Família de prioris conjugadas e use sequencial do teorema de Bayes

Prof. Caio Azevedo

Família Conjugada

- Sejam
  - $F = \{p(\boldsymbol{x}|\theta); \forall\ \theta\ \in \Theta\}$, uma família de modelos estatísticos (verossimilhança).
  - $p^* = \{p(\theta|\boldsymbol{\eta}) \equiv p(\theta); \forall\ \boldsymbol{\eta} \in \mathcal{B}\}$, uma família de prioris, em que $\boldsymbol{\eta}$ são os hiperparâmetros associados.
- Dizemos que a família p^* é conjugada à família F, se $\forall p(\boldsymbol{x}|\theta) \in F$ e $p(\theta) \in p^* \rightarrow p(\theta|\boldsymbol{x}) \in p^*$.
- Na prática, essencialmente, ao identificar a respectiva verossimilhança, em função de θ, como sendo correspondente ao núcleo de alguma família de distribuições, essa família será a respectiva conjugada natural.

Exemplo: Bernoulli

- Seja $X_i|\theta \overset{iid}{\sim}$ Bernoulli(θ). Temos que:

$$L(\theta) = \theta^{n\overline{x}}(1-\theta)^{n(1-\overline{x})}\mathbb{1}_{(0,1)}(\theta),$$

que corresponde ao núcleo de uma distribuição beta$(n\overline{x}+1, n(1-\overline{x})+1)$. Assim, a distribuição beta(a,b) é a família conjugada de prioris para o modelo Bernoulli. Obs: Se $X \sim$ beta(a,b) então $f_X(x) = \dfrac{1}{\beta(a,b)}x^{a-1}(1-x)^{b-1}\mathbb{1}_{(0,1)}(x)$

Exemplo: exponencial

- Seja $X_i|\theta \overset{iid}{\sim} \exp(\theta)$. Temos que:

$$L(\theta) = \theta^{-n} e^{-n\overline{x}/\theta} \mathbb{1}_{(0,\infty)}(\theta)$$

que corresponde ao núcleo de uma distribuição $\text{IG}(n-1, n\overline{x})$. Assim, a distribuição $\text{IG}(a,b)$ é a família conjugada de prioris para o modelo exponencial. Obs: Se $X \sim \text{IG}(a,b)$ então
$f_X(x) = \dfrac{b^a}{\Gamma(a)} x^{-(a+1)} e^{-b/x} \mathbb{1}_{(0,\infty)}(x)$

Exemplo: Poisson

- Seja $X_i|\theta \overset{iid}{\sim}$ Poisson(θ). Temos que:

$$L(\theta) = \frac{\theta^{n\overline{x}} e^{-n\theta}}{\prod_{i=1}^{n} x_i!} \mathbb{1}_{(0,\infty)}(\theta) \propto \theta^{n\overline{x}} e^{-n\theta} \mathbb{1}_{(0,\infty)}(\theta)$$

que corresponde ao núcleo de uma distribuição gama$(n\overline{x} + 1, 1/n)$. Assim, a distribuição gama(a, b) é a família de prioris conjugadas para o modelo Poisson. Obs: Se $X \sim$ gama(a, b) então
$f_X(x) = \dfrac{1}{b^a \Gamma(a)} x^{a-1} e^{-x/b} \mathbb{1}_{(0,\infty)}(x)$

Exemplo: Uniforme

- Seja $X_i|\theta \overset{iid}{\sim} \text{U}_{(0,\theta)}$. Temos que:

$$L(\theta) = \frac{1}{\theta^n}\prod_{i=1}^{n}\mathbb{1}_{(0,\theta)}(x_i)$$

Mas, note que ($\boldsymbol{x} = (x_1, .., x_n)'$):

$$\begin{aligned}
\prod_{i=1}^{n}\mathbb{1}_{(0,\theta)}(x_i) = 1 \quad & \leftrightarrow \quad 0 < x_i < \theta, \forall i \leftrightarrow 0 < \min(\boldsymbol{x}) < \max(\boldsymbol{x}) < \theta \\
& \leftrightarrow \quad 0 < y_1 < y_n < \theta \\
& \leftrightarrow \quad \mathbb{1}_{(0,\theta)}(y_n)\mathbb{1}_{(0,y_n)}(y_1) = \mathbb{1}_{(y_1,\theta)}(y_n)\mathbb{1}_{(0,\theta)}(y_1) = 1
\end{aligned}$$

