

UNIVERSIDADE FEDERAL DO PARANÁ

Matheus Henrique Tavares

Comparação entre o Método dos
Momentos-L e métodos estatísticos
clássicos de ajuste de distribuições a dados
hidrológicos

Curitiba
2015

Matheus Henrique Tavares

Comparação entre o Método dos
Momentos-L e métodos estatísticos
clássicos de ajuste de distribuições a dados
hidrológicos

Trabalho apresentado como requisito parcial à obtenção do grau de bacharel em Engenharia Ambiental, Setor de Tecnologia da Universidade Federal do Paraná.

Orientador: Paulo Justiniano Ribeiro Jr.

Co-orientador: Maurício Felga Gobbi

Curitiba
2015

Lista de Figuras

| | | |
|------|--|----|
| 5.1 | Densidade empírica da amostra aleatória gerada pela distribuição gama. . . | 11 |
| 5.2 | Função distribuição acumulada empírica da amostra aleatória gerada pela distribuição gama. | 11 |
| 5.3 | Comparativo das distribuições gama teóricas dos parâmetros estimados pelo Método dos Momentos (MM), Método da Máxima Verossimilhança (MMV) e Método dos Momentos-L (MML) com a amostra aleatória e sua distribuição teórica. | 12 |
| 5.4 | Densidade empírica da amostra aleatória gerada pela distribuição de Weibull. 14 | |
| 5.5 | Função distribuição acumulada empírica da amostra aleatória gerada pela distribuição de Weibull. | 15 |
| 5.6 | Comparativo das distribuições de Weibull teóricas dos parâmetros estimados pelo Método dos Momentos (MM), Método da Máxima Verossimilhança (MMV) e Método dos Momentos-L (MML) com a amostra aleatória e sua distribuição teórica. | 16 |
| 5.7 | Série histórica dos dados mensais de cota máxima da estação fluviométrica Barra do Mandacaia, em Telêmaco Borba (PR). | 18 |
| 5.8 | Histograma dos dados mensais de cota máxima da estação fluviométrica Barra do Mandacaia, em Telêmaco Borba (PR). | 18 |
| 5.9 | Densidade empírica dos dados de cota máxima da estação fluviométrica Barra do Mandacaia, em Telêmaco Borba (PR). | 19 |
| 5.10 | Comparativo das distribuições gama teóricas dos parâmetros estimados pelo Método dos Momentos (MM), Método da Máxima Verossimilhança (MMV) e Método dos Momentos-L (MML) com os dados de cota máxima. | 19 |
| 5.11 | Série histórica dos dados de vazão da estação fluviométrica São Francisco, em Corumbá (MS). | 20 |
| 5.12 | Histograma dos dados de vazão máxima da estação fluviométrica São Francisco, em Corumbá (MS). | 20 |
| 5.13 | Densidade dos dados de vazão máxima da estação fluviométrica São Francisco, em Corumbá (MS). | 21 |
| 5.14 | Comparativo das distribuições de Weibull teóricas dos parâmetros estimados pelo Método da Máxima Verossimilhança (MMV) e Método dos Momentos-L (MML) com os dados de vazão máxima. | 22 |

Sumário

| | | |
|----------|--|-----------|
| 1 | Introdução | 1 |
| 2 | Objetivos | 3 |
| 2.1 | Objetivo Geral | 3 |
| 2.2 | Objetivos específicos | 3 |
| 3 | Revisão Bibliográfica | 4 |
| 3.1 | Método dos Momentos | 4 |
| 3.2 | Método da Máxima Verossimilhança | 5 |
| 3.3 | Método dos Momentos-L | 6 |
| 4 | Metodologia | 8 |
| 5 | Análises Preliminares | 9 |
| 5.1 | Simulações | 9 |
| 5.1.1 | Caso 1: Dados simulados da distribuição gama | 9 |
| 5.1.2 | Caso 2: Dados simulados da distribuição de Weibull | 13 |
| 5.2 | Dados reais | 17 |
| 5.2.1 | Caso 3: Dados de cota máxima mensais | 17 |
| 5.2.2 | Caso 4: Dados de vazão máxima anuais | 17 |
| 6 | Cronograma | 23 |
| | Referências Bibliográficas | 24 |

1. Introdução

O ajuste de distribuições estatísticas a dados é, oficialmente, datado de 1894, quando Karl Pearson usou um método, batizado de Método dos Momentos, para ajustar uma mistura de duas distribuições normais a dados de razão de tamanho corporal de caranguejos (Pearson, 1894). Hoje em dia, são inúmeras as áreas em que são feitos ajustes de curvas, como na biologia, na economia e na meteorologia, assim como o número de métodos empregados para tal.

