Famílias conjugadas e alguns modelos paramétricos

Prof. Vinícius D. Mayrink
EST088 - Inferência Bayesiana

Sala: 4073 Email: vdm@est.ufmg.br

1^o semestre de 2025

3.1 - Famílias conjugadas

O tema a ser tratado nesta parte do curso é muito importante para nos orientar sobre como escolher distribuições *a priori* de maneira apropriada para garantir facilidade nas contas do Teorema de Bayes que irão determinar a distribuição *a posteriori*.

Em muitos problemas práticos iremos optar por **distribuições a priori conjugadas**. Para explicar este assunto, iniciamos com a definição a seguir.

Definição: Seja $\mathcal{F} = \{f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \ : \ \theta \in \Theta\}$ uma família de distribuições relacionadas à amostra $\boldsymbol{Y}$. Uma classe $\mathcal{C}$ de distribuições é reconhecida como família conjugada em relação a $\mathcal{F}$, se para todo $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \in \mathcal{F}$ e $f_\theta(\theta) \in \mathcal{C}$, temos $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \in \mathcal{C}$.

Em outras palavras, se a distribuição *a priori* $f_\theta(\theta)$ escolhida é membro de uma família de distribuições que conjugam com o modelo estabelecido pela verossimilhança $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})$, então a distribuição *a posteriori* $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y})$ também será membro da mesma família na qual $f_\theta(\theta)$ é integrante.

Cuidado! A classe $\mathcal{C}$ pode ser bastante ampla.

Se $\mathcal{C} = \{\text{todas as distribuições}\}$ e $\mathcal{F}$ é qualquer família de distribuições amostrais. Note que $\mathcal{C}$ é conjugada com respeito a $\mathcal{F}$ visto que qualquer distribuição *a posteriori* será membro de $\mathcal{C}$.

A classe $\mathcal{C}$ poderia ser também muito restrita.

Suponha, por exemplo, que $\mathcal{C} = \{f_\theta(\theta) \ : \ f_\theta(\theta = \theta_0) = 1, \ \theta_0 \in \mathcal{A}\}$ para algum conjunto não nulo $\mathcal{A}$.

Isto significa que: $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \propto f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \ f_\theta(\theta) = \begin{cases} f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \times 1, & \text{se } \theta = \theta_0. \\ f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \times 0, & \text{se } \theta \neq \theta_0. \end{cases}$

Portanto segue que: $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) = \begin{cases} k \times f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}), & \text{se } \theta = \theta_0. \\ 0, \ \text{se } \theta \neq \theta_0. \end{cases}$

Conclusão: $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) = 1$, se e somente se $\theta = \theta_0$, portanto, $\mathcal{C}$ é conjugada a qualquer família de distribuições.

Veja que quando uma probabilidade nula é atribuída a um subconjunto dos possíveis valores de θ, nenhuma informação observada irá mudar a especificação (o que é claramente inadequado).

Regra de Cromwell (Lindley): Devemos sempre associar uma probabilidade *a priori* não nula para cada possível valor de θ, mesmo que alguns deles sejam julgados como muito improváveis.

Exemplo: Considere $Y_1, Y_2, \cdots, Y_n$ uma amostra aleatória tal que $Y_i|\theta \sim \text{Bernoulli}(\theta)$. Dado θ, adote independência condicional dos Y_i's.

Verossimilhança: $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) = \theta^{\sum_{i=1}^{n} y_i} (1-\theta)^{n-\sum_{i=1}^{n} y_i}$.

Espaço paramétrico: $\theta \in [0, 1]$.

A distribuição *a priori* escolhida pelo pesquisador é discreta e coloca massa de probabilidade em apenas 4 valores dentro do invervalo $[0, 1]$:

Denote: $\theta_1 = 0.1$, $\theta_2 = 0.2$, $\theta_3 = 0.4$ e $\theta_4 = 0.9$.

$f_\theta(\theta_1) = 0.1$, $f_\theta(\theta_2) = 0.2$, $f_\theta(\theta_3) = 0.5$ e $f_\theta(\theta_4) = 0.2$.

Regra de Bayes para obter a distribuição *a posteriori*:

$$f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \;=\; \frac{\theta^{\sum_{i=1}^n y_i}\;(1-\theta)^{n-\sum_{i=1}^n y_i}\;\;f_\theta(\theta)}{\sum_{j=1}^4 \theta_j^{\sum_{i=1}^n y_i}\;(1-\theta_j)^{n-\sum_{i=1}^n y_i}\;\;f_\theta(\theta_j)}$$

$$f_{\theta|\boldsymbol{Y}}(\theta_1|\boldsymbol{y}) \;=\; \frac{(0.1)^{\sum_{i=1}^n y_i}\;(1-0.1)^{n-\sum_{i=1}^n y_i}\;\;f_\theta(0.1)}{\sum_{j=1}^4 \theta_j^{\sum_{i=1}^n y_i}\;(1-\theta_j)^{n-\sum_{i=1}^n y_i}\;\;f_\theta(\theta_j)} \;\propto\; (0.1)^{\sum_{i=1}^n y_i}\;(0.9)^{n-\sum_{i=1}^n y_i}\;\;0.1.$$

O cálculo acima foi feito para um dos valores com probabilidade não nula na especificação *a priori*. A próxima avaliação considera o valor $\theta = 0.5$ para o qual foi atribuído probabilidade 0 *a priori*.

$$f_{\theta|\boldsymbol{Y}}(0.5|\boldsymbol{y}) \;=\; \frac{(0.5)^{\sum_{i=1}^n y_i}\;(1-0.5)^{n-\sum_{i=1}^n y_i}\;\;f_\theta(0.5)}{\sum_{j=1}^4 \theta_j^{\sum_{i=1}^n y_i}\;(1-\theta_j)^{n-\sum_{i=1}^n y_i}\;\;f_\theta(\theta_j)}$$

$$\propto\; (0.5)^{\sum_{i=1}^n y_i}\;(0.5)^{n-\sum_{i=1}^n y_i}\;\;0 = 0.$$

Conclusão: para qualquer tamanho amostral n, a probabilidade *a posteriori* de $\theta = 0.5$ será 0, pois foi admitida uma probabilidade nula para esse valor *a priori*.

A classe $\mathcal{C}$ precisa ser ampla o bastante para assegurar uma conveniente elicitação *a priori* e, ao mesmo tempo, restrita o bastante para que a definição de conjugação seja útil.

Definição: Seja X uma variável aleatória associada à seguinte classe de distribuições $\mathcal{C} = \{f_{X|\lambda}(x) \ : \ \lambda \in \Lambda\}$. A classe $\mathcal{C}$ é dita *fechada sob multiplicação ou amostragem* se para todo $f_{X|\lambda_1}(x)$ e $f_{X|\lambda_2}(x) \in \mathcal{C}$, existe um k tal que $k \times f_{X|\lambda_1}(x) \times f_{X|\lambda_2}(x) \in \mathcal{C}$.

Exemplo: Seja $\mathcal{C}$ a classe de distribuições Gama(a_i, b_i) com $a_i > 0$ e $b_i > 0$. Denote $\lambda_i = (a_i, b_i)$.

$$f_{X|\lambda_1}(x) \times f_{X|\lambda_2}(x) \ = \ \frac{b_1^{a_1}}{\Gamma(a_1)}\, x^{a_1-1} \exp\{-b_1 x\} \ \frac{b_2^{a_2}}{\Gamma(a_2)}\, x^{a_2-1} \exp\{-b_2 x\}$$

$$\propto x^{a_1+a_2-2} \exp\{-(b_1+b_2)x\}.$$

Temos o núcleo da Ga$(a_1 + a_2 - 1, \ \ b_1 + b_2)$.

O "fechamento" é muito importante na busca por famílias conjugadas.

Suponha que iremos escolher nossa especificação *a priori* sobre θ dentro da classe de distribuições $\mathcal{C} = \{f_{\theta|\lambda}(\theta) \ : \ \lambda \in \Lambda\}$. Admita também que esta classe é fechada sob multiplicação e amostragem.

Se o núcleo da verossimilhança $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})$ se assemelha ao núcleo da classe $\mathcal{C}$, que é fechada sob multiplicação, então as distribuições *a priori* e *a posteriori* irão necessariamente pertencer à mesma classe $\mathcal{C}$ (dizemos que existe **conjugação**).

Procedimento para determinar a classe conjugada:

1. Identificar a classe $\mathcal{C}$ de distribuições para θ cujos membros são proporcionais a $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})$.
2. Verificar se $\mathcal{C}$ é fechada sob multiplicação e amostragem.

Se, além disso, para qualquer verossimilhança $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})$ obtida de uma família $\mathcal{F}$, existe uma constante k definindo $f_\theta(\theta) = k\ f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})$, então a classe $\mathcal{C}$, de todas as p.d.f.'s ou p.m.f.'s deste tipo, é dita família conjugada natural com respeito ao modelo com verossimilhança $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})$.

Exemplo: Seja $(Y_i|\theta) \sim \text{Bernoulli}(\theta)$, $i = 1, 2, \ldots, n$. O teorema de Bayes fornece: $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \;\propto\; f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\; f_\theta(\theta) \;\propto\; \theta^t (1-\theta)^{n-t}\; f_\theta(\theta)$, sendo $t = \sum_{i=1}^n y_i$.

Desejamos determinar uma distribuição *a priori* conjugada para θ.

Considere $\quad k^{-1} \;=\; \dfrac{\Gamma(t+1)\Gamma(n-t+1)}{\Gamma(t+1+n-t+1)}$.

Veja que: $\quad k\; f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \;=\; \frac{\Gamma(t+1+n-t+1)}{\Gamma(t+1)\Gamma(n-t+1)}\; \theta^t (1-\theta)^{n-t}$.

Temos acima a p.d.f. da Beta$(t+1, n-t+1)$. Portanto, a classe de distribuições Beta com parâmetros inteiros (t é inteiro) é uma família conjugada natural para o modelo Bernoulli.

Se considerarmos todos os possíveis valores positivos para os parâmetros da Beta (não somente os valores inteiros), a classe de distribuições Beta será ampliada, e ainda teremos uma distribuição conjugada, mas ela não é a conjugada natural.

Famílias conjugadas naturais são úteis visto que um significado objetivo pode ser atribuído aos **hiperparâmetros** envolvidos. No último exemplo, suponha que exista um experimento prévio para o qual obtemos t_0 sucessos em n_0 realizações. Então a conjugada natural será a Beta$(t_0 + 1, n_0 - t_0 + 1)$.

