Universidade de Aveiro - Estatística Computacional e Simulação

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

Diogo Pedrosa(94358), Rita Ferrolho(88822)

Ano Letivo 2021/2022

1 EXERCÍCIO 1

Neste exercício pretende-se usar os dados contidos em caudais.txt, referentes aos caudais máximos do rio Ocmulgee, registados na estação hidrolóigca de Hawkinsville, entre 1949 e 1984, para encontrar a estimativa de máxima verosimilhança para os parâmetros μ e σ. Sabe-se que o modelo que se ajusta aos dados é a distribuição de Gumbel, com função de densidade de probabilidade dada por:

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

Com este intuito, será usado o método de Estimação de Máxima Verosimilhança, um método que consiste no cálculo do valor de um parâmetro que maximiza a verosimilhança, para o seguinte tipo de distribuição:

$$f(x|\theta), \theta \in \theta \subseteq R^k \tag{2}$$

Numa amostra aleatória, $X = (X_1, X_2, ...X_n)$, proveniente da população $f(x|\theta)$, a função de verosimilhança é dada por:

$$L(x|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = f(x|\theta) \tag{3}$$

Isto implica que, dada uma amostra observada, $x = (x_1, x_2, ...x_n)$, a estimativa de máxima verosimilhança de um parâmetro θ é definida por:

$$\hat{\theta} = T(x) = argmax(L(\theta|x)) \tag{4}$$

Assumindo que a função logarítmica é monótona, a expressão anterior escreve-se da seguinte forma:

$$\hat{\theta} = T(x) = argmax(log(L(\theta|x))) \tag{5}$$

O valor máximo é otido como a solução da equação de verosimilhança, sob as condições de regularidade, isto é, para $\frac{d^2}{d\theta^2}log(L(\theta|x)) < 0$:

$$\frac{d}{d\theta}logL((\theta|x)) = 0 \tag{6}$$

Assim, deve-se descobrir as equações de verosimilhança para determinar a solução por implementação de código.

Seja $X = (X_1, X_2, ...X_n)$ uma amostra aleatória proveniente da população com distribuição Gumbel, cuja função de densidade de probabilidade foi previamente demonstrada, é necessário identificar os parâmetros σ e μ, do problema bidimensional $\theta = (\sigma, \mu)$.

Com este objetivo, começa-se por calcular a função de Verosimilhança:

$$L(x|\sigma, \mu) =$$
$$= \prod_{i=1}^{n} f(x_i|\sigma, \mu) =$$
$$= \prod_{i=1}^{n} \frac{1}{\sigma} exp\{-\frac{x-\mu}{\sigma} - exp[\frac{x-\mu}{\sigma}]\} =$$

$$= \frac{1}{\sigma^n} exp\{\sum_{i=1}^{n}(-\frac{x_i-\mu}{\sigma} - exp\{\sum_{i=1}^{n} -\frac{x_i-\mu}{\sigma}\})\} \tag{7}$$

A seguir, calcula-se a função Log-Verosimilhança, aplicando o logaritmo à função de verosimilhança:

$$log(L(\theta)) =$$
$$= log(L(x|\sigma, \mu)) =$$

$$= -n \times log(\sigma) - \sum_{i=1}^{n} \frac{x_i-\mu}{\sigma} - \sum_{i=1}^{n} exp\{-\frac{x_i-\mu}{\sigma}\} \tag{8}$$

Calculando as equações normais para o cálculo das estimativas de máxima verosimilhança, por derivação parcial da função Log-Verosimilhança em relação a σ e em relação μ:

$$\begin{cases} \frac{\partial}{\partial\sigma} log(L(x|\sigma, \mu)) = 0 \\ \frac{\partial}{\partial\mu} log(L(x|\sigma, \mu)) = 0 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} -\frac{n}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n}((x_i-\mu)(1 - exp\{-\frac{x_i-\mu}{\sigma}\})) = 0 \\ \frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} exp\{\frac{x_i-\mu}{\sigma}\} = 0 \end{cases} \tag{9}$$

Como a resolução analítica das equações é um processo demorado, Não existe uma solução analítica explicita para a estimativa de método de verosimilhança, $\hat{\sigma}_{m.v}$. Portanto, é preciso de recorrer a métodos computacionais. Para isso usou-se o R Studio, mais concretamente, as funções pré-definidas optim, optimize e uniroot. Caso seja possível escrever μ em função de σ, as funções optim, optimize e uniroot podem ser usadas para resolver as equações normais. Caso não haja essa possibilidade, apenas a função optim pode ser usada.