Exemplo: Uniforme

- Logo, temos que:

$$L(\theta) = \frac{1}{\theta^n}\mathbb{1}_{(0,\theta)}(y_n)\mathbb{1}_{(0,y_n)}(y_1) \propto \frac{1}{\theta^n}\mathbb{1}_{(y_n,\infty)}(\theta),$$

que corresponde ao núcleo de uma distribuição Pareto$(n-1, y_n)$.

Obs: Se $X \sim$ Pareto(a,b) então $f_X(x) = \dfrac{ab^a}{x^{a+1}}\mathbb{1}_{[b,\infty)}(x)$

Exemplo: Normal

- Seja $X_i|\sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$, ($\mu$ conhecido). Temos que

$$\begin{aligned} L(\sigma^2) &= \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2} \mathbb{1}_{(0,\infty)}(\sigma^2) \\ &\propto \left(\sigma^2\right)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2} \mathbb{1}_{(0,\infty)}(\sigma^2) \end{aligned}$$

que corresponde ao núcleo de uma distribuição $\text{IG}(n/2-1, \sum_{i=1}^n (x_i-\mu)^2/2)$. Assim, a distribuição $\text{IG}(a,b)$ é a família conjugada de prioris para o modelo normal com a média conhecida. Obs: Se $X \sim \text{IG}(a,b)$ então

$$f_X(x) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} e^{-b/x} \mathbb{1}_{(0,\infty)}(x)$$

Exemplo: Normal

- Seja $X_i|\mu \overset{iid}{\sim} N(\mu, \sigma^2)$, ($\sigma^2$ conhecido). Temos que

$$
\begin{aligned}
L(\mu) &= \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \mathbb{1}_{\mathcal{R}}(\mu) \\
&\propto e^{-\frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - 2\mu\sum_{i=1}^{n} x_i + n\mu^2\right)} \mathbb{1}_{\mathcal{R}}(\mu) \\
&\propto e^{-\frac{n}{\sigma^2}\left(\mu^2 - 2\mu\overline{x}\right)} \mathbb{1}_{\mathcal{R}}(\mu) \propto e^{-\frac{n}{\sigma^2}(\mu-\overline{x})^2} \mathbb{1}_{\mathcal{R}}(\mu)
\end{aligned}
$$

que corresponde ao núcleo de uma distribuição $\mathrm{N}(\overline{x}, \sigma^2/n)$. Assim, a distribuição $\mathrm{N}(a, b^2)$ é a família conjugada de prioris para o modelo normal com a variância conhecida. Obs: Se $X \sim \mathrm{N}(a, b^2)$ então $f_X(x) = \frac{1}{\sqrt{2\pi b^2}} e^{-\frac{1}{2b^2}(x-a)^2} \mathbb{1}_{(0,\infty)}(x)$

Família Conjugada na família exponencial

- Dizemos que uma variável aleatória $X|\boldsymbol{\theta}$ pertence à família exponencial (k paramétrica) se sua distribuição puder ser escrita como

$$p(x|\boldsymbol{\theta}) = \exp\left\{\sum_{j=1}^{k} c_j(\boldsymbol{\theta})t_j(x) + d(\boldsymbol{\theta})\right\} a(x)\mathbb{1}_A(x)$$

em que A não depende de θ.

- Seja $X_i|\boldsymbol{\theta} \overset{i.i.d}{\sim} X|\boldsymbol{\theta}, i = 1, ..., n$, então

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \exp\left\{\sum_{j=1}^{k} c_j(\boldsymbol{\theta})\sum_{i=1}^{n} t_j(x_i) + nd(\boldsymbol{\theta})\right\} \prod_{i=1}^{n} a(x_i)\mathbb{1}_A(\boldsymbol{x})$$

Família Conjugada na família exponencial (cont.)