Nomenclatura (esclarecendo): hiperparâmetro é um parâmetro que indexa a distribuição *a priori* especificada para um outro parâmetro que, por sua vez, indexa a distribuição de $\boldsymbol{Y}$.

No modelo Bayesiano $\quad Y_i \sim \text{Bernoulli}(\theta) \quad$ e $\quad \theta \sim \text{Beta}(a, b)$,
temos que a e b são hiperparâmetros.

Para fixar melhor os conceitos apresentados sobre conjugação, nos próximos slides iremos avaliar algumas famílias conjugadas importantes.

As principais famílias conjugadas.

Distribuição Binomial: A família de distribuições Beta é conjugada com o modelo Bernoulli ou Binomial.

O caso Bernoulli já foi trabalhado em vários exemplos discutidos neste curso (rever slides).

O caso Binomial será detalhado em duas situações a seguir.

Situação 1: Seja $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ uma amostra aleatória tal que $Y_i|\theta \sim \text{Bernoulli}(\theta)$ são i.i.d. por independência condicional. Admita que a informação amostral que é passada ao pesquisador é $W = \sum_{i=1}^{n} Y_i$ e n.

Note que W é o número de sucessos em n experimentos Bernoullis. Temos $W \sim \text{Binomial}(n, \theta)$, pois é a soma de ensaios Bernoullis independentes.

Espaço paramétrico: $\theta \in [0, 1]$.

Verossimilhança: $f_{W|\theta}(w) = \binom{n}{w} \theta^w (1-\theta)^{n-w}$. Sem produtório!

Distribuição *a priori*: $\theta \sim \text{Beta}(a, b)$.

Distribuição *a posteriori*: $f_{\theta|W}(\theta|w) \propto f_{W|\theta}(w)\, f_{\theta}(\theta)$

$$f_{\theta|W}(\theta|w) \propto \binom{n}{w}\theta^{w}(1-\theta)^{n-w}\, \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

Termos em destaque são constantes (não dependem de θ) e podem ser incorporados na constante normalizadora.

$$\propto \theta^{w}(1-\theta)^{n-w}\, \theta^{a-1}(1-\theta)^{b-1}$$

$$\propto \theta^{a+w-1}(1-\theta)^{b+n-w-1}.$$

A expressão acima é o núcleo da Beta(a^*, b^*)
com $a^* = a + w$ e $b^* = b + n - w$.

As distribuições *a priori* $\theta \sim \text{Beta}(a, b)$ e *a posteriori* $\theta|W \sim \text{Beta}(a^*, b^*)$ são da família Beta.

Situação 2: Seja $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ uma amostra aleatória tal que $Y_i|\theta \sim \text{Binomial}(m, \theta)$ são i.i.d. por independência condicional. Veja que Y_i é o número de sucessos em m ensaios Bernoullis.

Espaço paramétrico: $\theta \in [0, 1]$.

Verossimilhança: $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) = \prod_{i=1}^{n} \binom{m}{y_i} \theta^{y_i} (1-\theta)^{m-y_i}$. Com produtório!

Distribuição *a priori*: $\theta \sim \text{Beta}(a, b)$.

Distribuição *a posteriori*: $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \propto f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\, f_\theta(\theta)$

$$f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \propto \prod_{i=1}^{n} \binom{m}{y_i} \theta^{y_i} (1-\theta)^{m-y_i} \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}$$

$$\propto \left[\prod_{i=1}^{n} \binom{m}{y_i} \right] \theta^{\sum_{i=1}^{n} y_i} (1-\theta)^{\sum_{i=1}^{n} (m-y_i)} \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}.$$

Termos em destaque são constantes (não dependem de θ) e podem ser incorporados na constante normalizadora.

$$f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \;\propto\; \theta^{\sum_{i=1}^n y_i}(1-\theta)^{nm-\sum_{i=1}^n y_i}\;\theta^{a-1}(1-\theta)^{b-1}.$$

$$\propto\; \theta^{a+\sum_{i=1}^n y_i-1}(1-\theta)^{b+nm-\sum_{i=1}^n y_i-1}.$$

É possível identifica acima o núcleo da Beta(a^*,b^*)
com $a^* = a + \sum_{i=1}^n y_i$ e $b^* = b + nm - \sum_{i=1}^n y_i$.

Novamente, as distribuições *a priori* $\theta \sim$ Beta(a, b) e *a posteriori* $\theta|\boldsymbol{Y} \sim$ Beta(a^*, b^*) são da família Beta.

Distribuição Exponencial: A família de distribuições Gama é conjugada com o modelo Exponencial.

Este caso também já foi discutido neste curso (rever slides).

Seja $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ uma amostra aleatória da Exp(θ) com média $1/\theta > 0$. Assuma independência condicional dos Y_i's dado θ.

Espaço paramétrico: $\theta \in \mathbb{R}^+$.

Verossimilhança: $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \;=\; \prod_{i=1}^n \theta \exp\{-\theta y_i\} \;=\; \theta^n \exp\{-\theta \sum_{i=1}^n y_i\}$.

Distribuição *a priori*: $\theta \sim \text{Ga}(a,b)$.

Distribuição *a posteriori*: $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \;\propto\; f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\, f_\theta(\theta)$

$\propto\; \theta^n \exp\{-\theta \sum_{i=1}^n y_i\}\, \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp\{-b\theta\}$

Termos em destaque são constantes (não dependem de θ) e podem ser incorporados na constante normalizadora.

$\propto\; \theta^{a+n-1} \exp\{-(b + \sum_{i=1}^n y_i)\theta\}$.

Este é o núcleo da $\text{Ga}(a^*,b^*)$, com $a^* = a + n$ e $b^* = b + \sum_{i=1}^n y_i$.

Conclusão: as distribuições *a priori* $\theta \sim \text{Ga}(a,b)$ e *a posteriori* $\theta|\boldsymbol{Y} \sim \text{Ga}(a^*,b^*)$ são da família Gama.

$E(\theta|\boldsymbol{Y}) = a^*/b^* = \frac{a+n}{b+\sum_{i=1}^n y_i} = \frac{b}{b+\sum_{i=1}^n y_i}\ \frac{a}{b}\ +\ \frac{\sum_{i=1}^n y_i}{b+\sum_{i=1}^n y_i}\ \frac{n}{\sum_{i=1}^n y_i}.$

Veja que $n \to \infty$ implica em $E(\theta|\boldsymbol{Y}) \approx 1/\bar{y}$.

Perceba que a distribuição Exponencial é um caso particular da distribuição Gama.

Se $Y_i|\theta \sim \text{Ga}(1, \theta)$, então $Y_i|\theta \sim \text{Exp}(\theta)$.

f.d.p.'s: $\frac{\theta^1}{\Gamma(1)}\ y_i^{1-1}\ \exp\{-\theta y_i\}\ =\ \theta\ \exp\{-\theta y_i\}$ para $y_i > 0$ e $\theta > 0$.

Seja $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ uma amostra aleatória da $\text{Ga}(\kappa, \theta)$ com parâmetro de forma κ conhecido. Admita independência condicional dos Y_i's dado θ. A família de distribuições Gama é conjugada com a família $\text{Ga}(\kappa,\theta)$. Note que o parâmetro alvo a ser estimado é o θ.

Exercício: Mostre este resultado! Obtenha a distribuição *a posteriori*.

Distribuição Poisson: A família de distribuições Gama é conjugada com o modelo Poisson.

Reveja os slides! Este caso também já foi discutido anteriormente.

Considere $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ uma amostra aleatória da Poisson(θ). Adote independência condicional entre os Y_i's dado a taxa θ.

Espaço paramétrico: $\theta \in \mathbb{R}^+$.

Verossimilhança: $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) = \prod_{i=1}^n \frac{\theta^{y_i}}{y_i!} \exp\{-\theta\} = \frac{\theta^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!} \exp\{-n\theta\}$.

Distribuição *a priori*: $\theta \sim \text{Ga}(a, b)$.

Distribuição *a posteriori*: $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \propto f_{\boldsymbol{Y}|\theta}(\boldsymbol{y})\, f_\theta(\theta)$

$$\propto \frac{\theta^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!} \exp\{-n\theta\} \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp\{-b\theta\}$$

Termos em destaque são constantes (não dependem de θ) e podem ser incorporados na constante normalizadora.

$$f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \propto \theta^{\sum_{i=1}^n y_i} \exp\{-n\theta\}\, \theta^{a-1} \exp\{-b\theta\}.$$

$$\propto \theta^{a+\sum_{i=1}^n y_i - 1} \exp\{-(b+n)\theta\}.$$

Temos aqui o núcleo da $\text{Ga}(a^*,b^*)$, com $a^* = a + \sum_{i=1}^n y_i$ e $b^* = b + n$.

Conclusão: as distribuições *a priori* $\theta \sim \text{Ga}(a,b)$ e *a posteriori* $\theta|\boldsymbol{Y} \sim \text{Ga}(a^*, b^*)$ são da família Gama.

$$E(\theta|\boldsymbol{Y}) = a^*/b^* = \frac{a+\sum_{i=1}^n y_i}{b+n} = \frac{b}{b+n}\,\frac{a}{b} + \frac{n}{b+n}\,\bar{y}.$$

Veja que $n \to \infty$ implica em $E(\theta|\boldsymbol{Y}) \approx \bar{y}$.

Distribuição Geométrica: A família de distribuições Beta é conjugada com o modelo Geométrico. Seja $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ uma amostra tal que $Y_i|\theta \sim$ Geométrica(θ). Adote independência condicional dos Y_i's dado θ.

A Geométrica refere-se ao caso em que realizam-se sucessivos experimentos (sucesso e fracasso) até a ocorrência do 1^o sucesso. Y_i é o n^o de fracassos antes do evento sucesso. Esta é uma variável aleatória discreta com suporte $y_i \in \{0, 1, 2, 3, \cdots\}$.

Espaço paramétrico: $\theta \in [0, 1]$ é a probabilidade de sucesso.

Verossimilhança: $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) = \prod_{i=1}^{n} \theta(1-\theta)^{y_i} = \theta^n \, (1-\theta)^{\sum_{i=1}^{n} y_i}$.

Distribuição *a priori*: $\theta \sim$ Beta(a, b).

Distribuição *a posteriori*: $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \propto f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \, f_\theta(\theta)$

$$\propto \theta^n \, (1-\theta)^{\sum_{i=1}^{n} y_i} \, \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} \propto \theta^{a+n-1} \, (1-\theta)^{b+\sum_{i=1}^{n} y_i - 1}$$

Este é o núcleo da Beta(a^*, b^*), com $a^* = a + n$ e $b^* = b + \sum_{i=1}^{n} y_i$.