Na hidrologia, isto também se aplica. O ajuste de distribuições de probabilidade a dados hidrológicos é uma das formas mais comuns de resumir uma série de dados (como a vazão de um rio) em uma função (Chow et al., 1988). Tendo uma série histórica de dados de uma região, seja ela, por exemplo, de mínimas de vazão anuais ou de máximas de precipitação diária mensais, é possível se ajustar uma curva que represente estes dados, de forma que análises estatísticas e previsões possam ser feitas para esta região; isto é útil, por exemplo, na previsão de eventos extremos, como secas e enchentes, e de vazões de projeto para projetos de estruturas vulneráveis a enchentes, como pontes e galerias de água pluvial (Serrano, 1997).

Para um ajuste eficaz de distribuição, dois itens são necessários: dados de qualidade e um método consistente com a análise. No caso dos dados, eles devem ser representativos do comportamento hidrológico e meteorológico da área de estudo. Logo, a série histórica a ser utilizada deve conter um intervalo mínimo de dados contínuo, de forma a conter os padrões de variação intersazonal e interanual da região; em geral, este número mínimo é de 20 anos (Chow et al., 1988). A respeito do método ou métodos empregados, estes devem ter uma eficácia mínima que os ajustes feitos sejam consistentes com a amostra de dados disponíveis, para assegurar que a curva possa representar o comportamento hidrológico da região estudada.

Embora, em geral, a consistência dos dados hidrológicos seja estudada e respeitada no ajuste de distribuições, algumas vezes o método não é. Na literatura, há dois métodos já consolidados e que geralmente são aplicados para estes casos: o Método dos Momentos e o Método da Máxima Verossimilhança. O que ocorre é que muitas vezes estes métodos são aplicados sem critério ou sem alguma forma de avaliação da performance deles na sumarização do conjunto de dados disponível. Cousineau (2009), por exemplo, empregou alguns métodos, entre eles o Método dos Momentos e o Método da Máxima Verossimilhança, para amostras pequenas usando simulações de Monte Carlo, e concluiu que o Método da Máxima Verossimilhança foi bastante ineficiente e que deveria ser até evitado, com erros de estimação até duas vezes pior que os demais métodos.

O Método dos Momentos-L foi recentemente proposto por Hosking (1990) visando substituir estes métodos clássicos em análises de dados hidrológicos, devido a estudos como o de Cousineau que mostram que eles não são eficazes em situações comuns a este tipo de dados: amostras pequenas e presença de *outliers* (ou valores atípicos). O Método dos Momentos-L foi desenvolvido a partir do trabalho de Greenwood et al. (1979), e tem se tornado popular na hidrologia devido a diversos trabalhos na área (Delicado e Gorla, 2008;

Goda et al., 2010) que mostram que ele é bastante eficiente nesse contexto.

Sendo assim, o que este trabalho se propõe a fazer é um estudo comparativo entre estes métodos de ajustes de distribuição a dados hidrológicos.

2. Objetivos

2.1 Objetivo Geral

Comparar o Método dos Momentos-L, mais recentemente proposto na literatura, com os métodos já estabelecidos da Máxima Verossimilhança e o Métodos dos Momentos no ajuste de distribuições de probabilidade, com ênfase em dados de vazão máxima e mínima e de precipitação máxima e mínima.

2.2 Objetivos específicos

1. Revisar os fundamentos dos diferentes métodos empregados neste trabalho e sua implementação computacional.
2. Avaliar a performance do Método dos momentos-L em comparação com a dos Métodos dos Momentos e da Máxima Verossimilhança na obtenção de parâmetros de distribuições de probabilidade comumente usadas na hidrologia em simulações com diferentes cenários.
3. Comparar a performance dos três métodos no ajuste de curvas a dados hidrológicos reais, de séries históricas de estações fluviométricas e pluviométricas.

3. Revisão Bibliográfica

O intuito desta revisão é dar as bases dos métodos paramétricos que serão utilizados neste trabalho: o Método dos Momentos, o Método da Máxima Verossimilhança e o Método dos Momentos-L. Também serão revisados na literatura alguns trabalhos que fizeram estudos comparativos sobre a performance destes métodos em situações específicas.