Como foi possível obter a função μ em função de σ, então as funções optimize e uniroot podem ser usadas, além de optim.

Conseguiu-se determinar a função μ em função de σ:

$$\mu_{m.v} = \sigma log(\frac{n}{\sum_{i=1}^{n} exp\{-\frac{x_i}{\sigma}\}}) \tag{10}$$

Como foi possível obter a função (10), então as funções optimize e uniroot podem ser utilizadas, para além da função optim.

Assumindo que a estimativa de máxima verosimilhança de μ é desconhecida, o problema é classificado como um problema bidimensional, em que se pretende obter as estimativas de μ e σ. Para o resolver computacionalmente, pode-se utilizar a função optim. Depois de consultar a sua documentação, por execução do comando "?optim" na consola do R Studio, utilizaram-se os seguintes argumentos:

optim(par, fn, upper, maximum)

- par - Valores iniciais dos parâmetros para serem otimizados. Neste parâmetro atribuiu-se o vetor das estimativas iniciais de μ e σ;
- fn - Função de maximização ou de minimização. Por defeito, é realizada a minimização, portanto, como se pretende maximizar, atribuiu-se a este argumento o simétrico da função Log-Semelhança (10), $-log(L(x|\sigma, \mu))$;
- N - Neste parâmetro usou-se como valor o comprimento do vetor de dados do ficheiro caudais.txt.

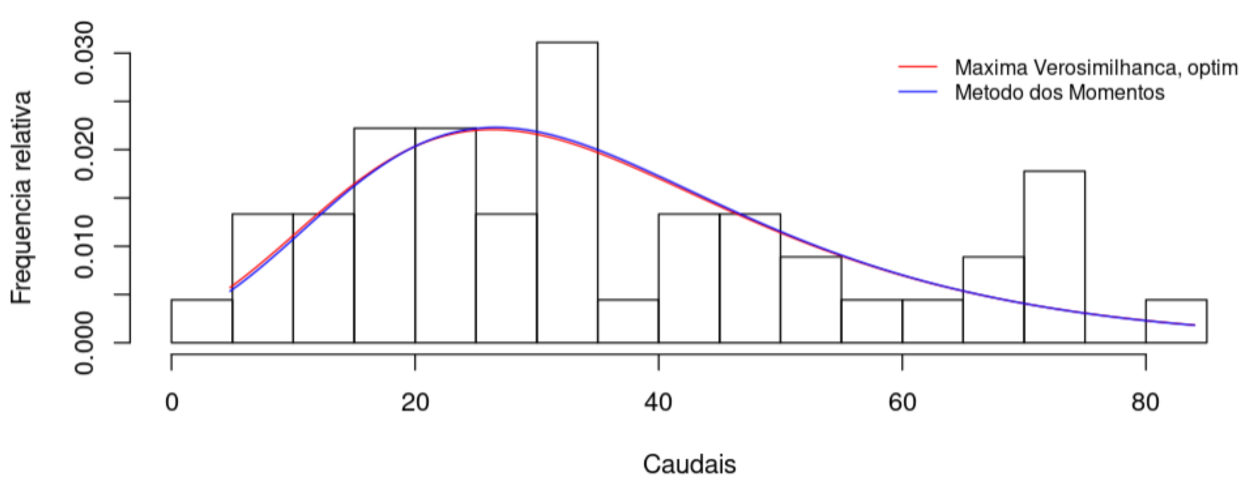


Figure 1: Histograma da amostra. A linha azul representa a função correspondente à distribuição de Gumbel, com parâmetros obtidos pelo método dos momentos. A linha vermelha representa a mesma função, mas obtida com a função optim.