Conclusão: a distribuição *a priori* $\theta \sim$ Beta(a, b) e *a posteriori* $\theta|\boldsymbol{Y} \sim$ Beta(a^*, b^*) são da família Beta.

Distribuição $N(\mu, \sigma^2)$ com variância conhecida: A família de distribuições Normais é conjugada com o modelo Normal para estimar a média μ.

Seja $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ uma amostra aleatória da $N(\mu, \sigma^2)$. Admita independência condicional dos Y_i's dado μ.

Espaço paramétrico: $\quad \mu \in \mathbb{R}$.

Verossimilhança: $\quad f_{\boldsymbol{Y}|\mu}(\boldsymbol{y}) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\}$.

$= \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\{-\frac{1}{2\sigma^2}(y_i^2 - 2\mu y_i + \mu^2)\}$.

$= (2\pi\sigma^2)^{-n/2} \exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n} y_i^2 - 2\mu \sum_{i=1}^{n} y_i + n\mu^2)\}$.

$= (2\pi\sigma^2)^{-n/2} \exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n} y_i^2)\} \exp\{-\frac{1}{2\sigma^2}(-2\mu \sum_{i=1}^{n} y_i + n\mu^2)\}$.

Distribuição *a priori*: $\quad \mu \sim N(m, v)$.

Distribuição *a posteriori*: $f_{\mu|\boldsymbol{Y}}(\mu|\boldsymbol{y}) \propto f_{\boldsymbol{Y}|\mu}(\boldsymbol{y})\, f_\mu(\mu)$

$$\propto (2\pi\sigma^2)^{-n/2} \exp\{-\tfrac{1}{2\sigma^2}(\textstyle\sum_{i=1}^n y_i^2)\} \exp\{-\tfrac{1}{2\sigma^2}(-2\mu \textstyle\sum_{i=1}^n y_i + n\mu^2)\} \times$$

$$\times(2\pi v)^{-1/2} \exp\{-\tfrac{1}{2v}(\mu - m)^2\}.$$

$$\propto (2\pi\sigma^2)^{-n/2} \exp\{-\tfrac{1}{2\sigma^2}(\textstyle\sum_{i=1}^n y_i^2)\} \exp\{-\tfrac{1}{2\sigma^2}(-2\mu \textstyle\sum_{i=1}^n y_i + n\mu^2)\} \times$$

$$\times(2\pi v)^{-1/2} \exp\{-\tfrac{1}{2v} m^2\} \exp\{-\tfrac{1}{2v}(\mu^2 - 2\mu m)\}.$$

Termos destacados em vermelho são constantes (não dependem de μ). Eles serão incorporados na constante normalizadora.

$$f_{\mu|\boldsymbol{Y}}(\mu|\boldsymbol{y}) \propto \exp\{-\frac{1}{2}(-2\mu\frac{\sum_{i=1}^n y_i}{\sigma^2} + \frac{n}{\sigma^2}\mu^2)\} \exp\{-\frac{1}{2}(\frac{1}{v}\mu^2 - 2\mu\frac{m}{v})\}.$$

$$\propto \exp\{-\frac{1}{2}[\mu^2(\frac{n}{\sigma^2} + \frac{1}{v}) - 2\mu(\frac{\sum_{i=1}^n y_i}{\sigma^2} + \frac{m}{v})]\}.$$

$$\propto \exp\{-\frac{1}{2}\,(\frac{n}{\sigma^2} + \frac{1}{v})\,[\mu^2 - 2\mu\,(\frac{n}{\sigma^2} + \frac{1}{v})^{-1}\,(\frac{\sum_{i=1}^n y_i}{\sigma^2} + \frac{m}{v})]\}.$$

Temos aqui o núcleo da distribuição $N(m^*, v^*)$, com:

- O trecho em vermelho fornece: $v^* = (\frac{n}{\sigma^2} + \frac{1}{v})^{-1}$.
- O trecho em verde estabelece: $m^* = v^*(\frac{\sum_{i=1}^n y_i}{\sigma^2} + \frac{m}{v})$.

Conclusão: a distribuição *a priori* $\mu \sim N(m, v)$ e *a posteriori* $\mu|\boldsymbol{Y} \sim N(m^*, v^*)$ são da família Normal.

Estudo simulado usando o modelo $N(\mu, \sigma^2)$ com σ^2 conhecido:

Passos para gerar dados artificiais (comandos R):

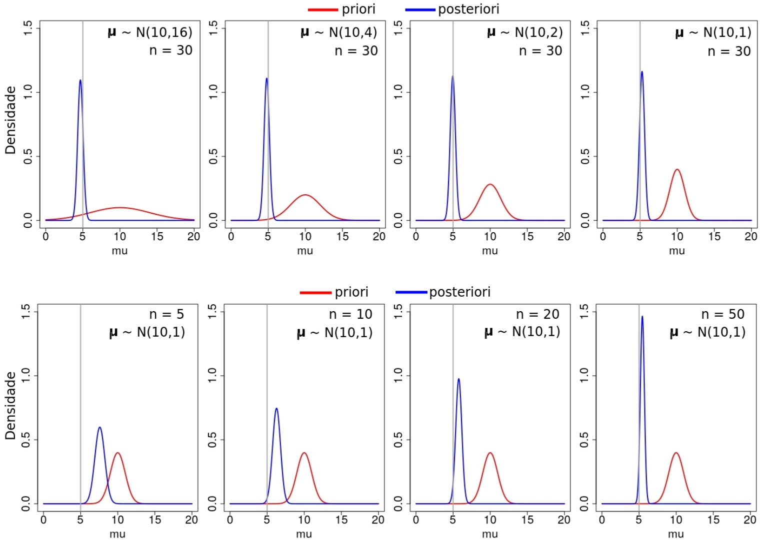
```
mureal = 5                           # defina o verdadeiro valor de mu;
sig2real = 4                         # defina o verdadeiro valor de sigma2;
n = 30                               # defina o tamanho amostral;
y = rnorm(n, mureal, sqrt(sig2real)) # gere observações da Normal.
```

Na prática, não conhecemos o valor real de μ. Ele é usado acima apenas para gerar a amostra $\boldsymbol{y}$. O valor real $\sigma^2 = 4$ é conhecido nesta aplicação.

Os gráficos a seguir mostram a distribuição *a posteriori* perante diferentes especificações *a priori* e diferentes valores de n.

Comandos para cálculo da f.d.p. *a priori* e *a posteriori*:

```
mu = seq(0.01, 20, 0.01)
prior = dnorm(mu, m, sqrt(v))
vs = 1 / ( n/sig2real + 1/v )
ms = vs * ( sum(y)/sig2real + m/v )
post = dnorm(mu, ms, sqrt(vs))
```



priori
posteriori
μ ~ N(10,16)
n = 30
μ ~ N(10,4)
n = 30
μ ~ N(10,2)
n = 30
μ ~ N(10,1)
n = 30
Densidade
mu
priori
posteriori
n = 5
μ ~ N(10,1)
n = 10
μ ~ N(10,1)
n = 20
μ ~ N(10,1)
n = 50
μ ~ N(10,1)
Densidade
mu

Comentários sobre os gráficos no slide anterior:

- As distribuições *a priori* estão centradas no valor 10. Note que $\mu_{\text{real}} = 5$, então a informação *a priori* está equivocada.
- Nos casos $n = 30$, conforme $Var(\mu)$ diminui, a f.d.p. *a posteriori* é muito pouco afetada; há um ligeiro deslocamento à direita no cenário mais informativo com $\mu \sim N(10, 1)$. A verossimilhança tende a ser dominante quando $n = 30$.
- Para $n = 5$, a especificação equivocada $\mu \sim N(10, 1)$ tem grande impacto e vicia o resultado *a posteriori* em azul (f.d.p. azul concentra massa de probabilidade longe de $\mu_{\text{real}} = 5$).
- Conforme n aumenta, a verossimilhança tende a dominar a informação *a priori* e assim a f.d.p. *a posteriori* passa a concentrar sua massa de probabilidade próximo de $\mu_{\text{real}} = 5$.

Distribuição $N(\mu, \sigma^2)$ com média conhecida: A família de distribuições Gama Inversa é conjugada com o modelo Normal para estimar a variância σ^2.

Seja $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ uma amostra aleatória da $N(\mu, \sigma^2)$. Admita independência condicional entre os Y_i's dado σ^2.

Espaço paramétrico: $\sigma^2 \in \mathbb{R}^+$.

Verossimilhança: $f_{\boldsymbol{Y}|\sigma^2}(\boldsymbol{y}) = \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\}$.

$= \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\{-\frac{1}{2\sigma^2}(y_i^2 - 2\mu y_i + \mu^2)\}$.

$= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2)\}$.

Distribuição *a priori*: $\sigma^2 \sim \text{GI}(a, b)$.

GI = Gama Inversa.

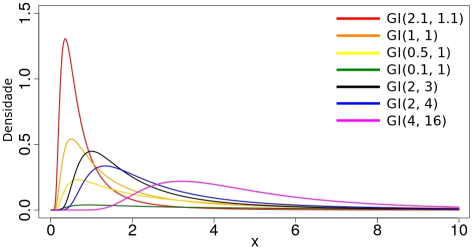
Definição: Seja X uma variável aleatória contínua com distribuição **Gama Inversa** (ou Gama Invertida), iremos denotar $X \sim \text{GI}(a, b)$ com $a > 0$ e $b > 0$. Sua f.d.p é dada por:

$$\frac{b^a}{\Gamma(a)}\, x^{-a-1}\, \exp\{-b/x\} \quad \text{para} \quad x > 0.$$

Além disso temos: $E(X) = \frac{b}{a-1}$ e $Var(X) = \frac{b^2}{(a-1)^2\,(a-2)}$

$E(X)$ existe para $a > 1$ e $Var(X)$ existe para $a > 2$.

Observação: Se $W \sim \text{Ga}(a, b)$, então $1/W \sim \text{GI}(a, b)$.



