostral ela dominará a distribuição *a priori*. temos uma distribuição *a priori* vaga.
  - $Var(\mu \mid \mathrm{x}) \to \sigma^2/n$
  - Ou seja, a distribuição *a posteriori* de μ converge para uma distribuição Normal$(\bar{x}, \sigma^2/n)$.

Dist. Conjugadas: Caso Normal com variância conhecida

IMPORTANTE:

- Perceba que se $V = 0 \Rightarrow \mu \sim \mathrm{Normal}(M, 0)$, que é uma distribuição degenerada em M.
- Voce, *a priori*, está completamente convencido que o valor real de μ é M.
  - *A posteriori*, tem-se $E(\mu \mid \mathrm{x}) = M$ e $Var(\mu \mid \mathrm{x}) = 0$
  - A distribuição *a posteriori* de μ é $\mu \mid \boldsymbol{x} \sim \mathrm{Normal}(M, 0)$
  - os dados não modificam o seu conhecimento inicial

Lei de Cromwell: Atribua probabilidade zero apenas para eventos que você sabe que não podem numca ocorrer pois os dados, por mais forte que seja a informação que trazem, não são capazes de modificar/atualizar uma probabilidade nula.

Dist. Conjugadas: Caso Normal com variância conhecida

IMPORTANTE: Se X_1 e X_2, dado μ, são variáveis aleatórias i.i.d com distribuição Normal(μ, σ^2) e se, *a priori*, você estabeleceu que $\mu \sim \mathrm{Normal}(M, V)$, podemos dizer que X_1 é independente de X_2?

Como X_1 e X_2, dado μ, são variáveis aleatórias independentes temos que $Cov(X_1, X_2 \mid \mu) = 0$.

$$\begin{aligned} Cov(X_1, X_2) &= E(Cov(X_1, X_2 \mid \mu)) + COV(E(X_1 \mid \mu), E(X_2 \mid \mu)) \\ &= E(0) + Cov(\mu, \mu) = Var(\mu) = V. \end{aligned} \tag{8}$$

Note que a variância *a priori* para μ mensura a correlação entre X_1 e X_2 *a priori*.

Dist. Conjugadas: Família Exponencial($\mathcal{FE}$)

Definição de Família Exponencial ($\mathcal{FE}$): Dizemos que uma distribuição de probabilidade $f(x \mid \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$ e $x \in \mathcal{X}$, pertence à **Família Exponencial** se assume a seguinte forma:

$$f(x \mid \theta) = c(\theta)h(x)\exp\{R(\theta)^t T(x)\}, \qquad (9)$$

onde $c(.)$ e $h(.)$ são funções não-negativas e $R(.)$ e $T(.)$ são funções vetoriais k-dimensionais.

Exemplo 1: Se $X|\lambda \sim Poisson(\lambda)$, temos que

$$f(x \mid \lambda) = \frac{e^{-\lambda}\lambda^x}{x!} = (x!)^{-1}e^{-\lambda}\exp\{x ln(\lambda)\}. \qquad (10)$$

- Pertence à família exponencial basta fazer$h(x) = (x!)^{-1}$, $c(\theta) = e^{-\lambda}$, $T(x) = x$ e $R(\theta) = ln(\lambda)$.
- $T(x) = x$ é a **Estatítica Suficiente** para λ.

Dist. Conjugadas: Família Exponencial($\mathcal{FE}$)

Exemplo 2: Se $X|\theta \sim \textit{Uniforme}(a,b)$, temos que

$$f(x \mid \lambda) = \frac{1}{b-a}1\{x \in (a,b)\} = \frac{1}{b-a}\exp\{ln(1\{x \in (a,b)\})\} \quad (11)$$

- Não pertence à família exponencial.