Para a introdução dos métodos, primeiro fazem-se necessárias algumas definições. X é uma variável aleatória, com função de densidade $f(x)$ e função distribuição cumulativa $F(x)$. O valor esperado de X é definido como:

$$E[X] = \mu = \int_{-\infty}^{\infty} x \, dF(x) = \int_{-\infty}^{\infty} x f(x) \, dx \quad (3.1)$$

Isto definido, introduz-se os métodos.

3.1 Método dos Momentos

O Método dos Momentos foi criado em 1894, quando Pearson “importou” os momentos da física à estatística para estimar os coeficientes de duas distribuições normais ajustadas a dados de razão entre a largura da frente e o comprimento do corpo de 1000 caranguejos (Pearson, 1894). Desde então, tem sido um dos métodos mais utilizados para ajuste de curvas a dados em várias áreas do conhecimento.

Os momentos de uma população são os valores esperados de potências da variável aleatória em questão, e são definidos como:

$$\mu_r = E[X^r] = \int_{-\infty}^{\infty} x^r f(x) \, dx \quad (3.2)$$

onde μ_r é o r -ésimo momento da população.

O primeiro momento é a média da população. Os momentos centrados são momentos em torno da média, e dados por:

$$\mu'_r = E[X - E(X)]^r = \int_{-\infty}^{\infty} (x - \mu)^r f(x) \, dx \quad (3.3)$$

Muito da forma de distribuições de probabilidade é explicado por medidas que são definidas em torno dos momentos. A variância, σ^2 , é definida como segundo momento centrado, e o desvio padrão é a raiz quadrada da variância. O coeficiente de variação CV, uma medida de dispersão da distribuição em torno da média, é dado por:

$$CV = \frac{\sigma}{\mu} \quad (3.4)$$

A obliquidade γ , uma medida de assimetria, e a curtose κ , uma medida de dispersão da distribuição, são definidas como razões entre momentos:

3.2. Método da Máxima Verossimilhança

$$\gamma = \frac{\mu_3}{\mu_2^{3/2}} \quad (3.5)$$

$$\kappa = \frac{\mu_4}{\mu_2^2} \quad (3.6)$$

Os momentos de uma amostra são dados de forma análoga por:

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r \quad (3.7)$$

onde m_r é o r -ésimo momento da amostra.

O Método dos Momentos consiste em se estimar os momentos da população pelos momentos amostrais, isto é, igualar os momentos amostrais dos dados aos momentos da distribuição escolhida para o ajuste, e estimar assim os parâmetros da distribuição. Ele é largamente usado pela facilidade de implementação, em especial para distribuições com apenas dois parâmetros. Porém, os momentos dão mais peso a grandes diferenças, devido a sua definição (exponenciação dos x_i), e por isso tende a não ter bom desempenho em amostras com muitos *outliers* (Hosking e Wallis, 1997).

3.2 Método da Máxima Verossimilhança

O Método da Máxima Verossimilhança foi oficialmente estruturado, e posteriormente popularizado, por R. A Fisher entre 1912 e 1922 (apesar dele já ter sido aplicado anteriormente, mas sob outros nomes) (Hald, 1999; Fisher, 1912, 1921). Este método foi proposto depois de Fisher questionar o Método dos Momentos de Pearson, que o considerava um estimador ineficaz (Fisher, 1937).

O método consiste em, dada uma amostra, se obter os parâmetros da distribuição escolhida para o ajuste que têm maior probabilidade de ser a distribuição que gerou a amostra. Em termos matemáticos, os parâmetros estimados são os que maximizam a função de verossimilhança, definida como:

$$\mathcal{L}(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) \quad (3.8)$$

Onde θ é o parâmetro ou conjunto de parâmetros a serem estimados. Como a função deve ser maximizada, o log da função de verossimilhança, chamada de função de log-verossimilhança, também pode ser maximizado, facilitando a álgebra, já que o logaritmo transforma o produtório em um somatório. A função de log-verossimilhança é dada, então, por:

$$\ell(x_1, x_2, \dots, x_n, \theta) = \sum_{i=1}^n \log f(x_i, \theta) \quad (3.9)$$

Para a maximização da função, pode ser empregada análise matemática (derivadas parciais iguais a zero) ou computacional, empregando métodos iterativos.

O Método da Máxima Verossimilhança é, em geral, mais eficaz na estimação dos parâmetros, devido a sua definição. Contudo, para amostras pequenas, este método nem sempre estima os parâmetros satisfatoriamente (Delicado e Gorla, 2008). Öztekin (2005), por exemplo, analisou dados de vazão máxima de 50 rios na Turquia, estimando os parâmetros da distribuição generalizada de Pareto por 5 métodos diferentes, e concluiu que o método dos momentos, em geral, foi superior.