Distribuição *a posteriori*: $f_{\sigma^2|\boldsymbol{Y}}(\sigma^2|\boldsymbol{y}) \propto f_{\boldsymbol{Y}|\sigma^2}(\boldsymbol{y})\, f_{\sigma^2}(\sigma^2)$

$$\propto (2\pi)^{-n/2}\,(\sigma^2)^{-n/2}\,\exp\{-\tfrac{1}{2\sigma^2}(\textstyle\sum_{i=1}^n y_i^2 - 2\mu\sum_{i=1}^n y_i + n\mu^2)\} \times$$

$$\times \tfrac{b^a}{\Gamma(a)}\,(\sigma^2)^{-a-1}\,\exp\{-b/\sigma^2\}.$$

Termos destacados em vermelho são constantes (não dependem de σ^2). Eles serão incorporados na constante normalizadora.

$$\propto (\sigma^2)^{-n/2}\,\exp\{-\tfrac{1}{2\sigma^2}(\textstyle\sum_{i=1}^n y_i^2 - 2\mu\sum_{i=1}^n y_i + n\mu^2)\,(\sigma^2)^{-a-1}\,\exp\{-b/\sigma^2\}.$$

$$\propto (\sigma^2)^{-(a+n/2)-1}\,\exp\{-\tfrac{1}{\sigma^2}[b + \tfrac{1}{2}(\textstyle\sum_{i=1}^n y_i^2 - 2\mu\sum_{i=1}^n y_i + n\mu^2)]\}.$$

Temos aqui o núcleo da distribuição GI(a^*,b^*), com:

- O trecho em vermelho fornece: $a^* = a + n/2$.
- O trecho em verde estabelece: $b^* = b + \frac{1}{2}(\sum_{i=1}^n y_i^2 - 2\mu\sum_{i=1}^n y_i + n\mu^2)$.

Conclusão: a distribuição *a priori* $\sigma^2 \sim \text{GI}(a,b)$ e *a posteriori* $\sigma^2|\boldsymbol{Y} \sim \text{GI}(a^*,b^*)$ são da família Gama Inversa.

Estudo simulado usando o modelo $N(\mu, \sigma^2)$ com μ conhecido:

Gerando dados artificiais (comandos R):

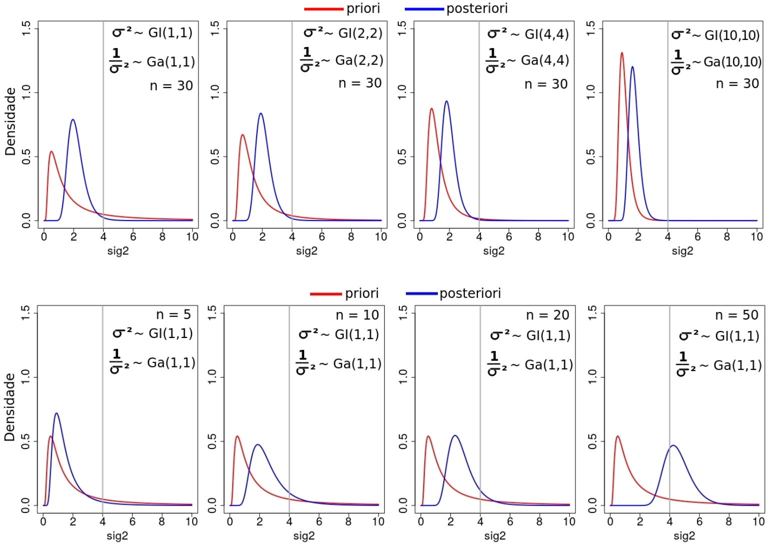
```
mureal = 5                             # defina o verdadeiro valor de mu;
sig2real = 4                           # defina o verdadeiro valor de sigma2;
n = 30                                 # defina o tamanho amostral;
y = rnorm(n, mureal, sqrt(sig2real)) # gere observações da Normal.
```

Na prática, σ^2_{real} é desconhecido. Neste problema, o pesquisador teria acesso apenas à amostra $\boldsymbol{y}$ e ao valor $\mu_{\text{real}} = 5$.

Os gráficos a seguir mostram a distribuição *a posteriori* perante diferentes especificações *a priori* e diferentes valores de n.

Comandos para cálculo da f.d.p. *a priori* e *a posteriori*:

```
dGI = function(x,a,b){ x^{-a-1} * exp(-b/x) * (b^a)/gamma(a) }
sig2 = seq(0.01, 10, 0.01)
prior = dGI(sig2, a, b)
as = a + n/2
bs = b + 0.5*(sum(y^2) -2*mureal*sum(y) + n*mureal^2)
post = dGI(sig2, as, bs)
```


priori
posteriori
σ²~ GI(1,1)
1/σ² ~ Ga(1,1)
n = 30
σ²~ GI(2,2)
1/σ² ~ Ga(2,2)
n = 30
σ²~ GI(4,4)
1/σ² ~ Ga(4,4)
n = 30
σ²~ GI(10,10)
1/σ² ~ Ga(10,10)
n = 30
n = 5
n = 10
n = 20
n = 50
σ²~ GI(1,1)
1/σ² ~ Ga(1,1)
Densidade
sig2

Comentários sobre os gráficos no slide anterior:

- Todas as especificações *a priori* indicam $E(1/\sigma^2) = a/b = 1$ (lembre-se que $1/\sigma^2_{\text{real}} = 1/4$, i.e., priori equivocada).
- $a = b$ aumentando $\Rightarrow$ $Var(1/\sigma^2) = a/b^2$ diminuindo.
- $1/\sigma^2 \sim \text{Ga}(10, 10) \Rightarrow Var(1/\sigma^2)$ pequena e $E(1/\sigma^2) = 1 \neq 1/4$; f.d.p. *a posteriori* é muito influenciada e desloca para a esquerda.
- Para $n = 30$, a verossimilhança não teve força suficiente para dominar a informação *a priori*; f.d.p. *a posteriori* (azul) tem massa de probabilidade longe de $\sigma^2_{\text{real}} = 4$.
- $n = 5$ determina uma f.d.p. *a posteriori* bastante afetada pela informação inicial equivocada.
- $n = 50$ determina uma f.d.p. *a posteriori* que concentra-se ao redor de $\sigma^2_{\text{real}} = 4$.

Exercício: Use o `R` para simular e construir este tipo de gráfico. Assuma $\mu_{\text{real}} = 5$, $\sigma^2_{\text{real}} = 4$, $n = 30$ e a distribuição *a priori* $\sigma^2 \sim \text{GI}(2.1, 1.1)$. Calcule a esperança e a variância das distribuições *a priori* e *a posteriori* de σ^2.

Reparametrizando a distribuição Normal para trabalhar com a precisão:

Definição: Se σ^2 é um parâmetro de variância, dizemos que sua inversa (dada por $\phi = 1/\sigma^2$) é um parâmetro de precisão. Quanto maior for o valor da variância σ^2, menor será o valor da precisão ϕ (e vice versa). Visto que $\sigma^2 > 0$, temos $\phi > 0$.

Considere a variável aleatória $Y_i|\mu, \sigma^2 \sim N(\mu, \sigma^2)$. Podemos reformular sua f.d.p. para expressá-la em termos da precisão $\phi = 1/\sigma^2$. Isso é feito com segue:

$$
\begin{aligned}
f_{Y_i|\mu,\sigma^2}(y_i) &= (2\pi\sigma^2)^{-1/2} \exp\{-\tfrac{1}{2\sigma^2}(y_i - \mu)^2\}. \\
f_{Y_i|\mu,\phi}(y_i) &= [2\pi(1/\phi)]^{-1/2} \exp\{-\tfrac{1}{2(1/\phi)}(y_i - \mu)^2\}. \\
&= (2\pi/\phi)^{-1/2} \exp\{-\tfrac{\phi}{2}(y_i - \mu)^2\} \quad \text{para} \quad y_i \in \mathbb{R}.
\end{aligned}
$$

Tomando como base a reparametrização acima, o próximo slide volta a discutir a família conjugada para o modelo Normal com média conhecida.

Distribuição $N(\mu, 1/\phi)$ com média conhecida: A família de distribuições Gama é conjugada com o modelo Normal para estimar a precisão ϕ.

Assuma que $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ é amostra aleatória da $N(\mu, \sigma^2)$. Admita independência condicional dos Y_i's dado $\sigma^2 = 1/\phi$.

Espaço paramétrico: $\phi \in \mathbb{R}^+$.

Verossimilhança: $f_{\boldsymbol{Y}|\phi}(\boldsymbol{y}) = \prod_{i=1}^n (2\pi/\phi)^{-1/2} \exp\{-\frac{\phi}{2}(y_i - \mu)^2\}$.

$= \prod_{i=1}^n (2\pi/\phi)^{-1/2} \exp\{-\frac{\phi}{2}(y_i^2 - 2\mu y_i + \mu^2)\}$.

$= (2\pi)^{-n/2} (\phi)^{n/2} \exp\{-\frac{\phi}{2}(\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2)\}$.

Distribuição *a priori*: $\phi \sim \text{Ga}(a, b)$.

Distribuição *a posteriori*: $f_{\phi|\boldsymbol{Y}}(\phi|\boldsymbol{y}) \propto f_{\boldsymbol{Y}|\phi}(\boldsymbol{y})\, f_{\phi}(\phi)$

$$\propto (2\pi)^{-n/2}\, \phi^{n/2} \exp\{-\tfrac{\phi}{2}(\textstyle\sum_{i=1}^{n} y_i^2 - 2\mu \sum_{i=1}^{n} y_i + n\mu^2)\} \times$$

$$\times \tfrac{b^a}{\Gamma(a)}\, \phi^{a-1} \exp\{-b\phi\}.$$

Termos em vermelho são constantes (não dependem de ϕ). Eles serão incorporados na constante normalizadora.

$$f_{\phi|\boldsymbol{Y}}(\phi|\boldsymbol{y}) \propto \phi^{n/2} \exp\{-\tfrac{\phi}{2}(\textstyle\sum_{i=1}^{n} y_i^2 - 2\mu \sum_{i=1}^{n} y_i + n\mu^2)\ \phi^{a-1} \exp\{-b\phi\}.$$

$$\propto \phi^{(a+n/2)-1} \exp\{-\phi[b + \tfrac{1}{2}(\textstyle\sum_{i=1}^{n} y_i^2 - 2\mu \sum_{i=1}^{n} y_i + n\mu^2)]\}.$$

Temos aqui o núcleo da distribuição Ga(a^*,b^*), com:

- O trecho em vermelho fornece: $a^* = a + n/2$.
- O trecho em verde estabelece: $b^* = b + \frac{1}{2}(\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2)$.

Conclusão: a distribuição *a priori* $\phi \sim \text{Ga}(a, b)$ e *a posteriori* $\phi|\boldsymbol{Y} \sim \text{Ga}(a^*, b^*)$ são da família Gama.

Note que não era necessário ter feito todas estas últimas contas para obter a distribuição *a posteriori* no caso reparametrizado com ϕ.