Exemplo 3: Se $X|\theta \sim \textit{Normal}(\mu, \sigma^2)$, onde μ e σ^2 são desconhecidos. Neste caso, temos que

$$\begin{aligned} f(x \mid \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma}}\exp\left\{\frac{-1}{2\sigma^2}(x^2 - 2x\mu + \mu^2)\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma}}\exp\left\{\frac{-\mu^2}{2\sigma^2}\right\}\exp\left\{\left(\frac{-1}{2\sigma^2}, \frac{\mu}{\sigma^2}\right).(x^2, 2x)\right\} \end{aligned}$$

- Pertence à família exponencial. Basta considerar $h(x) = 1$, $c(\theta) = \frac{1}{\sqrt{2\pi\sigma}}\exp\left\{\frac{-\mu^2}{2\sigma^2}\right\}$, $T(x) = (x^2, 2x)$ e $R(\theta) = \left(\frac{-1}{2\sigma^2}, \frac{\mu}{\sigma^2}\right)$.
- $T(x) = (x^2, 2x)$ é a **Estatítica Suficiente** para (σ^2, μ).

Dist. Conjugadas: Família Exponencial($\mathcal{FE}$)

Conjugação na Família Exponencial ($\mathcal{FE}$): Seja

$$\mathcal{F} = \{f(\mathrm{x} \mid \theta) : \theta \in \Theta \subseteq \mathbb{R}^k, \mathrm{x} \in \mathcal{X}\},$$

a **Família Exponencial** de distribuições amostrais indexadas pelo parâmetro θ tal que

$$f(\mathrm{x} \mid \theta) = c(\theta)h(\mathrm{x}) \exp\{R(\theta)^t T(\mathrm{x})\}, \tag{12}$$

onde $c(.)$ e $h(.)$ são funções não-negativas e $R(.)$ e $T(.)$ são funções vetoriais k-dimensionais. Então, a família de distribuições de probabilidade definida no espaço paramétrico Θ, $\mathcal{P} = \{\pi(\theta \mid a) : a \in \mathcal{A}\}$ tal que

$$\pi(\theta \mid a) = \frac{c(\theta)^\alpha \exp\{R(\theta)^t \beta\}}{\int_\Theta c(\theta)^\alpha \exp\{R(\theta)^t \beta\} d\theta} \tag{13}$$

é conjugada natural da familia $\mathcal{F}$.

Distribuições Conjugadas: Família Exponencial

Se $X_1, \ldots, X_n$, dado θ, são condicionalmente independentes e identicamente com distribuição na família exponencial em (12) e, *a priori*, $\pi(\theta) \in \mathcal{P}$ dada em (13) , então a distribuição *a posteriori* é

$$\pi(\theta \mid x_1, \ldots, x_n, a) = \frac{c(\theta)^{\alpha+n} \exp\{R(\theta)^t(\boldsymbol{\beta} + T(x_1, \ldots, x_n))\}}{\int_\Theta c(\theta)^{\alpha+n} \exp\{R(\theta)^t(\boldsymbol{\beta} + T(x_1, \ldots, x_n))\} d\theta}$$

A distribuição preditiva *a priori* é

$$f(x_1, \ldots, x_n) = \left[\prod_{i=1}^{n} h(x_i)\right] \frac{\int_\Theta c(\theta)^{\alpha+n} \exp\{R(\theta)^t(\boldsymbol{\beta} + T(x_1, \ldots, x_n))\} d\theta}{\int_\Theta c(\theta)^{\alpha} \exp\{R(\theta)^t \boldsymbol{\beta}\} d\theta}$$

Distribuições Conjugadas: Família Exponencial

Exemplo (Caso Poisson-Gamma): Se $X_1, \ldots, X_n$, dado θ, são i.i.d com distribuição Poisson(θ), então a verosimilhança

$$f(\mathrm{x} \mid \theta) = \left[\prod_{i=1}^{n} \frac{1}{x_i!}\right] \exp\left\{\log(\theta) \sum_{i=1}^{n} x_i\right\} \exp\{-n\theta\}$$

pertence à Família Exponencial em que

$$c(\theta)^n = [\exp\{-\theta\}]^n, \qquad h(\boldsymbol{x}) = \left[\prod_{i=1}^{n} \frac{1}{x_i!}\right]$$

$$T(\boldsymbol{x}) = \sum_{i=1}^{n} x_i, \qquad R(\theta) = \log(\theta)$$

Distribuições Conjugadas: Família Exponencial

Como consequência do teorema anterior,

- a família de distribuições de probabilidade definida em Θ, que é conjugada natural família amostral Poisson, tem forma

$$\begin{aligned} \pi(\theta) &\propto \exp\{\log(\theta)\beta\}\exp\{\alpha\theta\} \\ &\propto \theta^{\beta}\exp\{-\alpha\theta\}, \end{aligned} \tag{14}$$

  que é o núcleo de uma distribução $\mathrm{Gama}(\beta-1,\alpha)$ com média $(\beta-1)/\alpha$.