Uma vez que a estimativa de máxima verosimilhança de μ foi obtida (10), é possível interpretar o problema como um problema de otimização unidimensional, em que apenas o valor de σ é estimado.

Uma das funções que pode ser usada para resolver este problema é optimize. Após a consulta da documentação desta função, utilizaram-se os seguintes argumentos:

optimize(f, lower, upper, maximum)

- f - A função a ser otimizada. Neste argumento foi utilizada a função de Log Verosimilhança (8);
- lower - o valor mínimo do intervalo a ser procurado. Neste argumento atribuiu-se o valor 0;
- upper - o valor máximo do intervalo a ser procurado. Neste argumento atribuiu-se o valor 100;
- maximum - valor booleano que indica se o problema é de maximização (True ou T) ou se é de minimização (False ou F). Por omissão, se nenhum valor for atribuído, assume-se que o problema é de minimização. Como se pretende resolver um problema de maximização, utilizou-se o valor T para este argumento.

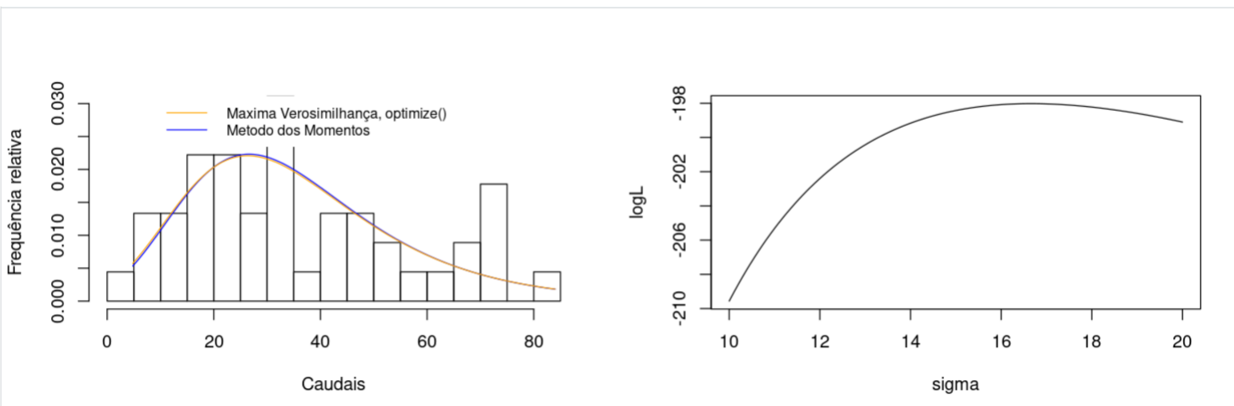
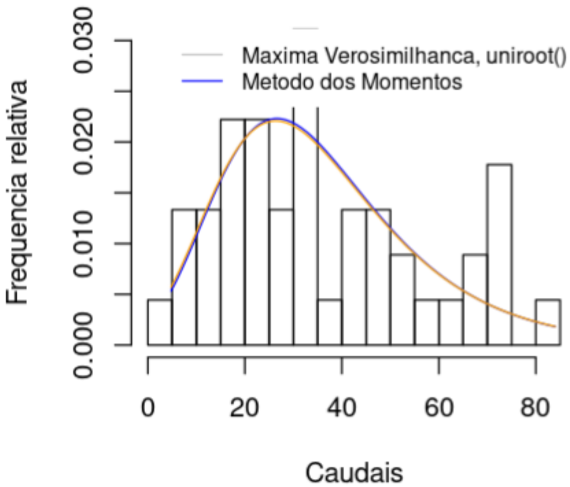


Figure 2: Gráfico da esquerda: Histograma da amostra. A linha azul representa a função da distribuição de Gumbel com parâmetros obtidos pelo método dos momentos. A linha laranja representa a mesma função, mas obtida com optimize. Gráfico da direita: Função a ser maximizada.