Bastava ter percebido que:
se $\sigma^2|\boldsymbol{Y} \sim \text{GI}(a^*, b^*)$, então $(\sigma^2)^{-1}|\boldsymbol{Y} = \phi|\boldsymbol{Y} \sim \text{Ga}(a^*, b^*)$.
As expressões de a^* e b^* são as mesmas com ou sem reparametrização.

Definição: Considere um par de variáveis aleatórias (X_1,X_2) e assuma que $X_1|X_2 \sim N(m, v/X_2)$, significando uma distribuição Normal com média m e variância v/X_2.

Suponha também que $X_2 \sim \text{Ga}(a, b)$.

Dizemos, neste caso, que $(X_1, X_2) \sim \text{Normal-Gama}(m, v, a, b)$, com $m \in \mathbb{R}$, $v > 0$, $a > 0$ e $b > 0$, e sua f.d.p. $f_{X_1,X_2}(x_1, x_2)$ é dada por:

$$\frac{b^a}{\Gamma(a)} \, (2\pi v)^{-1/2} \, x_2^{a-1/2} \exp\{-bx_2\} \, \exp\{-\frac{1}{2v}(x_1 - m)^2\} \text{ para } x_1 \in \mathbb{R} \text{ e } x_2 \in \mathbb{R}^+.$$

Neste contexto, $E(X_1) = m$, $E(X_2) = a/b$, $Var(X_1) = \frac{bv}{a-1}$ e $Var(X_2) = a/b^2$.

A Normal-Gama é uma família bivariada de distribuições contínuas contendo 4 parâmetros (m,v,a,b).

Esta distribuição será utilizada no próximo tópico.

Distribuição $N(\mu, 1/\phi)$ com média e variância desconhecidas: A família Normal-Gama é conjugada com o modelo Normal para estimar o par (μ, ϕ).

Considere $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ uma amostra aleatória da $N(\mu, 1/\phi)$ com os dois parâmetros desconhecidos. Admita independência condicional entre os Y_i's dado (μ, ϕ).

Espaço paramétrico: $\mu \in \mathbb{R}$ e $\phi \in \mathbb{R}^+$.

Verossimilhança: $f_{\boldsymbol{Y}|\mu,\phi}(\boldsymbol{y}) = \prod_{i=1}^{n} (2\pi/\phi)^{-1/2} \exp\{-\frac{\phi}{2}(y_i - \mu)^2\}$.

$= (2\pi)^{-n/2} \, \phi^{n/2} \exp\{-\frac{\phi}{2}(\sum_{i=1}^{n} y_i^2 - 2\mu \sum_{i=1}^{n} y_i + n\mu^2)\}$.

Distribuição *a priori*: $f_{\mu,\phi}(\mu, \phi) = f_{\mu|\phi}(\mu) \, f_{\phi}(\phi)$.

Este é um modelo bi-paramétrico, então a distribuição *a priori* a ser escolhida é uma distribuição conjunta conforme escrito acima.

Adote: $\mu|\phi \sim N(m, v/\phi)$ e $\phi \sim \text{Ga}(a, b)$.

Veja que esta escolha significa: $(\mu, \phi) \sim \text{Normal-Gama}(m, v, a, b)$.

Importante: esta especificação conjunta não estabelece independência entre μ e ϕ.

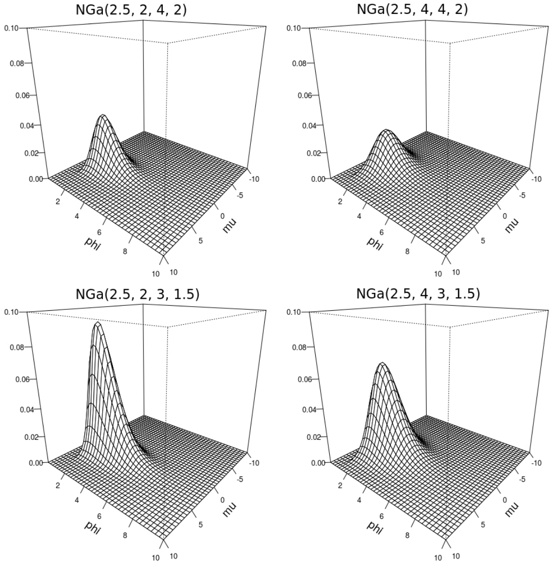
$$f_{\mu,\phi}(\mu, \phi) \; = \; (2\pi v/\phi)^{-1/2} \; \exp\{-\tfrac{\phi}{2v}(\mu - m)^2\} \quad \tfrac{b^a}{\Gamma(a)} \, \phi^{a-1} \exp\{-b\phi\}.$$

Distribuição *a posteriori*: $f_{\mu,\phi|\mathbf{Y}}(\mu, \phi|\mathbf{y}) \; \propto \; f_{\mathbf{Y}|\mu,\phi}(\mathbf{y}) \; f_{\mu,\phi}(\mu, \phi)$

$$\propto \; (2\pi)^{-n/2} \; \phi^{n/2} \; \exp\{-\tfrac{\phi}{2}(\textstyle\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2)\} \; \times$$

$$\times \; (2\pi v)^{-1/2} \; \phi^{1/2} \; \exp\{-\tfrac{\phi}{2v}(\mu^2 - 2\mu m)\} \; \exp\{-\tfrac{\phi}{2v} m^2\} \quad \tfrac{b^a}{\Gamma(a)} \, \phi^{a-1} \exp\{-b\phi\}.$$

Termos em vermelho são constantes (não dependem de μ ou ϕ). Eles serão incorporados na constante normalizadora.


NGa(2.5, 2, 4, 2)
NGa(2.5, 4, 4, 2)
NGa(2.5, 2, 3, 1.5)
NGa(2.5, 4, 3, 1.5)
0.10
0.08
0.06
0.04
0.02
0.00
phi
2
4
6
8
10
mu
-10
-5
0
5
10

A estratégia para continuar a conta é separar a expressão em duas partes conforme estabelece a seguinte fatoração da f.d.p. conjunta *a posteriori*:

$$f_{\mu,\phi|\boldsymbol{Y}}(\mu,\phi|\boldsymbol{y}) \;=\; f_{\mu|\phi,\boldsymbol{Y}}(\mu|\boldsymbol{y})\; f_{\phi|\boldsymbol{Y}}(\phi).$$

- Parte vermelha: depende de μ e pode conter ϕ.
- Parte verde: depende apenas de ϕ.

Diante desta visão, aplica-se a seguinte separação:

$$f_{\mu,\phi|\boldsymbol{Y}}(\mu,\phi|\boldsymbol{y}) \;\propto\; \phi^{1/2}\; \exp\{-\tfrac{\phi}{2}(-2\mu\textstyle\sum_{i=1}^n y_i + n\mu^2)\}\;\; \exp\{-\tfrac{\phi}{2v}(\mu^2 - 2\mu m)\} \times$$

$$\times\; \phi^{n/2}\; \exp\{-\tfrac{\phi}{2}(\textstyle\sum_{i=1}^n y_i^2)\}\; \exp\{-\tfrac{\phi}{2v} m^2\}\; \phi^{a-1}\; \exp\{-b\phi\}.$$

$$\propto\; \phi^{1/2}\; \exp\{-\tfrac{1}{2}[-2\mu\phi\textstyle\sum_{i=1}^n y_i + n\phi\mu^2 + \tfrac{\phi}{v}\mu^2 - 2\mu\tfrac{\phi m}{v}]\} \times$$

$$\times\; \phi^{a+n/2-1}\; \exp\{-\phi[b + \tfrac{\sum_{i=1}^n y_i^2}{2} + \tfrac{m^2}{2v}]\}.$$

$$f_{\mu,\phi|\boldsymbol{Y}}(\mu,\phi|\boldsymbol{y}) \propto \phi^{1/2} \exp\{-\tfrac{1}{2}[\mu^2(n\phi+\tfrac{\phi}{v}) - 2\mu(\phi\textstyle\sum_{i=1}^n y_i + \tfrac{\phi m}{v})]\} \times$$

$$\times\ \phi^{a+n/2-1} \exp\{-\phi[b + \tfrac{\sum_{i=1}^n y_i^2}{2} + \tfrac{m^2}{2v}]\}.$$

$$\propto \phi^{1/2} \exp\{-\tfrac{1}{2}\,(n+\tfrac{1}{v})\,\phi\,[\mu^2 - 2\mu\,(n+\tfrac{1}{v})^{-1}\,(\textstyle\sum_{i=1}^n y_i + \tfrac{m}{v})]\} \times$$

$$\times\ \phi^{a+n/2-1} \exp\{-\phi[b + \tfrac{\sum_{i=1}^n y_i^2}{2} + \tfrac{m^2}{2v}]\}.$$

Simplificando a notação, denote:

$v^* = (n+\frac{1}{v})^{-1}$ e $m^* = v^*(\sum_{i=1}^n y_i + \frac{m}{v})$.

Então:

$$f_{\mu,\phi|\boldsymbol{Y}}(\mu,\phi|\boldsymbol{y}) \propto \phi^{1/2} \exp\{-\tfrac{\phi}{2v^*}\,[\mu^2 - 2\mu\, m^* + (m^*)^2 - (m^*)^2]\} \times$$

$$\times\ \phi^{a+n/2-1} \exp\{-\phi[b + \tfrac{\sum_{i=1}^n y_i^2}{2} + \tfrac{m^2}{2v}]\}.$$

$$f_{\mu,\phi|\boldsymbol{Y}}(\mu,\phi|\boldsymbol{y}) \propto \phi^{1/2} \exp\{-\tfrac{\phi}{2v^*}\ [\mu^2 - 2\mu\ m^* \ + (m^*)^2]\} \exp\{+\tfrac{\phi}{2v^*}(m^*)^2\} \times$$

$$\times\ \phi^{a+n/2-1} \exp\{-\phi[b + \tfrac{\sum_{i=1}^n y_i^2}{2} + \tfrac{m^2}{2v}]\}.$$

$$\propto \phi^{1/2} \exp\{-\tfrac{\phi}{2v^*}\ (\mu - m^*)^2\} \times$$

$$\times\ \phi^{a+n/2-1} \exp\{-\phi[b + \tfrac{\sum_{i=1}^n y_i^2}{2} + \tfrac{m^2}{2v}]\} \exp\{+\tfrac{\phi}{2v^*}(m^*)^2\}.$$

$$\propto \phi^{1/2} \exp\{-\tfrac{\phi}{2v^*}\ (\mu - m^*)^2\}\ \phi^{a+n/2-1} \exp\{-\phi[b + \tfrac{\sum_{i=1}^n y_i^2}{2} + \tfrac{m^2}{2v} - \tfrac{(m^*)^2}{2v^*}]\}.$$

Novamente simplificando a notação denote:

$$a^* = a + n/2 \quad \text{e} \quad b^* = b + \frac{\sum_{i=1}^n y_i^2}{2} + \frac{m^2}{2v} - \frac{(m^*)^2}{2v^*}.$$

$$f_{\mu,\phi|\boldsymbol{Y}}(\mu,\phi|\boldsymbol{y}) = (2\pi v^*)^{-1/2}\ \phi^{1/2} \exp\{-\tfrac{\phi}{2v^*}\ (\mu - m^*)^2\} \times$$

$$\times \frac{(b^*)^{a^*}}{\Gamma(a^*)}\ \phi^{a^*-1} \exp\{-\phi b^*\}.$$

Os termos adicionados em preto são constantes (não dependem de μ ou ϕ).