- a distribuição *a posteriori* é

$$\begin{aligned} \pi(\theta \mid \boldsymbol{x}) &\propto \exp\left\{\log(\theta)(\beta+\sum_{i=1}^{n} x_i)\right\}\exp\{(\alpha+n)\theta\} \\ &\propto \theta^{\beta+\sum_{i=1}^{n} x_i}\exp\{-(\alpha+n)\theta\}, \end{aligned} \tag{15}$$

- $\Rightarrow$ *A posteriori*, $\theta \mid \boldsymbol{x} \sim Gama(\beta+\sum_{i=1}^{n} x_i - 1, \alpha+n)$

Distribuições Conjugadas: Família Exponencial

No caso Poisson-Gama, também se observa que a média *a posteriori* é uma combinação linear convexa do EMV para θ e a média *a priori*

$$\begin{aligned} E(\theta \mid \mathrm{x}) &= \frac{\beta + \sum_{i=1}^{n} x_i - 1}{\alpha + n} \\ &= \frac{\beta - 1}{\alpha + n} + \frac{\sum_{i=1}^{n} x_i}{\alpha + n} \\ &= \frac{\alpha}{\alpha + n} E(\theta) + \frac{n}{\alpha + n} \bar{X} \end{aligned} \tag{16}$$

Perceba que

- Se $n \to \infty \Rightarrow E(\theta \mid \mathrm{x}) \to \bar{X}$: a informação amostral sobre θ é forte e domina a informação *a priori*.
- Se $\alpha \to 0$ e $\beta \to 1 \Rightarrow E(\theta \mid \mathrm{x}) \to \bar{X}$: a informação *a priori* é fraca e é dominada pela informação amostral qualquer que seja n.

Exemplo (Modelo Linear): Denote o vector de covariáveis para o k-ésimo elemento da amostra por $\mathrm{X}_k \in \mathbb{R}^l$ dado por $(1, X_{k,1}, \ldots, X_{k,l-1})$, $k = 1, \ldots, n$, e por $\boldsymbol{\beta} \in \mathbb{R}^l$ o vetor fixos (β_0, $\ldots$, $\beta_{l-1})^t$. Assuma que (X_k, Y_k) segue a estrutura de um modelo linear

$$Y_k = \mathrm{X}_k\boldsymbol{\beta} + e_k \quad \mathrm{e} \quad e_k \overset{\mathrm{iid}}{\sim} \mathcal{N}(0, \sigma^2) \tag{17}$$

Assim, as variáveis $Y_1, \ldots, Y_n$ são independentes e

$$Y_k|\mathrm{X}_k, \boldsymbol{\beta}, \sigma^2 \overset{\mathrm{ind}}{\sim} \mathcal{N}(\mathrm{X}_k\boldsymbol{\beta}, \sigma^2). \tag{18}$$

Na análise conjugada do modelo de regressão assume-se,*a priori* que

$$\boldsymbol{\beta}|\sigma^2 \sim \mathcal{N}_l(\mathrm{m}, \sigma^2\mathrm{V}), \tag{19}$$
$$\sigma^2 \sim \mathcal{IG}(\nu/2, d/2), \tag{20}$$

onde ν e d são valores reais positivos, m é um vetor real $l \times 1$ e V é uma matrix $l \times l$, positiva definida.

Distribuições Conjugadas: Família Exponencial

Exemplo (Modelo Linear, cont.): Consequentemente, *a posteriori* a inferência é

$$\begin{aligned} \boldsymbol{\beta} \mid \sigma^2, \mathrm{y}, \mathrm{x} &\sim \mathcal{N}_l(\mathrm{m}^*, \sigma^2 \mathrm{V}^*), && (21) \\ \sigma^2 \mid \sigma^2, \mathrm{y}, \mathrm{x} &\sim \mathcal{IG}(\nu^*/2, d^*/2), && (22) \end{aligned}$$

onde

$$\begin{aligned} \mathrm{m}^* &= \mathrm{V}^* \left(\mathrm{V}^{-1}\mathrm{m} + \mathrm{x}^t\mathrm{x} \right), && (23) \\ \mathrm{V}^* &= \left(\mathrm{V}^{-1} + \mathrm{x}^t\mathrm{x} \right)^{-1}, && (24) \\ \nu^* &= \nu + (\mathrm{m})^t\mathrm{V}^{-1}\mathrm{m} + \mathrm{y}^t\mathrm{y} - (\mathrm{m}^*)^t(\mathrm{V}^*)^{-1}\mathrm{m}^*, && (25) \\ d^* &= d + j - i. && (26) \end{aligned}$$