3.3 Método dos Momentos-L

O Método dos Momentos-L foi proposto por J. R. M. Hosking em 1990 como uma alternativa a estes dois métodos descritos (Hosking, 1990). Ele foi desenvolvido a partir da idéia de “*momentos ponderados por probabilidade*” (tradução livre de “*probability weighted moments*”) de Greenwood et al. (1979). Esses momentos são definidos como:

$$M_{p,r,s} = E\{X^p F(X)^r [1 - F(X)]^s\} \quad (3.10)$$

Chamando de $\alpha_r = M_{1,0,r}$, de $\beta_r = M_{1,r,0}$, e fazendo $u = F(x)$, temos que:

$$\alpha_r = \int_0^1 x(u)(1-u)^r du \quad (3.11)$$

$$\beta_r = \int_0^1 x(u)u^r du \quad (3.12)$$

Os momentos-L são, então, definidos como combinações lineares destas integrais de $x(u)$, ponderados pelos “*polinômios modificados de Legendre*” (tradução livre de “*shifted Legendre polynomials*”), definidos como:

$$P_r^*(u) = \sum_{k=1}^r p_{r,k}^* u^k \quad (3.13)$$

onde

$$p_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} = \frac{(-1)^{r-k} (r+k)!}{(k!^2)(r-k)!} \quad (3.14)$$

Em termos dos momentos ponderados por probabilidade, os momentos-L são dados por:

$$\begin{aligned} \lambda_1 &= \alpha_0 = \beta_0 \\ \lambda_2 &= \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0 \\ \lambda_3 &= \alpha_0 - 6\alpha_1 + 6\alpha_2 = 2\beta_2 - 6\beta_1 + \beta_0 \\ \lambda_4 &= \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \end{aligned} \quad (3.15)$$

e, em geral:

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^r p_{r,k}^* \alpha_k = \sum_{k=0}^r P_{r,k}^* \beta_k \quad (3.16)$$

onde λ_{r+1} é o $(r+1)$ -ésimo momento-L.

Assim, o r -ésimo momento-L da população é definido em termos de $x(u)$ como:

$$\lambda_r = \int_0^1 x(u) P_{r-1}^*(u) du. \quad (3.17)$$

Os momentos-L amostrais são calculados de forma análoga. Sendo n o tamanho da amostra, e $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ a amostra ordenada de forma crescente, onde $x_{r:n}$ é o r -ésimo menor valor da amostra, os momentos-L amostrais são definidos como:

$$\begin{aligned} l_1 &= b_0 \\ l_2 &= 2b_1 - b_0 \\ l_3 &= 2b_2 - 6b_1 + b_0 \\ l_4 &= 20b_3 - 30b_2 + 12b_1 - b_0 \end{aligned} \quad (3.18)$$

3.3. Método dos Momentos-L

onde (de Landwehr et al., 1979)

$$b_r = \frac{1}{n} \sum_{j=r+1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_{j:n} \quad (3.19)$$

e, em geral

$$l_{r+1} = \sum_{k=0}^r p_{r,k}^* b_k \quad (3.20)$$

O Método dos Momentos-L funciona de forma similar ao Método do Momentos: usando o momento amostral como um estimador do momento da população, para se obter os parâmetros da função ajustada à amostra. E, assim como os momentos, os momentos-L definem quantidades que sumarizam a distribuição de probabilidade: λ_1 é a média da distribuição, λ_2 é a escala, τ é o CV-L, τ_3 é a obliquidade-L, e τ_4 é a curtose-L (Hosking, 1989), onde:

$$\tau = \frac{\lambda_2}{\lambda_1} \quad (3.21)$$

e

$$\tau_r = \frac{\lambda_r}{\lambda_2} \quad (3.22)$$

Comparando com os momentos convencionais, percebe-se que os momentos-L dão menos “*peso*” aos *outliers* da amostra, devido à exponenciação de $1 - u$ em vez de $x(u)$ (Hosking e Wallis, 1997).

4. Metodologia

Para a simulação de dados de distribuições de probabilidade e aplicação dos métodos, será utilizado o software *R* 3.2.0 (R Development Core Team, 2008). Com o pacote *FAdist* (Aucoin, 2015) serão incluídas algumas distribuições comuns em hidrologia que não existem no *Rbase*. Para a implementação do Método dos Momentos, será utilizado o pacote *moments* (Komsta e Novomestky, 2015); para a implementação do Método da Máxima Verossimilhança, serão utilizados os pacotes *MASS* (Venables e Ripley, 2002) e *fitdistrplus* (Delignette-Muller e Dutang, 2015); e para o método dos momentos-L, o pacote *lmomco* (Asquith, 2015).