Em alternativa ao comando otimize, pode-se usar a função uniroot. Uma vez realizada a consulta da sua documentação, utilizaram-se os seguintes argumentos:

uniroot(f, lower, upper)

- f - A função para qual se procura a raíz. A equação de verosimilhança derivada em σ, correspondente à segunda equação de (9), foi introduzida neste argumento;
- lower - o valor mínimo do intervalo a ser procurado. Neste argumento atribuiu-se o valor 1;
- upper - o valor máximo do intervalo a ser procurado. Neste argumento atribuiu-se o valor 100.



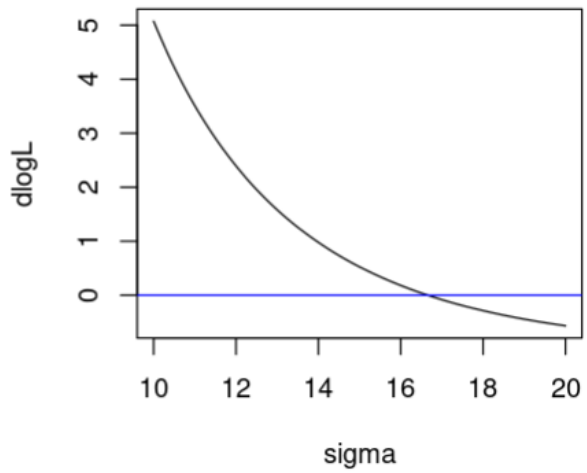


Figure 3: Gráfico da esquerda: Histograma da amostra. A linha azul representa a função da distribuição de Gumbel com parâmetros obtidos pelo método dos momentos. A linha laranja representa a mesma função, mas obtida com uniroot. Gráfico da direita: Função com raiz a ser determinada.

De seguida é mostrada uma tabela com as estimativas de μ e σ, para o Método dos Momentos e para o Método de Máxima Verosimilhança (com funções optim, optimize e uniroot):

Estimativas			
Método dos Momentos	optim	optimize	uniroot
$\hat{\mu}_{m.m} = 26.61160$	$\hat{\mu}_{optim} = 26.33968$	$\hat{\mu}_{optimize} = 26.34149$	$\hat{\mu}_{uniroot} = 26.34149$
$\hat{\sigma}_{m.m} = 16.48871$	$\hat{\sigma}_{optim} = 16.65487$	$\hat{\sigma}_{optimize} = 16.65789$	$\hat{\sigma}_{uniroot} = 16.65789$

Calculou-se ainda o erro relativo para cada uma destas estimativas, por comparação das estimativas obtidas com o Método de Máxima Verosimilhança com as estimativas obtidas no Método dos Momentos:

Erros relativos calculados para o Método de Máxima Verosimilhança		
optim	optimize	uniroot
1.021833	1.015022	1.015018
1.007686	1.025998	1.026012

Ao calcular a média dos valores de $\hat{\mu}$ e $\hat{\sigma}$ calculados pelo Método de Máxiama Verosimilhança, os valores finais estimados são os seguintes:

$$\hat{\mu} = 26.34089$$

$$\hat{\sigma} = 16.65688$$

2 EXERCÍCIO 2

Considera-se que os dados incluídos em caudais.txt se ajustam ao modelo da distribuição Gumbel, cuja função de densidade de probabilidade é dada pela equação (1). Com esta distribuição, pretende-se estimar o parâmetro μ, usando para isso a metodologia bayesiana aliada ao algoritmo de Metropolis-Hastings.

A distribuição condicional completa é a seguinte, em que $\int f_\mu(\mu) f_{\tilde{X}|\mu}(\tilde{x}) d\mu$ é a constante de normalização:

$$f_{\mu|\tilde{x}}(\mu) =$$

$$= \frac{f_{\mu,\tilde{x}}(\mu, \tilde{x})}{f_{\tilde{X}(\tilde{x})}}$$

$$= \frac{f_\mu(\mu) f_{\tilde{X}|\mu}(\tilde{x})}{\int f_\mu(\mu) f_{\tilde{X}|\mu}(\tilde{x}) d\mu} \propto f_\mu(\mu) f_{\tilde{X}|\mu}(\tilde{x}) \tag{11}$$

A distribuição conjunta consiste no produto da distribuição a priori, $f_\mu(\mu)$, com a função de verosimilhança, $f_{\tilde{X}|\mu}(\tilde{x})$:

$$f_\mu \tilde{x}(\mu) = f_\mu(\mu) \times f_{\tilde{X}|\mu}(\tilde{x}) \tag{12}$$

Admitindo que $\mu \sim U(100, 100)$, a distribuição a priori é a seguinte:

$$f_\mu(\mu) = \frac{1}{100 - (-100)} = 1/200 \tag{13}$$

Relembra-se que a equação de verosimilhança corresponde à equação (7).

