

# ESTATÍSTICA COMPUTACIONAL E SIMULAÇÃO

Projeto 3: Estimação(M.V. e bayesiana-métodos MCMC)

Ano letivo 2022/23

1. McCormick e Mathew(1983) examinaram questões de estimação relacionadas com o modelo

$$X_t = \gamma + \rho X_{t-1} + Y_t, t \geq 1$$

onde  $\gamma \geq 0$ ,  $0 \leq \rho < 1$  são parâmetros desconhecidos e  $Y_t$  são variáveis aleatórias i.i.d. A análise bayesiana deste modelo quando os  $Y_t$  são exponenciais com valor médio  $\alpha^{-1}$ ,  $\alpha > 0$  foi considerada por Pereira e Amaral Turkman(1983). Considere as seguintes distribuições condicionais completas:

- $h(\alpha|x, \gamma, \rho) \sim Ga(n, S_0 - \rho S_1 - (n-1)\gamma)$ .
- $h(\gamma|x, \alpha, \rho) \sim Exp_{esq}((n-1)\alpha, \gamma^*)$ , onde  $\gamma^* = \min(x_t - \rho x_{t-1})$ .
- $h(\rho|x, \alpha, \gamma) \sim Exp_{esq}(\alpha S_1, \rho^*)$ , onde  $\rho^* = \min(1, \frac{x_t - \gamma}{x_{t-1}})$ .

onde  $S_0 = \sum_{t=2}^n x_t$ ,  $S_1 = \sum_{t=2}^n x_{t-1}$  e  $f(x) \sim Exp_{esq}(\beta, \delta)$  significa  $f(x) = \frac{\beta e^{-\beta(\delta-x)}}{1-e^{-\beta\delta}}$ , para  $0 < x < \delta$ . Os seguintes dados representam valores dos teores de oxigénio dissolvido medidos na ponte de Angeja de Junho a Novembro de 1991:

4.0	4.1	3.9	4.4	3.2	4.0	3.7	4.2	4.5	4.3	3.6	1.9
3.3	1.9	2.9	2.7	2.4	2.9	3.8	3.5	2.7	3.9	2.8	3.3
2.9	3.8	4.4	5.1	5.2	7.2	6.2	4.8	4.0	2.7	4.4	3.4
4.2	4.8	5.3	4.5	4.1	4.0	2.9	0.8	5.2	7.3	5.1	5.3
7.1	8.1	7.8	6.9	7.5	6.0	5.0	5.3	4.8	4.3	5.8	4.6
4.5	4.1	4.6	6.4	6.3	6.2	6.2	6.8	7.5	7.4	7.0	6.7
7.5	6.1	5.7	5.4	5.3	4.0	3.7	2.5	0.8	1.3	3.3	4.1
5.7	4.3	3.5	3.8	2.0	3.8	4.1	1.8	3.0	4.7	6.2	6.0
5.3	4.4	3.4	4.7	4.5	3.7	4.3	1.6	2.9	3.6	3.7	3.9
4.6	5.0	5.3	4.7	6.5	5.7	5.8	8.0	7.4	6.1	7.6	

Usando esta amostra, implemente o algoritmo Gibbs, usando o pacote CODA para analisar a convergencia, com o objetivo de estimar os parâmetros e caracterizar aproximadamente as distribuições de  $\alpha|x$ ,  $\gamma|x$  e de  $\rho|x$ .

2. Considere os dados incluídos no ficheiro caudais.txt, referentes aos caudais máximos do rio Ocmulgee, registados na estação hidrológica de Hawkinsville, nos anos 1949 a 1984. Considere-se que o modelo que se ajusta é o da distribuição Gumbel,

$$f(x) = \frac{1}{\sigma} \exp\left\{-\frac{x-\mu}{\sigma} - \exp\left[-\frac{x-\mu}{\sigma}\right]\right\}, x \in R, \sigma > 0, \mu \in \mathbf{R}.$$

- (a) Mostre que a função de logverosimilhança é dada por

$$L^* = -n \log \sigma - \sum_{i=1}^n \left( \frac{x_i - \mu}{\sigma} + \exp\left\{-\frac{x_i - \mu}{\sigma}\right\} \right), \sigma > 0, \mu \in \mathbf{R}.$$

- (b) Após obter as equações normais para o cálculo das estimativas de m.v., mostre que da equação  $\frac{\partial}{\partial \mu} L^* = 0$  se obtém a relação  $\mu = \sigma \log\left(\frac{n}{\sum_{i=1}^n \exp\left(-\frac{x_i}{\sigma}\right)}\right)$ .
- (c) Usando funções apropriadas do R, e comentando a sua escolha, determine estimativas de máxima verosimilhança para os parâmetros; usando, se possível uniroot e optimize, e a

função optim.

OBS: no caso de precisar de estimativas iniciais, pode usar as que se obtêm pelo método dos momentos, dadas por  $\hat{\sigma}_{m.m.} = \frac{\sqrt{6}}{\pi}s_x$  e  $\hat{\mu}_{m.m.} = \bar{x} - 0.5772\sigma_{m.m.}$ .

- (d) Considere o caso particular da distribuição Gumbel referida anteriormente, considerando  $\sigma$  conhecido e dado por  $\hat{\sigma}_{m.v.}$ . Utilizando os mesmos dados pretende-se estimar  $\mu$  usando a metodologia bayesiana; para isso supõe-se que  $\mu$  é também variável aleatória e vai-se admitir que  $\mu \sim U(0, 100)$ . Gere valores da distribuição condicional completa, usando o algoritmo de Metropolis-Hastings e a distribuição do qui-quadrado com  $|X_t|$  graus de liberdade para proponente para  $X_{t+1}$ . Considerando período de aquecimento, espaçamento entre observações para limitar a correlação, analise a convergência da cadeia de Markov associada. Compare com o valor correspondente dado pela estimativa de máxima verosimilhança.

OBS: Pretende gerar valores da distribuição condicional completa de  $\mu$ , dada por  $f_{\mu|\mathbf{x}}(\mu) = \pi(\mu)f_{\mathbf{X}|\mu}(\mathbf{x})$ , com  $\pi(\mu)$  a representar a distribuição *a priori* e  $f_{\mathbf{X}|\mu}(\mathbf{x})$  a expressão da função verosimilhança da amostra  $\mathbf{x}$ .