Conjugação fora da $\mathcal{FE}$

Se a família de distribuições amostrais $\mathcal{F}$ é a $\mathrm{Uniforme}(0, \theta)$, então a família Pareto de distribuições de probabilidade para θ é conjugada natural da família $\mathcal{F}$.

Prova: Seja $X_1, \ldots, X_n$ uma amostra que, dado, θ é i.i.d. tal que $X_i \mid \theta \overset{iid}{\sim} \mathrm{Uniforme}(0, \theta)$. Então a função de verossimilhança é

$$
\begin{aligned}
f(\mathbf{x} \mid \theta) &= \prod_{i=1}^{n} \frac{1}{\theta} 1(x_i \in (0, \theta)) \\
&= (\theta)^{-n} 1(max\{x_1, \ldots, x_n\} < \theta)
\end{aligned}
$$

Construindo uma família fechada por amostragem com respeito a $\mathcal{F}$: Para que isto ocorra a família de distribuições $\mathcal{P}$ sobre θ tem que ter núcleo proporcional ao núcleo da verossimilhança, isto é,

$$\pi(\theta) \propto Kernell(f(\mathbf{x} \mid \theta)) = \left(\frac{1}{\theta}\right)^{??} 1(\theta > ??)$$

Conjugação fora da $\mathcal{FE}$

Incluindo parâmetros pra torná-la flexivel: A familia $\mathcal{P}$ sobre θ candidata a família conjugada natural de $\mathcal{F}$ tem a seguinte fdp

$$\pi(\theta) = \beta\alpha^{\beta}\,(\theta)^{-(\beta+1)}\,1(\theta > \alpha),$$

ou seja, $\theta \sim Pareto(\alpha, \beta)$.

Provando o fechamento por produto: Tomemos a distribuição *a priori* para θ pertencente à família Parto, i.é, $\theta \sim Pareto(\alpha, \beta)$. Daí, a distribuição *a posteriori* é

$$\begin{aligned}\pi(\theta \mid \boldsymbol{x}) &= \frac{(\theta)^{-n}\,1(max\{x_1,\ldots,x_n\} < \theta) * \beta\alpha^{\beta}\,(\theta)^{-(\beta-1)}\,1(\theta > \alpha)}{\int (\theta)^{-n}\,1(max\{x_1,\ldots,x_n\} < \theta) * \beta\alpha^{\beta}\,(\theta)^{-(\beta-1)}\,1(\theta > \alpha)d\theta} \\ &= \frac{\theta^{-(n+\beta+1)}1(\theta > max[max\{x_1,\ldots,x_n\},\alpha])}{\int_{\theta > max[max\{x_1,\ldots,x_n\},\alpha]}\theta^{-(n+\beta+1)}d\theta}\end{aligned}$$

Conjugação fora da $\mathcal{FE}$

Note que, no denominador, estamos integrando o núcleo de uma distribuição $Pareto(\alpha^*, \beta + n)$ onde $\alpha^* = max[max\{x_1, \ldots, x_n\}, \alpha]$.

A integral do denominator é $[(\alpha^*)^{\beta+n}(\beta + n)]^{-1}$.

A distribuição *a posteriori* para θ é

$$\theta \sim Pareto(\alpha^*, \beta + n)$$

onde $\alpha^* = max[max\{x_1, \ldots, x_n\}, \alpha]$.

Como a família $\mathcal{P}$ é fechada sob a amostragem de $\mathcal{F}$ e $\mathcal{P}$ é fechada por produto, provamos que as famílias $Uniforme(0, \theta)$ e $Pareto(\alpha, \beta)$ são conjugadas naturais.