Desenvolvendo a equação da distribuição conjunta, para $\hat{\sigma}_{m.v} = 16.66?$, calculado na alínea anterior:

$$f_\mu \tilde{x}(\mu) =$$

$$= f_\mu(\mu) \times f_{\tilde{X}|\mu}(\tilde{x}) =$$

$$= \frac{1}{200} \times \frac{1}{16.66?^n} exp\{\sum -\frac{x_i - \mu}{16.66?} - \sum exp\{-\frac{x_i - \mu}{16.66?}\}\} \propto exp\{\sum -\frac{x_i - \mu}{16.66?} - \sum exp\{-\frac{x_i - \mu}{16.66?}\}\} \tag{14}$$

De seguida, gerou-se valores da distribuição condicional completa encontrada, usando o algoritmo de Metropolis- Hastings e a distribuição qui-quadrado com $|X_t|$ graus de liberdade para proponente para X_{t+1}. Implementou-se o seguinte algoritmo:

1 - Considera-se x_0 tal que f(x_0) > 0 e guarda-se em xt[1] = x_0. Foi através da função candidata que se gerou o x_0.
2 - Repete-se para t = 1 (ei = 2,..., T), onde T é o tamanho da cadeia de Markov.

- Gera-se y da distribuição proponente para o parâmetro μ, q(y—x_t) X^2($Qui-QuadradocomX_t$ graus de liberdade);

- Gera-se NPA U U(0,1);
- Calcula-se α (x_t,Y), com os dados observados, com f a função distribuição condicional completa e q a fdp de X^2x$_t; SeU < \alpha(x_t$,Y) aceitar y e x_{t+1} = y;
- Caso contrário, $x_{t+1} = x_t$;
- Guarda-se x_{t+1} em x[i] = x_{t+1};
- Incrementa-se e voltar ao passo 2.

Abaixo apresentamos a cadeia de Markov e o Histograma da cadeia de Markov com a densidade do nucleo.

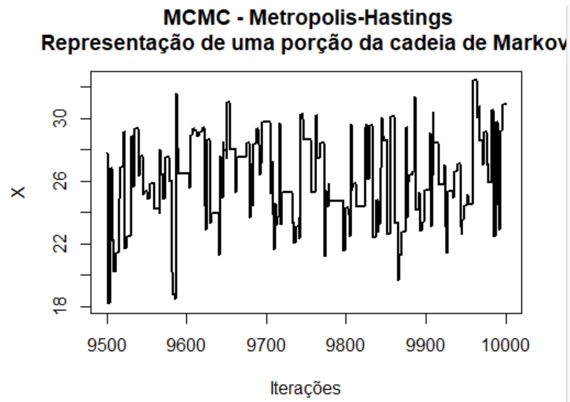


Figure 4: Representação de uma cadeia de Markov

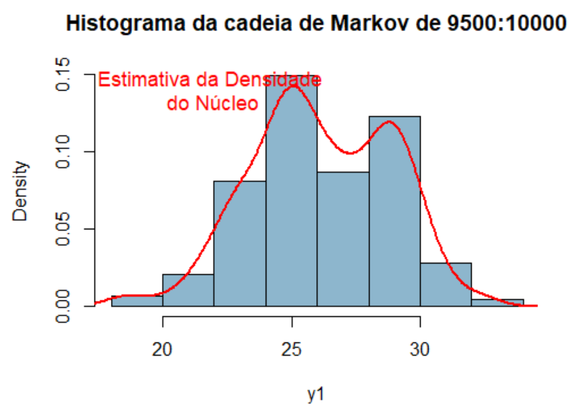


Figure 5: Histograma da cadeia de markov e da densidade do nuc2leo(Linha vermelha)

A estimativa de μ foi de 26,09. Obtivemos este resultado fazendo a média da cadeia de Markov gerada na figura 4. A figura 5 confirma-nos a estimativa, pois a densidade do nucleo acompanha, de certa forma, o formato do histograma. No entanto, o valor que obtivemos para a taxa de rejeição foi de 60,95% o que é uma elevada taxa de rejeição, o que pode indicar que o método utilizado não foi eficiente.