- Neste caso, a família conjugada de prioris é dada por

$$p(\theta) = \exp\left\{\sum_{j=1}^{k} c_j(\boldsymbol{\theta})\alpha_j + \beta d(\boldsymbol{\theta})\right\} k(\beta, \boldsymbol{\alpha})\mathbb{1}_{\Theta}(\boldsymbol{\theta})$$

em que $k(\beta, \boldsymbol{\alpha})$ é uma constante de normalização apropriada e $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)'$.

Família Conjugada na família exponencial (cont.)

- Consequentemente, a posteriori será dada por (exercício):

$$
\begin{aligned}
p(\theta|\boldsymbol{x}) &= \exp\left\{\sum_{j=1}^{k} c_j(\boldsymbol{\theta})\left[\alpha_j + \sum_{i=1}^{n} t_j(x_i)\right] + (\beta + n)d(\boldsymbol{\theta})\right\} \times \\
&\times \quad k(\beta + n, \boldsymbol{\alpha}, \boldsymbol{t})\mathbb{1}_{\Theta}(\boldsymbol{\theta})
\end{aligned}
$$

em que $k(\beta + n, \boldsymbol{\alpha}, \boldsymbol{t})$ é uma constante de normalização apropriada e $\boldsymbol{t} = (t_1, ..., t_n)'$.

Uso sequencial do teorema de Bayes

- Sejam $\boldsymbol{x}_1 = (x_1, ..., x_n)'$ e $\boldsymbol{x}_2 = (x_{n+1}, ..., x_{n+m})'$ amostras de $X|\theta$, tomadas sequencialmente e defina $\boldsymbol{x}^* = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$.
- Dessa forma, temos que

$$p(\theta|\boldsymbol{x}_1) = \frac{p(\boldsymbol{x}_1|\boldsymbol{\theta})p(\theta)}{p(\boldsymbol{x}_1)}, p(\boldsymbol{x}_1) = \int_{\Theta} p(\boldsymbol{x}_1|\boldsymbol{\theta})p(\theta)d\theta$$

- Por outro lado, temos que

$$p(\theta|\boldsymbol{x}^*) = \frac{p(\boldsymbol{x}^*|\boldsymbol{\theta})p(\theta)}{p(\boldsymbol{x}^*)}, p(\boldsymbol{x}^*) = \int_{\Theta} p(\boldsymbol{x}^*|\boldsymbol{\theta})p(\theta)d\theta$$

Uso sequecial do teorema de Bayes (cont.)

- Agora, note que $p(\boldsymbol{x}^*|\theta) = p((\boldsymbol{x}_1', \boldsymbol{x}_2')'|\theta) = p(\boldsymbol{x}_2|\theta, \boldsymbol{x}_1)p(\boldsymbol{x}_1|\theta)$. Logo, vem que

$$
\begin{aligned}
p(\theta|\boldsymbol{x}^*) &= \frac{p(\boldsymbol{x}_2|\theta, \boldsymbol{x}_1)p(\boldsymbol{x}_1|\theta)p(\theta)}{p(\boldsymbol{x}^*)} = \frac{p(\boldsymbol{x}_2|\theta, \boldsymbol{x}_1)}{\frac{p(\boldsymbol{x}^*)}{p(\boldsymbol{x}_1)}} \frac{p(\boldsymbol{x}_1|\theta)p(\theta)}{p(\boldsymbol{x}_1)} \\
&= \frac{p(\boldsymbol{x}_2|\theta, \boldsymbol{x}_1)}{\int_\Theta p(\boldsymbol{x}_2|\theta, \boldsymbol{x}_1)\frac{p(\boldsymbol{x}_1|\theta)p(\theta)}{p(\boldsymbol{x}_1)} d\theta} p(\theta|\boldsymbol{x}_1) \\
&= \frac{p(\boldsymbol{x}_2|\theta, \boldsymbol{x}_1)}{\int_\Theta p(\boldsymbol{x}_2|\theta, \boldsymbol{x}_1)p(\theta|\boldsymbol{x}_1) d\theta} p(\theta|\boldsymbol{x}_1) = \frac{p(\boldsymbol{x}_2|\theta, \boldsymbol{x}_1)p(\theta|\boldsymbol{x}_1)}{p(\boldsymbol{x}_2)}
\end{aligned}
$$