Logo $(\mu,\phi)|\boldsymbol{Y} \sim$ Normal-Gama(m^*, v^*, a^*, b^*) com

$v^* = (n + \frac{1}{v})^{-1}, \quad m^* = v^*(\sum_{i=1}^n y_i + \frac{m}{v}),$

$a^* = a + n/2$ e $b^* = b + \frac{\sum_{i=1}^n y_i^2}{2} + \frac{m^2}{2v} - \frac{(m^*)^2}{2v^*}.$

Conclusão: as distribuições *a priori* $(\mu,\phi) \sim$ Normal-Gama(m, v, a, b) e *a posteriori* $(\mu,\phi)|\boldsymbol{Y} \sim$ Normal-Gama(m^*, v^*, a^*, b^*) são da família Normal-Gama.

Ressalta-se novamente que assumir a $(\mu,\phi) \sim$ Normal-Gama *a priori* estabelece uma <u>dependência</u> entre estes parâmetros.

Pergunta: Como fica a análise Bayesiana para o caso Normal com μ e ϕ desconhecidos sob a suposição de independência *a priori* entre μ e ϕ?

A implicação desta suposição é que $f_{\mu,\phi}(\mu,\phi) = f_\mu(\mu)\, f_\phi(\phi).$

Adote $\mu \sim N(m, v)$ e $\phi \sim \text{Ga}(a, b).$

$$f_{\mu,\phi}(\mu,\phi) \;=\; (2\pi v)^{-1/2}\; \exp\{-\tfrac{1}{2v}(\mu-m)^2\} \;\; \tfrac{b^a}{\Gamma(a)}\, \phi^{a-1} \exp\{-b\phi\}.$$

Distribuição *a posteriori*: $\quad f_{\mu,\phi|\mathbf{Y}}(\mu,\phi|\mathbf{y}) \;\propto\; f_{\mathbf{Y}|\mu,\phi}(\mathbf{y})\; f_{\mu,\phi}(\mu,\phi)$

$$\propto\; (2\pi)^{-n/2}\; \phi^{n/2}\; \exp\{-\tfrac{\phi}{2}(\textstyle\sum_{i=1}^n y_i^2 - 2\mu\sum_{i=1}^n y_i + n\mu^2)\} \times$$

$$\times\; (2\pi v)^{-1/2}\; \exp\{-\tfrac{1}{2v}(\mu^2 - 2\mu m)\}\; \exp\{-\tfrac{1}{2v}m^2\} \;\; \tfrac{b^a}{\Gamma(a)}\; \phi^{a-1} \exp\{-b\phi\}.$$

Termos em vermelho são constantes (não dependem de μ ou ϕ). Eles serão incorporados na constante normalizadora.

Fatoração da f.d.p. conjunta *a posteriori*:

$$f_{\mu,\phi|\mathbf{Y}}(\mu,\phi|\mathbf{y}) \;=\; f_{\mu|\phi,\mathbf{Y}}(\mu|\mathbf{y})\; f_{\phi|\mathbf{Y}}(\phi).$$

- Parte vermelha: depende de μ e pode conter ϕ.
- Parte verde: depende apenas de ϕ.

$$f_{\mu,\phi|\mathbf{Y}}(\mu,\phi|\mathbf{y}) \propto \exp\{-\tfrac{\phi}{2}(-2\mu\sum_{i=1}^n y_i + n\mu^2)\} \;\; \exp\{-\tfrac{1}{2v}(\mu^2 - 2\mu m)\} \times$$
$$\times\; \phi^{n/2} \; \exp\{-\tfrac{\phi}{2}(\sum_{i=1}^n y_i^2)\} \; \phi^{a-1} \; \exp\{-b\phi\}.$$
$$\propto \exp\{-\tfrac{1}{2}(-2\mu\phi\sum_{i=1}^n y_i + n\phi\mu^2 + \tfrac{1}{v}\mu^2 - 2\mu\tfrac{m}{v})\} \times$$
$$\times\; \phi^{a+n/2-1} \; \exp\{-\phi[b + \tfrac{\sum_{i=1}^n y_i^2}{2}]\}.$$
$$\propto \exp\{-\tfrac{1}{2}[\mu^2(n\phi + \tfrac{1}{v}) - 2\mu(\phi\sum_{i=1}^n y_i + \tfrac{m}{v})]\} \times$$
$$\times\; \phi^{a+n/2-1} \; \exp\{-\phi[b + \tfrac{\sum_{i=1}^n y_i^2}{2}]\}.$$
$$\propto \exp\{-\tfrac{1}{2}(n\phi + \tfrac{1}{v})\;[\mu^2 - 2\mu\;(n\phi + \tfrac{1}{v})^{-1}(\phi\sum_{i=1}^n y_i + \tfrac{m}{v})]\} \times$$
$$\times\; \phi^{a+n/2-1} \; \exp\{-\phi[b + \tfrac{\sum_{i=1}^n y_i^2}{2}]\}.$$

Denote: $v_\phi^* = (n\phi + \frac{1}{v})^{-1}$, $\quad m_\phi^* = v_\phi^*\,(\phi\sum_{i=1}^n y_i + \frac{m}{v})$,

$\tilde{a} = a + n/2$ e $\tilde{b} = b + \sum_{i=1}^n y_i^2/2$.

O subscrito "ϕ" foi usado para lembrar que as expressões dependem de ϕ.

$$f_{\mu,\phi|\boldsymbol{Y}}(\mu,\phi|\boldsymbol{y}) \propto \exp\{-\tfrac{1}{2v_\phi^*}\,[\mu^2 - 2\mu\, m_\phi^* + (m_\phi^*)^2 - (m_\phi^*)^2]\} \times$$

$$\times\ \phi^{\tilde{a}-1}\,\exp\{-\phi\tilde{b}\}.$$

$$\propto \exp\{-\tfrac{1}{2v_\phi^*}\,[\mu^2 - 2\mu\, m_\phi^* + (m_\phi^*)^2]\}\ \exp\{+\tfrac{(m_\phi^*)^2}{2v_\phi^*}\} \times \phi^{\tilde{a}-1}\,\exp\{-\phi\tilde{b}\}.$$

$$\propto (v_\phi^*)^{-1/2}\,\exp\{-\tfrac{1}{2v_\phi^*}\,(\mu - m_\phi^*)^2\} \times (v_\phi^*)^{1/2}\exp\{+\tfrac{(m_\phi^*)^2}{2v_\phi^*}\} \times \phi^{\tilde{a}-1}\,\exp\{-\phi\tilde{b}\}.$$

O termo **vermelho** é o núcleo de uma Normal. Já o termo **verde** é o núcleo de uma Gama. Perceba que o termo central (em **preto**) depende de ϕ e, portanto, não pode ser incorporado na constante normalizadora. Além disso, ele não pode ser combinado com o termo **vermelho** ou **verde** para formar uma distribuição de probabilidade conhecida.

Conclusão: Não haverá conjugação quando assumimos independência *a priori* para (μ,ϕ). Usar a suposição de independência nos coloca em uma situação em que não sabemos identificar a distribuição *a posteriori* conjunta $f_{\mu,\phi|\boldsymbol{Y}}(\mu,\phi|\boldsymbol{y})$.

Voltaremos a este problema mais adiante no curso.

Distribuição Multinomial: A família Dirichlet é conjugada com o modelo Multinomial para estimar o vetor de probabilidades $\theta = \{\theta_1, \theta_2, \cdots, \theta_q\}$.

Suponha que $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ é uma amostra aleatória de tamanho n relativa a um experimento em que Y_i indica em qual categoria, de q possíveis, o elemento i faz parte. Podemos escrever que $Y_i = j$ se o i-ésimo experimento resultou na categoria $j \in \{1, 2, \cdots, q\}$.

Seja $W = \{W_1, W_2, \cdots, W_q\}$ um conjunto de variáveis aleatórias, tal que $W_j = \sum_{i=1}^{n} 1_{\{Y_i = j\}}$ é o número de experimentos que indicaram categoria j. Admita que θ_j é a probabilidade de um elemento qualquer ser classificado na categoria j. Notação: $\theta = \{\theta_1, \theta_2, \cdots, \theta_q\}$.

Importante: $\sum_{j=1}^{q} W_j = n$ e $\sum_{j=1}^{q} \theta_j = 1$.

Dizemos que W tem distribuição Multinomial(n,θ) com a seguinte f.m.p.

$$f_{W|\theta}(w) = \frac{n!}{\prod_{j=1}^{q} w_j!} \prod_{j=1}^{q} \theta_j^{w_j} \quad \text{para} \quad \theta_j \in [0, 1], \quad w_j \in \{0, 1, 2, \cdots, n\}.$$

Exemplo: Um dado de 6 faces ($q = 6$ categorias) é lançado $n = 100$ vezes. Neste caso, $Y_i = j$ se o i-ésimo lançamento indicou a face $j \in \{1, 2, 3, 4, 5, 6\}$.

Uma observação do vetor aleatório $W = \{W_1, W_2, W_3, W_4, W_5, W_6\}$ indica o número de vezes que cada face ocorreu.

É natural, neste contexto, supor que $\theta_j = 1/6$ para todo j.

Considere a simulação abaixo usando comandos do R.

```
# Simulando os 100 lançamentos do dado.
y = sample(1:6, 100, replace = TRUE, prob = rep(1/6,6))

# Contando o número de ocorrência de cada face.
w = table(y)

 1   2   3   4   5   6
16  15  18  17  18  16

sum(w) # resulta em 100
```

Lembrete: A distribuição Beta é sempre atribuída para variáveis aleatórias contínuas limitadas no intervalo $[0,1]$. Conforme visto anteriormente, ela é uma opção adequada como distribuição *a priori* para a desconhecida probabilidade de sucesso θ_1, em um experimento do tipo sucesso e fracasso.