No final foi se analisar a convergência da cadeia de Markov. Primeiro realizou-se o diagnóstico de Geweke que permite investigar se duas partes não-sobrepostas da cadeia são da mesma distribuição, comparando as médias de ambas as partes. Assim, são testadas as seguintes hipóteses para as partes que constituem os primeiros 10% e nos ultimos 50% da cadeia: H0:$\mu_{10\%} = \mu_{50\%}$ vs H1:$\mu_{10\%} \neq \mu_{50\%}$

Foi obtido um Z_{score} = -0.07952 como temos que | $T_{observado}$ | < 1,96 a hipótese de que as duas partes da cadeia pertencem à mesma distribuição não é rejeitada, o que indicia que o tamanho da cadeia de Markov é o suficientemente grande para ocorrer convergência dos valores de µ gerados.

Depois realizou-se o diganóstico de Gelman-Rumin(figura 6)

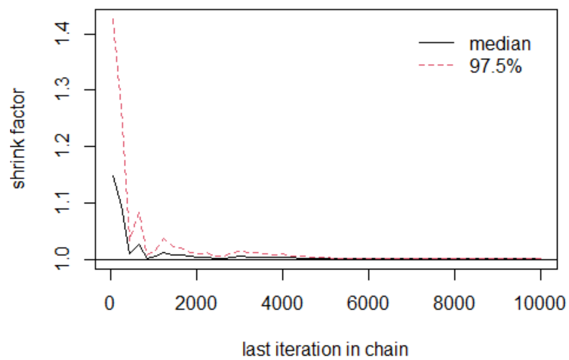


Figure 6: Diagnóstico Gelman-Rumin

O diagnóstico de Gelman-Rubin sugere a convergência para a distribuição estacionária, a figura 3 confirma-nos isso. Isto tudo sugere uma convergência da cadeia.
Para analisar a simetria da densidade da cadeia de Markov tem-se a figura 7 que sugere que os valores de µ têm uma distribuição ligeiramente assimétrica

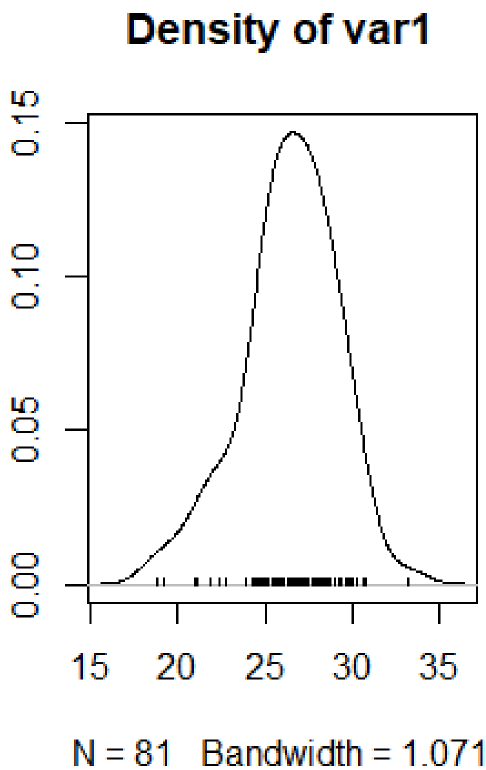


Figure 7: Gráfico da densidade dos valores da cadeia de Markov

Para concluir, os resultados obtidos pela metodologia bayesiana foram muito semelhantes aos resultados obtidos pela metodologia da máxima verosomilhança pelo que pode-se dizer que foram bons resultados.