Inicialmente serão simulados alguns dados de distribuições de probabilidade, e os métodos serão aplicados para se obter os parâmetros destas amostras aleatórias de distribuições. Então, será avaliada a performance dos mesmos em cada caso; em especial, sob algumas condições comuns em dados hidrológicos, como amostras pequenas ($n < 30$) ou com presença de *outliers*.

Num segundo momento, serão utilizados dados reais de séries históricas de máximas e mínimas de vazão e precipitação, que poderão ser mensais ou anuais. Os dados serão da base de dados da Agência Nacional de Águas (ANA), disponível na plataforma *HidroWeb* (ANA, 2005).

5. Análises Preliminares

Foram incluídas 4 análises dos ajustes de parâmetros pelos três métodos: duas com simulações e duas com dados reais.

5.1 Simulações

5.1.1 Caso 1: Dados simulados da distribuição gama

A função de distribuição gama é uma distribuição de dois parâmetros, β , chamado de parâmetro de taxa, e k , chamado de parâmetro de forma da função. A função gama é definida então como:

$$f(x; k, \beta) = \frac{x^{k-1} \beta^k e^{-x\beta}}{\Gamma(k)} \quad (5.1)$$

Para a distribuição gama, o primeiro e o segundo momentos são definidos por:

$$\begin{aligned} \mu &= \frac{k}{\beta} \\ \mu_2 = \text{Var}[X] &= \frac{k(k+1)}{\beta^2} \end{aligned} \quad (5.2)$$

Igualando os momentos amostrais aos momentos populacionais, temos os estimadores dos parâmetros da distribuição:

$$\begin{aligned} \hat{k} &= \frac{\bar{x}^2}{s^2} \\ \hat{\beta} &= \frac{\bar{x}}{s^2} \end{aligned} \quad (5.3)$$

A função de log-verossimilhança para a distribuição gama é definida por:

$$\ell(x_i; k, \beta) = nk \log \beta - n \log \Gamma(k) + (k-1) \sum_{i=1}^n \log x_i - \beta \sum_{i=1}^n x_i \quad (5.4)$$

Derivando em relação a k , temos:

$$\frac{\partial}{\partial k} \ell(x_i; \hat{k}, \hat{\beta}) = n(\log \hat{\beta} - \frac{d}{dk} \log \Gamma(\hat{k})) + \sum_{i=1}^n x_i = 0 \quad (5.5)$$

Derivando novamente a equação 5.4, agora em relação a β , temos:

$$\frac{\partial}{\partial \beta} \ell(x_i; \hat{k}, \hat{\beta}) = n \frac{\hat{k}}{\hat{\beta}} - \sum_{i=1}^n x_i = 0 \quad (5.6)$$

5.1. Simulações

Resolvendo a equação 5.6, obtemos $\hat{\beta} = \frac{\hat{k}}{\bar{x}}$. Substituindo na equação 5.5, e sabendo que:

$$\frac{d}{dk} \log \Gamma(\hat{k}) = \psi(\hat{k})$$

temos:

$$n(\log \hat{k} - \log(\bar{x}) - \psi(\hat{k})) + \sum_{i=1}^n \log x_i \quad (5.7)$$

A equação 5.7 não pode ser resolvida analiticamente, e usam-se métodos iterativos para se obter \hat{k} . Uma vez obtido \hat{k} , $\hat{\beta}$ é calculado rapidamente por substituição.

Os dois primeiros momentos-L da distribuição gama são definidos por (Hosking, 1986, 1990):

$$\begin{aligned} \lambda_1 &= \frac{k}{\beta} \\ \lambda_2 &= \frac{\Gamma(k + \frac{1}{2})}{\beta \sqrt{\pi} \Gamma(k)} \end{aligned} \quad (5.8)$$

E a obliquidade-L amostral é dada por:

$$\tau_3 = 6I_{1/3}(k, 2k) - 3 \quad (5.9)$$

onde $I_x(p, q)$ é a função beta incompleta.