A probabilidade de fracasso seria denotada por $(1-\theta_1)=\theta_2$.

Veja que: $\theta=\{\theta_1,\theta_2\}$ é um par de probabilidades tal que $\theta_1+\theta_2=1$.

A distribuição Dirichlet é uma generalização da distribuição Beta, servindo para descrever o comportamento de um vetor de probabilidades $\theta=\{\theta_1,\theta_2,\cdots,\theta_q\}$ com dimensão $q>2$ e satisfazendo $\sum_{j=1}^q \theta_j=1$.

Definição: Se θ segue a distribuição **Dirichlet** com vetor paramétrico $a=(a_1,\ldots,a_q)^\top$, tal que $a_i>0$ e $\sum_{i=1}^q a_i=\alpha$, então sua f.d.p. será:

$$f_\theta(\theta)=\frac{\Gamma(\alpha)}{\prod_{j=1}^q \Gamma(a_j)} \prod_{j=1}^q \theta_j^{a_j-1} \quad \text{para} \quad \theta_j\in[0,1] \quad \text{e} \quad \sum_{j=1}^q \theta_j=1.$$

α é geralmente chamado de parâmetro de concentração.

A família Dirichlet com parâmetros inteiros $a = (a_1, \ldots, a_q)^\top$, notação Dir($a$), é conjugada natural com o modelo multinomial.

Novamente, pouco perdemos ao estender a conjugação natural considerando todas as distribuições Dirichlet, ou seja, sem a restrição de $a = (a_1, \ldots, a_q)^\top$ contendo apenas inteiros.

Verossimilhança Multinomial: $f_{W|\theta}(w) = \dfrac{n!}{\prod_{j=1}^{q} w_j!} \prod_{j=1}^{q} \theta_j^{w_j}$.

Distribuição *a priori* Dir(a): $f_\theta(\theta) = \dfrac{\Gamma(\alpha)}{\prod_{j=1}^{q} \Gamma(a_j)} \prod_{j=1}^{q} \theta_j^{a_j - 1}$

Os trechos destacados em verde são os núcleos (parte que depende do parâmetro alvo θ). Perceba a semelhança entre eles.

Esta semelhança é ponto chave para obter a conjugação.

Distribuição *a posteriori*:

$$f_{\theta|W}(\theta|w) \;\propto\; f_{W|\theta}(w)\, f_\theta(\theta) \;\propto\; \frac{n!}{\prod_{j=1}^{q} w_j!} \prod_{j=1}^{q} \theta_j^{w_j} \;\; \frac{\Gamma(\alpha)}{\prod_{j=1}^{q} \Gamma(a_j)} \; \prod_{j=1}^{q} \theta_j^{a_j-1}$$

Trechos vermelhos são constantes (não dependem de θ) que serão incorporadas à constante de normalização.

$$f_{\theta|W}(\theta|w) \;\propto\; \prod_{j=1}^{q} \theta_j^{w_j} \;\; \prod_{j=1}^{q} \theta_j^{a_j-1} \;\propto\; \prod_{j=1}^{q} \theta_j^{a_j+w_j-1}.$$

Este é o núcleo da Dirichlet com parâmetro $a^* = (a_1 + w_1, a_2 + w_2, \cdots, a_q + w_q)$.

$$f_{\theta|W}(\theta|w) \;=\; \frac{\Gamma(\alpha+n)}{\prod_{j=1}^{q} \Gamma(a_j + w_j)} \; \prod_{j=1}^{q} \theta_j^{a_j+w_j-1}. \qquad \text{Denota-se } \theta|w \sim \text{Dir}(a^*).$$

Conclusão: as distribuições *a priori* $\theta \sim \text{Dir}(a)$ e *a posteriori* $\theta|W \sim \text{Dir}(a^*)$ são da família Dirichlet.

Conjugação na família exponencial.

Definição: A família exponencial é formada pelas funções (f.d.p. ou f.m.p.) que podem ser escritas na seguinte configuração

$$f_{Y|\theta}(y) \;=\; s(y)\, t(\theta)\, \exp\{a(y)\, b(\theta)\},$$

em que $s(\bullet)$ e $t(\bullet)$ são funções que assumem valores em $\mathbb{R}^+$.
Ademais $a(\bullet)$ e $b(\bullet)$ são funções vetoriais de dimensão K.

Esta expressão pode ser rescrita como:

$$f_{Y|\theta}(y) \;=\; \exp\{a(y)\, b(\theta) + c(\theta) + d(y)\},$$

sendo $\quad c(\theta) = \ln[t(\theta)] \quad$ e $\quad d(y) = \ln[s(y)]$.

Resultado de inferência: Pelo teorema da fatoração de Neyman, o termo $a(y)$ é uma estatística suficiente para θ.

Várias distribuições conhecidas fazem parte da família exponencial, tais como: Poisson, Binomial, Binomial Negativa, Rayleigh, Exponencial, Gama, Normal e Normal Inversa (citando apenas algumas).

O suporte $\{y;\ f_{Y|\theta}(y) > 0\}$ de uma distribuição membro da família exponencial não pode depender de θ. Assim, por exemplo, a distribuição U(0,θ) não é membro desta família, pois não podemos identificar $a(y)$ e $b(\theta)$ através da f.d.p.:

$$f_{Y|\theta}(y) \ = \ \frac{1}{\theta} \ 1_{\{0<y<\theta\}} \quad \text{em que} \quad 1_{\{\bullet\}} \text{ é função indicadora.}$$

Mostrar que uma distribuição faz parte da família exponencial não é difícil. Veja três exemplos a seguir.

Exemplo: Considere uma variável aleatória discreta com distribuição Poisson; ou seja, $Y \sim \text{Poisson}(\theta)$ com $\theta > 0$. Sua f.m.p. é dada por

$$f_{Y|\theta}(y) \;=\; \frac{\theta^y}{y!}\exp\{-\theta\} \;=\; \frac{1}{y!}e^{-\theta}\exp\{y\ln(\theta)\} \;\text{ para }\; y = 0, 1, 2, \ldots$$

Note que: $\quad s(y) = 1/y!, \quad t(\theta) = e^{-\theta}, \quad a(y) = y \quad \text{e} \quad b(\theta) = \ln(\theta).$

Veja que $a(y)$ e $b(\theta)$ são escalares $(K = 1)$ neste caso.

Exemplo: Seja $Y \sim \text{Binomial}(m, \theta)$ com $0 < \theta < 1$ e $m = n^o$ conhecido de ensaios Bernoulli independentes. A f.m.p. é como segue

$$f_{Y|\theta}(y) \;=\; \binom{m}{y}\theta^y(1-\theta)^{m-y} \;=\; \binom{m}{y}(1-\theta)^m\exp\left\{y\ln\left(\frac{\theta}{1-\theta}\right)\right\} \;\text{ para }\; y = 0, 1, \ldots, m.$$

Temos aqui: $\quad s(y) = \binom{m}{y}, \quad t(\theta) = (1-\theta)^m, \quad a(y) = y \quad \text{e} \quad b(\theta) = \ln\left(\frac{\theta}{1-\theta}\right).$

Observe que $a(y)$ e $b(\theta)$ são escalares $(K = 1)$ neste caso.

Lembrete: $\text{Binomial}(m = 1, \theta) = \text{Bernoulli}(\theta)$.

Exemplo: Considere $Y \sim N(\mu, \sigma^2)$ com $\mu \in \mathbb{R}$ e $\sigma^2 \in \mathbb{R}^+$.
Neste caso bi-paramétrico, denota-se $\theta = \{\mu, \sigma^2\}$.

$$
\begin{aligned}
f_{Y|\mu,\sigma^2}(y) &= (2\pi\sigma^2)^{-1/2} \, \exp\left\{ -\frac{1}{2\sigma^2}(y-\mu)^2 \right\} \quad \text{para} \quad y \in \mathbb{R}, \\
&= (2\pi\sigma^2)^{-1/2} \, \exp\left\{ -\frac{1}{2\sigma^2}(y^2 - 2y\mu + \mu^2) \right\}, \\
&= (2\pi\sigma^2)^{-1/2} \, \exp\left\{ -\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} \right\}, \\
&= (2\pi)^{-1/2} \, (\sigma^2)^{-1/2} \, \exp\left\{ -\frac{\mu^2}{2\sigma^2} \right\} \, \exp\left\{ (y^2, y) \left(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2} \right)^\top \right\}.
\end{aligned}
$$

Temos: $s(y) = (2\pi)^{-1/2}$, $t(\theta) = (\sigma^2)^{-1/2} \, \exp\left\{ -\frac{\mu^2}{2\sigma^2} \right\}$,
$a(y) = (y^2, y)$, $b(\theta) = \left(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2} \right)^\top$.
Perceba que $a(y)$ e $b(\theta)$ são vetores ($K = 2$).

Considere $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ uma amostra aleatória proveniente de uma distribuição membro da família exponencial. Esta distribuição é indexada pelo parâmetro θ (escalar ou vetor) e existe independência condicional dos Y_i's dado θ.

Podemos escrever: $f_{Y_i|\theta}(y_i) \; = \; s(y_i)\; t(\theta)\; \exp\{a(y_i)\; b(\theta)\}$.

Verosimilhança: $\quad f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \; = \; \prod_{i=1}^{n} s(y_i)\; t(\theta)\; \exp\{a(y_i)\; b(\theta)\}$.

$= \; \left[\prod_{i=1}^{n} s(y_i)\right]\; t(\theta)^n\; \exp\left\{\left[\sum_{i=1}^{n} a(y_i)\right]\; b(\theta)\right\}$.

Distribuição *a priori* conjugada para θ:

Note que o núcleo (parte atrelada a θ) da função de verossimilhança é

$f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \; \propto \; t(\theta)^n\; \exp\left\{\left[\sum_{i=1}^{n} a(y_i)\right]\; b(\theta)\right\}$.

Visando obter uma especificação *a priori* semelhante a este núcleo, denote $\alpha = n$ e $\beta = \sum_{i=1}^{n} a(y_i)$. A partir disso, escreva

$f_\theta(\theta) \; \propto \; t(\theta)^\alpha\; \exp\{\beta\; b(\theta)\}, \quad$ ou seja $\quad f_\theta(\theta) \; = \; C_N\; t(\theta)^\alpha\; \exp\{\beta\; b(\theta)\}$,

sendo $1/C_N = \int_\Theta\; t(\theta)^\alpha\; \exp\{\beta\; b(\theta)\}\; d\theta$.