Conjugação na família de mistura finita de distribuições

Seja $\mathcal{F}$ uma família de distribuições amostrais indexadas pelo parâmetro θ, isto é,

$$\mathcal{F} = \{f(\mathrm{x} \mid \theta) : \theta \in \Theta, \mathrm{x} \in \mathcal{X}\}.$$

Se a distribuição *a priori* para θ é uma mistura finita de distribuições, ou seja, se $\pi(\theta) = \sum_{i=1}^{k} \omega_i \pi_i(\theta)$, $\omega_i \in (0, 1)$, $\sum_{i=1}^{k} \omega_i = 1$, e $\pi_i(\theta)$ é uma função densidade ou função de probabilidade, então

- a distribuição preditiva *a priori* $f(\boldsymbol{x})$ é também uma mistura finita de distribuições dada por

$$f(\boldsymbol{x}) = \sum_{i=1}^{k} \omega_i f_i(\boldsymbol{x}),$$

$f_i(\mathrm{x}) = \int_{\Theta} f(\mathrm{x} \mid \theta)\pi_i(\theta) d\theta$ é a distribuição preditiva *a priori* com respeito ao componente $\pi_i(\theta)$ da mistura.

Conjugação na família de mistura finita de distribuições

Prova: Da definição de distribuição preditiva *a priori* segue que

$$
\begin{aligned}
f(\boldsymbol{x}) &= \int_{\Theta} f(\boldsymbol{x} \mid \theta)\pi(\theta)d\theta \\
&= \int_{\Theta} f(\boldsymbol{x} \mid \theta)\sum_{i=1}^{k}\omega_i\pi_i(\theta)d\theta \\
&= \sum_{i=1}^{k}\omega_i\int_{\Theta} f(\boldsymbol{x} \mid \theta)\pi_i(\theta)d\theta \\
&= \sum_{i=1}^{k}\omega_i f_i(\boldsymbol{x})
\end{aligned}
\tag{27}
$$

Conjugação na família de mistura finita de distribuições

- a distribuição *a posteriori* é uma mistura finita tal que

$$\pi(\theta \mid \mathrm{x}) = \sum_{i=1}^{k} \omega_i^* \pi_i(\theta \mid \mathrm{x})$$

onde $\omega_i^* = \frac{\omega_i f_i(\mathrm{x})}{\sum_{i=1}^{k} \omega_i f_i(\mathrm{x})}$ e $f_i(\mathrm{x}) = \int_\Theta f(\mathrm{x} \mid \theta)\pi_i(\theta) d\theta$ é a distribuição preditiva *a priori* com respeito ao componente $\pi_i(\theta)$ da mistura.

Conjugação na família de mistura finita de distribuições

Prova: usando o teorema de Bayes temos que

$$
\begin{aligned}
\pi(\theta \mid \boldsymbol{x}) &= \frac{f(\boldsymbol{x} \mid \theta)\pi(\theta)}{\int_\Theta f(\boldsymbol{x} \mid \theta)\pi(\theta)d\theta} \\
&= \frac{f(\boldsymbol{x} \mid \theta)\sum_{i=1}^k \omega_i \pi_i(\theta)}{\int_\Theta f(\boldsymbol{x} \mid \theta)\sum_{i=1}^k \omega_i \pi_i(\theta)d\theta} \\
&= \frac{\sum_{i=1}^k \omega_i f(\boldsymbol{x} \mid \theta)\pi_i(\theta)}{\sum_{i=1}^k \omega_i f_i(\boldsymbol{x})} \\
&= \sum_{i=1}^k \frac{\omega_i}{\sum_{i=1}^k \omega_i f_i(\boldsymbol{x})} f(\boldsymbol{x} \mid \theta)\pi_i(\theta) \\
&= \sum_{i=1}^k \frac{\omega_i \int_\Theta f(\mathrm{x} \mid \theta)\pi_i(\theta)d\theta}{\sum_{i=1}^k \omega_i f_i(\boldsymbol{x})} \frac{f(\boldsymbol{x} \mid \theta)\pi_i(\theta)}{\int_\Theta f(\mathrm{x} \mid \theta)\pi_i(\theta)d\theta} \\
&= \sum_{i=1}^k \frac{\omega_i f_i(\mathrm{x})}{\sum_{i=1}^k \omega_i f_i(\boldsymbol{x})} \pi_i(\theta \mid \boldsymbol{x})
\end{aligned}
$$