Os estimadores dos parâmetros da distribuição são aproximados por: chamando de $t = \frac{l_2}{l_1}$, então

$$\begin{aligned} \text{se } 0 < t < \frac{1}{2}, \text{ então } z = \pi t^2 \text{ e} \\ \hat{k} \approx \frac{1 - 0.3080z}{z - 0.05812z^2 + 0.01765z^3}; \end{aligned} \quad (5.10)$$

$$\begin{aligned} \text{se } \frac{1}{2} \leq t < 1, \text{ então } z = 1 - t \text{ e} \\ \hat{k} \approx \frac{0.7213z - 0.5947z^2}{1 - 2.1817z + 1.2113z^2}; \end{aligned} \quad (5.11)$$

$$\hat{\beta} = \frac{l_1}{\hat{k}} \quad (5.12)$$

Foi gerada uma amostra aleatória da distribuição gama, com $n = 80$, $k = 3$ e $\beta = 0.75$. A densidade empírica e a função distribuição cumulativa empírica da amostra gerada podem ser vistas nas figuras 5.1 e 5.2.

Na estimativa dos parâmetros, o Método dos Momentos estimou $\beta = 0.7489$ e $k = 3.0643$; o Método da Máxima Verossimilhança estimou $\beta = 0.6958$ e $k = 2.8467$; e o Método dos Momentos-L estimou $\beta = 0.7405$ e $k = 3.0297$.

Na figura 5.3 foram desenhadas junto da densidade empírica da amostra aleatória a distribuição gama teórica que gerou a amostra e as distribuições gama teóricas com os parâmetros estimados pelos três métodos.

O código *R* que implementou toda esta análise pode ser visto na listagem 5.1.

5.1. Simulações

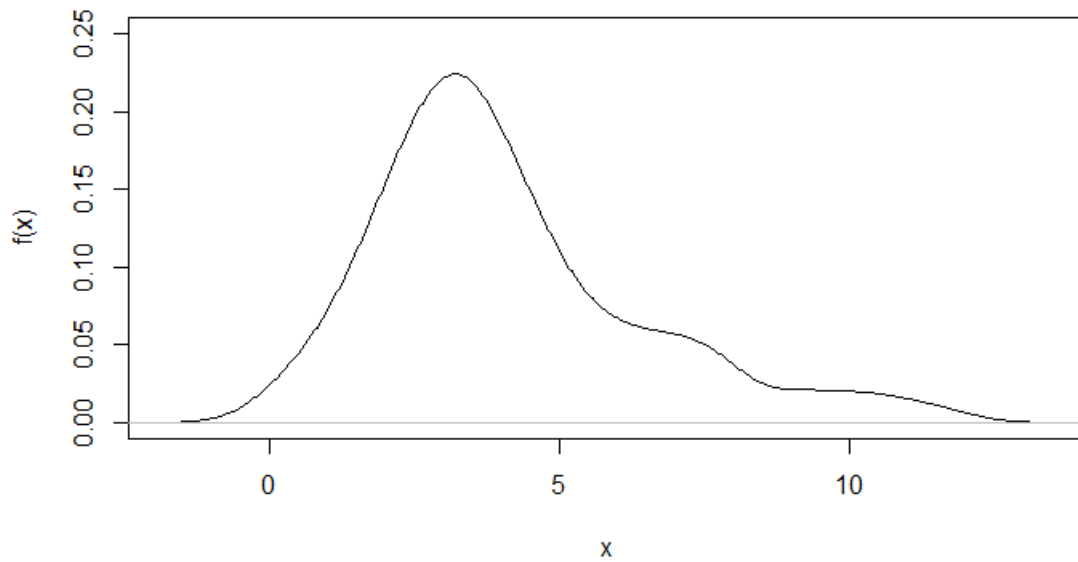


Figura 5.1: Densidade empírica da amostra aleatória gerada pela distribuição gama.