A distribuição *a posteriori* $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y})$ é obtida via Teorema de Bayes como segue:

$$\frac{\left[\prod_{i=1}^{n} s(y_i)\right] \ t(\theta)^n \ \exp\left\{\left[\sum_{i=1}^{n} a(y_i)\right] \ b(\theta)\right\} \ \ C_N \ \ t(\theta)^{\alpha} \ \exp\left\{\beta \ b(\theta)\right\}}{\int_{\Theta} \ \left[\prod_{i=1}^{n} s(y_i)\right] \ t(\theta)^n \ \exp\left\{\left[\sum_{i=1}^{n} a(y_i)\right] \ b(\theta)\right\} \ \ C_N \ \ t(\theta)^{\alpha} \ \exp\left\{\beta \ b(\theta)\right\} \ d\theta}.$$

Termos em vermelho são constantes (não dependem de θ) e irão cancelar.

$$f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \ = \ \frac{t(\theta)^{\alpha+n} \ \exp\left\{\left[\sum_{i=1}^{n} a(y_i)\right] \ b(\theta) \ + \ \beta \ b(\theta)\right\}}{\int_{\Theta} \ t(\theta)^{\alpha+n} \ \exp\left\{\left[\sum_{i=1}^{n} a(y_i)\right] \ b(\theta) \ + \ \beta \ b(\theta)\right\} d\theta}$$

$$= \ \frac{t(\theta)^{\alpha+n} \ \exp\left\{\left[\beta + \sum_{i=1}^{n} a(y_i)\right] \ b(\theta)\right\}}{\int_{\Theta} \ t(\theta)^{\alpha+n} \ \exp\left\{\left[\beta + \sum_{i=1}^{n} a(y_i)\right] \ b(\theta)\right\} d\theta}$$

$$= \ C_N^* \ \ t(\theta)^{\alpha+n} \ \exp\left\{\left[\beta + \sum_{i=1}^{n} a(y_i)\right] \ b(\theta)\right\},$$

sendo $1/C_N^* \ = \ \int_{\Theta} \ t(\theta)^{\alpha+n} \ \exp\left\{\left[\beta + \sum_{i=1}^{n} a(y_i)\right] \ b(\theta)\right\} \ d\theta.$

Para um modelo da família exponencial temos

- *a priori*: $f_\theta(\theta) \ = \ C_N \ \ t(\theta)^\alpha \ \exp\{\beta \ b(\theta)\}$.
- *a posteriori*: $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) = \ C_N^* \ \ t(\theta)^{\alpha+n} \ \exp\left\{\left[\beta + \sum_{i=1}^n a(y_i)\right] \ b(\theta)\right\}$

As duas distribuições acima estão na mesma família, portanto, há conjugação.

<u>Exemplo</u>: Seja $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ uma amostra aleatória da Bernoulli(θ). Sabe-se que esta distribuição é membro da família exponencial com:
$s(y_i) = \binom{1}{y_i}$, $\ t(\theta) = 1 - \theta$, $\ a(y_i) = y_i \ $ e $\ b(\theta) = \ln\left(\frac{\theta}{1-\theta}\right)$.

Reveja o exemplo com a Binomial alguns slides atrás.

Verossimilhança: $\quad f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \ = \ \prod_{i=1}^n s(y_i) \ t(\theta) \ \exp\{a(y_i) \ b(\theta)\}$

$= \ \prod_{i=1}^n \binom{1}{y_i} \ (1-\theta) \ \exp\left\{y_i \ \ln\left(\frac{\theta}{1-\theta}\right)\right\}$

Distribuição *a priori* conjugada: $f_\theta(\theta) \propto t(\theta)^\alpha \exp\{\beta\ b(\theta)\}$

Usando $t(\theta) = 1 - \theta$ e $b(\theta) = \ln\left(\frac{\theta}{1-\theta}\right)$ teremos

$$f_\theta(\theta) \propto (1-\theta)^\alpha \exp\left\{\beta \ln\left(\frac{\theta}{1-\theta}\right)\right\} \propto (1-\theta)^\alpha \exp\left\{\ln\left(\frac{\theta}{1-\theta}\right)^\beta\right\}$$

$$\propto (1-\theta)^\alpha \frac{\theta^\beta}{(1-\theta)^\beta} \propto \theta^\beta (1-\theta)^{\alpha-\beta}.$$

O núcleo acima indica que a distribuição Beta$(\beta+1, \alpha-\beta+1)$ é a candidata para obter conjugação.

Distribuição *a posteriori*:

$$f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \propto t(\theta)^{\alpha+n} \exp\left\{\left[\beta + \textstyle\sum_{i=1}^n a(y_i)\right] b(\theta)\right\},$$ lembrete: $a(y_i) = y_i$,

$$\propto (1-\theta)^{\alpha+n} \exp\left\{\left[\beta + \textstyle\sum_{i=1}^n y_i\right] \ln\left(\frac{\theta}{1-\theta}\right)\right\}$$

$$f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \;\propto\; (1-\theta)^{\alpha+n} \, \exp\left\{ \ln\left(\frac{\theta}{1-\theta}\right)^{\beta+\sum_{i=1}^n y_i} \right\}$$

$$\propto\; (1-\theta)^{\alpha+n} \, \frac{\theta^{\beta+\sum_{i=1}^n y_i}}{(1-\theta)^{\beta+\sum_{i=1}^n y_i}} \;\propto\; \theta^{\beta+\sum_{i=1}^n y_i} \, (1-\theta)^{\alpha+n-\beta-\sum_{i=1}^n y_i}.$$

O resultado final é o núcleo da Beta($\beta + \sum_{i=1}^n y_i$, $\alpha + n - \beta - \sum_{i=1}^n y_i$).

Exercícios:

- Assim como foi feito nos slides para o modelo Bernoulli(θ), escreva a distribuição *a priori* e *a posteriori* para os seguintes membros da família exponencial: (a) Poisson(θ) e (b) Gama(λ_1,λ_2) com $\theta = \{\lambda_1, \lambda_2\}$.
- Determine a formulação da distribuição preditiva $f_{\boldsymbol{Y}}(\boldsymbol{y})$ *a priori* para a família exponencial.

Conjugação fora da família exponencial.

A conjugação não é uma propriedade exclusiva das distribuições da família exponencial. O caso discutido a seguir envolve a $\text{U}(0,\theta)$, que não pertence a essa família.

Seja $\boldsymbol{Y} = \{Y_1, Y_2, \cdots, Y_n\}$ uma amostra aleatória tal que $Y_i|\theta \sim \text{U}(0,\theta)$ para $\theta > 0$. Adote independência condicional dos Y_i's dado θ.

$f_{Y_i|\theta}(y_i) \;=\; 1/\theta \quad \text{se} \quad y_i \in (0,\theta).$ A f.d.p. será 0 caso contrário.

Esta expressão pode ser rescrita usando a função indicadora $1_{\{0<y_i<\theta\}}$:

$$f_{Y_i|\theta}(y_i) \;=\; \frac{1}{\theta}\; 1_{\{0<y_i<\theta\}}$$

Verossimilhança: $\quad f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) \;=\; \prod_{i=1}^n \frac{1}{\theta}\; 1_{\{0<y_i<\theta\}} \;=\; \frac{1}{\theta^n} \prod_{i=1}^n 1_{\{0<y_i<\theta\}}.$

Perceba que $\prod_{i=1}^{n} 1_{\{0<y_i<\theta\}} = 1,$ se e somente se:

$0 < y_1 < \theta$ e $0 < y_2 < \theta$ e $0 < y_3 < \theta$ e $\cdots$ e $0 < y_n < \theta.$

Denote $y_{(n)} = \max\{y_1, y_2, \cdots, y_n\}$ e veja que se $y_{(n)} < \theta$ temos certamente que todos os y_i's são menores que θ.

Então $\prod_{i=1}^{n} 1_{\{0<y_i<\theta\}} = 1$ ocorre quando $1_{\{y_{(n)}<\theta\}} = 1_{\{\theta>y_{(n)}\}} = 1.$

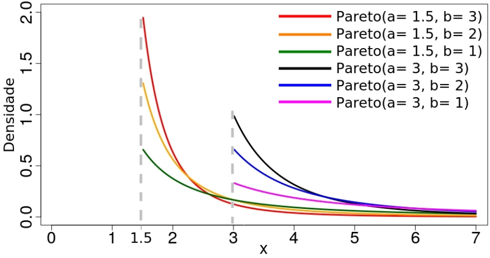
Isso permite escrever: $f_{\boldsymbol{Y}|\theta}(\boldsymbol{y}) = \dfrac{1}{\theta^n} 1_{\{\theta>y_{(n)}\}}$

Definição: Uma variável aleatória contínua X com distribuição Pareto(a,b), $a > 0$ é parâmetro de escala e $b > 0$ é parâmetro de forma, tem a seguinte f.d.p.

$$f_X(x) = \frac{b\ a^b}{x^{b+1}}\ 1_{\{x>a\}}.$$

Note a semelhança com o núcleo da verossimilhança se adotarmos *a priori*

$\theta \sim$ Pareto(a, b) com f.d.p. $f_\theta(\theta) = \dfrac{b\ a^b}{\theta^{b+1}}\ 1_{\{\theta>a\}}.$



Distribuição *a posteriori*: $f_{\theta|\boldsymbol{Y}}(\theta|\boldsymbol{y}) \propto \dfrac{1}{\theta^n} 1_{\{\theta > y_{(n)}\}} \dfrac{b\, a^b}{\theta^{b+1}} 1_{\{\theta > a\}}.$

Termos em vermelho são constantes (não dependem de θ) e serão incorporadas na constante de normalização.

$$\propto \frac{1}{\theta^{n+b+1}} 1_{\{\theta > y_{(n)}\}} 1_{\{\theta > a\}} \propto \frac{1}{\theta^{n+b+1}} 1_{\{\theta > \max\{y_1,\cdots,y_n,a\}\}}.$$

Este é o núcleo da Pareto(a^*,b^*) com $a^* = \max\{y_1, \cdots, y_n, a\}$ e $b^* = n + b$.

Conclusão: as distribuições *a priori* $\theta \sim$ Pareto(a, b) e *a posteriori* $\theta|\boldsymbol{Y} \sim$ Pareto(a^*, b^*) são da família Pareto.