Conjugação na família de mistura finita de distribuições

Exemplo: Se $X_1, \dots, X_n$, dado θ, são i.i.d com distribuição Poisson(θ). Assuma que a distribuição *a priori* para θ é a seguinte mistura finita de distribuições gama:

$$\pi(\theta) = pF_{G_1}(\alpha_1, \beta_1) + (1-p)F_{G_2}(\alpha_2, \beta_2),$$

onde $F_{G_i}(\alpha_i, \beta_i)$ representa a densidade de uma distribuição $Gama(\alpha_i, \beta_i)$, $p \in (0,1)$, $\alpha_i > 0$ e $\beta_i > 0$, $i = 1, 2$. Encontre (a) a distribuição preditiva *a priori* e (b) a distribuição *a posteriori* de θ.

Solução: Considerando o resultado anterior precisamos encontra as distribuições *a posteriori* e preditiva *a priori* para cada componente da mistura.

a) Dist. preditiva *a priori* :

Dist. preditiva *a priori* para o componente j:

$$
\begin{aligned}
f_j(\boldsymbol{x}) &= \int_0^\infty (\prod_{i=1}^n 1/x_i!) \exp\{-n\theta\}\theta^{\sum_{i=1}^n x_i} \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)}\theta^{\alpha_j-1}\exp\{-\beta_j\theta\}d\theta \\
&= \left(\prod_{i=1}^n 1/x_i!\right) \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} \int_0^\infty \theta^{\sum_{i=1}^n x_i+\alpha_j-1}\exp\{-(\beta_j+n)\theta\}d\theta \\
&= \left(\prod_{i=1}^n 1/x_i!\right) \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} \frac{\Gamma(\sum_{i=1}^n x_i+\alpha_j)}{(\beta_j+n)^{\sum_{i=1}^n x_i+\alpha_j}}, \; x_i = 0, 1, \ldots, i = 1, \ldots, n.
\end{aligned}
$$

Dist. preditiva *a priori*

$$
f(\boldsymbol{x}) = \left(\prod_{i=1}^n 1/x_i!\right) \left\{ p\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \frac{\Gamma(\sum_{i=1}^n x_i+\alpha_1)}{(\beta_i+n)^{\sum_{i=1}^n x_i+\alpha_1}} + (1-p)\frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \frac{\Gamma(\sum_{i=1}^n x_i+\alpha_2)}{(\beta_2+n)^{\sum_{i=1}^n x_i+\alpha_2}} \right\}
$$

para todo $x_i = 0, 1, \ldots, i = 1, \ldots, n$.

b) Dist. *a posteriori* **:**

Distribuição *a posteriori* para cada componente da mistura: Como as distribuições Gama e Poisson são conjugadas naturais temos que para cada componente da mistura

$$\pi_j(\theta|\boldsymbol{x}) \propto \theta^{\sum_{i=1}^{n} x_i+\alpha_j-1} \exp\{-(\beta_j+n)\theta\}.$$

Assim, *a posteriori*, cada componente j da mistura é uma $Gama(\sum_{i=1}^{n} x_i+\alpha_j, \beta_j+n)$

A distribuição *a posteriori* para θ é uma mistura finita das distribuições $Gama(\sum_{i=1}^{n} x_i+\alpha_1, \beta_1+n)$ e $Gama(\sum_{i=1}^{n} x_i+\alpha_2, \beta_2+n)$ onde o peso da primeira componente da mistura é

$$p^* = \frac{p\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\frac{\Gamma(\sum_{i=1}^{n} x_i+\alpha_1)}{(\beta_i+n)^{\sum_{i=1}^{n} x_i+\alpha_1}}}{\left\{p\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\frac{\Gamma(\sum_{i=1}^{n} x_i+\alpha_1)}{(\beta_i+n)^{\sum_{i=1}^{n} x_i+\alpha_1}}+(1-p)\frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)}\frac{\Gamma(\sum_{i=1}^{n} x_i+\alpha_2)}{(\beta_2+n)^{\sum_{i=1}^{n} x_i+\alpha_2}}\right\}}$$