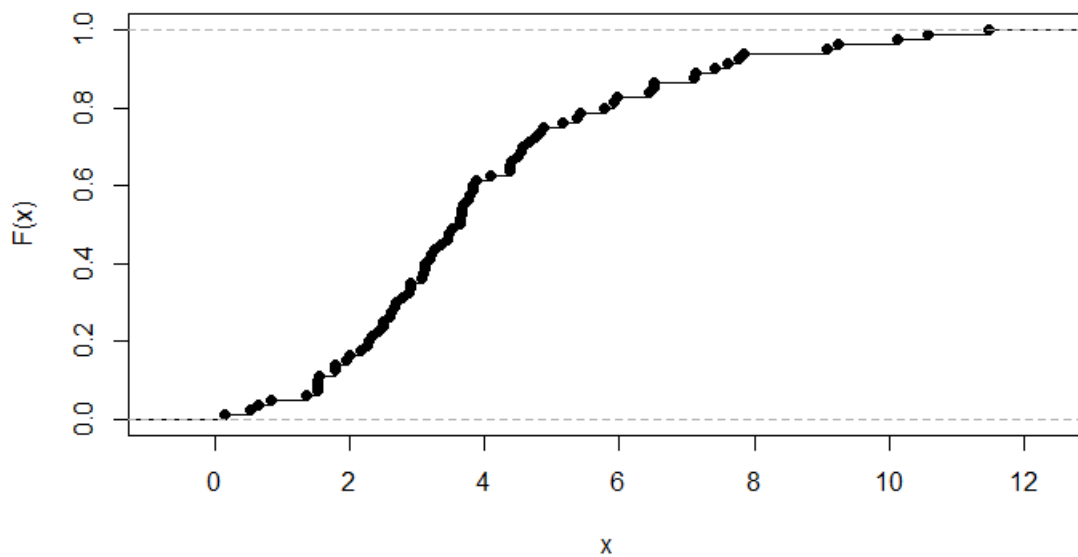


Figura 5.2: Função distribuição acumulada empírica da amostra aleatória gerada pela distribuição gama.

5.1. Simulações

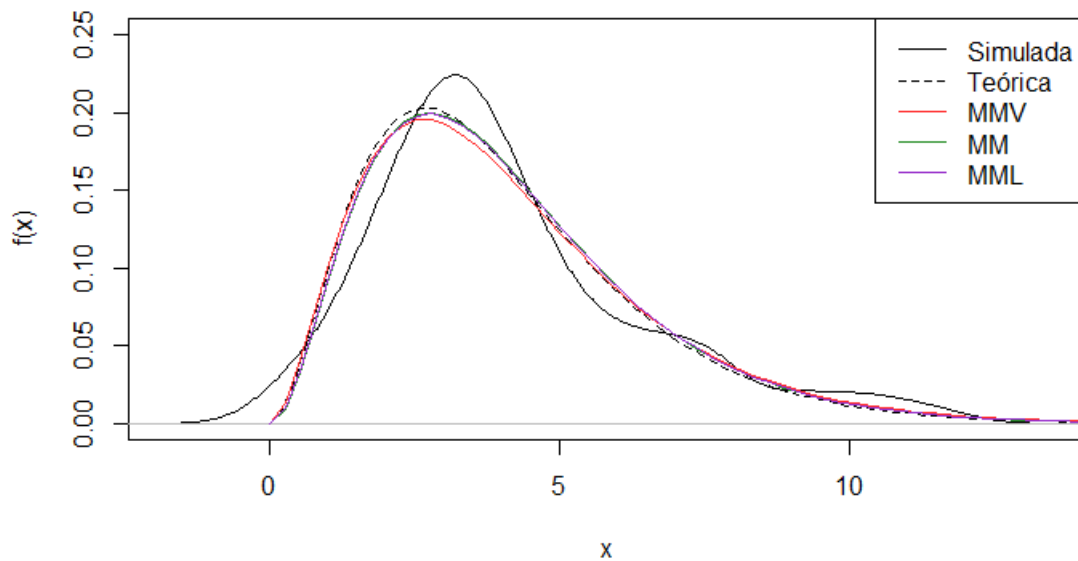


Figura 5.3: Comparativo das distribuições gama teóricas dos parâmetros estimados pelo Método dos Momentos (MM), Método da Máxima Verossimilhança (MMV) e Método dos Momentos-L (MML) com a amostra aleatória e sua distribuição teórica.

Listagem 5.1: Simulação de uma amostra aleatória da distribuição gama

1º Simulando uma amostra da distribuição gama

```
> #Gerando uma semente
> set.seed(3000)
> # Definindo o tamanho da amostra
> n = 80
> # Definindo o parâmetro beta
> rt = 0.75
> # Definindo o parâmetro k
> sh = 3.0
> # Simulando a distribuição
> x.gam <- rgamma(n=n,rate=rt,shape=sh)
```

2º Método da Máxima Verossimilhança

```
> # Função de Log-verossimilhança
> logl <- function(beta,k,dados,n) {
  nll <- -n*k*log(beta)+n*log(gamma(k))-(k-1)*sum(log(dados))+beta*sum(dados)
  return(nll)
}
> #Otimizador
> est <- mle(minuslogl=logl,start=list(beta=2,k=1),
  fixed=list(dados=x.gam, n=length(x.gam)), method="L-BFGS-B",
  lower=c(0.001,0.001), upper=c(Inf,Inf))
```


5.1. Simulações

```
> # Parâmetros estimados pelo método
> MVrt <- getElement(coef(est),1)
> MVsh <- getElement(coef(est),2)

3º Método dos Momentos

> # Cálculo de beta
> est.sh <- function(dados,mean,n) {
  esh <- (mean^2)/((1/n)*sum((dados^2)-(mean^2)))
  return(esh)
}
> MMsh = est.sh(x.gam,mean,n)
> # Cálculo de k
> est.rt <- function(dados,mean,n) {
  ert <- (mean)/((1/n)*sum((dados^2)-(mean^2)))
  return(ert)
}
> MMrt = est.rt(x.gam,mean,n)
```

4º Método dos Momentos-L

```
> # Cálculo dos momentos-L amostrais
> lmoments <- lmoms(x.gam)
> # Estimando os parâmetros a partir dos momentos-L
para.lm <- lmom2par(lmoments, type="gam")
> MLMsh <- para.lm[[2]][[1]]
> MLMrt <- 1/para.lm[[2]][[2]]
```

5.1.2 Caso 2: Dados simulados da distribuição de Weibull

A função de distribuição de Weibull é uma distribuição de três parâmetros, β , o parâmetro de taxa, k , o parâmetro de forma, e ξ , o parâmetro de localidade da função. A distribuição de Weibull é definida como:

$$f(x; k, \beta, \xi) = \frac{k}{\beta} \left(\frac{x - \xi}{\beta} \right)^{k-1} e^{-(x/\beta)^k} \quad (5.13)$$

Para a distribuição de Weibull, os primeiro, segundo e terceiro momentos são definidos por:

$$\mu = \xi + \beta \Gamma\left(1 + \frac{1}{k}\right) \quad (5.14)$$

$$\mu_2 = \beta^2 \left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right) \right] \quad (5.15)$$

$$\mu_3 = \beta^3 \left[\Gamma\left(1 + \frac{3}{k}\right) - 3\Gamma\left(1 + \frac{1}{k}\right)\Gamma\left(1 + \frac{2}{k}\right) + 2\Gamma^3\left(1 + \frac{1}{k}\right) \right] \quad (5.16)$$

A distribuição de Weibull não permite a estimação dos parâmetros diretamente a partir dos momentos amostrais, mas sim encontrando as raízes da equação 5.17, para se obter k :

$$\frac{\mu_3}{\mu_2^{3/2}} = b_1 \quad (5.17)$$

onde b_1 é a obliquidade da amostra.

5.1. Simulações

Os outros parâmetros são obtidos em seguida, por substituição, nas equações 5.14 e 5.15.

Para a distribuição de Weibull, a função de log-verossimilhança é definida como:

$$l(x_i, k, \beta, \xi) = n(\log k - k \log \lambda) + (k - 1) \sum_{i=1}^n (\log(x_i - \xi)) - \frac{1}{\beta^k} \sum_{i=1}^n (x_i - \xi)^k \quad (5.18)$$

A obtenção dos parâmetros ocorre por iteração desta equação diretamente, para se obter nos 3 parâmetros que maximizam a função de log-verossimilhança.

Os três primeiros momentos-L da distribuição de Weibull são definidos por (Goda et al., 2010):

$$\lambda = \beta \Gamma(1 + \frac{1}{k}) + \xi \quad (5.19)$$

$$\lambda_2 = \beta \Gamma(1 + \frac{1}{k})(1 - \frac{1}{2^{1/k}}) \quad (5.20)$$

$$\lambda_3 = \beta \Gamma(1 + \frac{1}{k})(1 - \frac{3}{2^{1/k}} + \frac{2}{3^{1/k}}) \quad (5.21)$$

Similar ao Método dos Momentos, o valor de k são as raízes da equação 5.22:

$$\frac{1 - \frac{3}{2^{1/k}} + \frac{2}{3^{1/k}}}{1 - \frac{1}{2^{1/k}}} = t_3 \quad (5.22)$$

onde t_3 é a obliquidade-L da amostra.

Os outros parâmetros são obtidos, então, por substituição nas equações 5.20 e 5.21.

Foi gerada uma amostra aleatória da distribuição de Weibull, com $n = 50$, $k = 5$, $\beta = 1$ e $\xi = 0$. A densidade empírica e a função distribuição cumulativa empírica da amostra gerada podem ser vistas nas figuras 5.4 e 5.5.

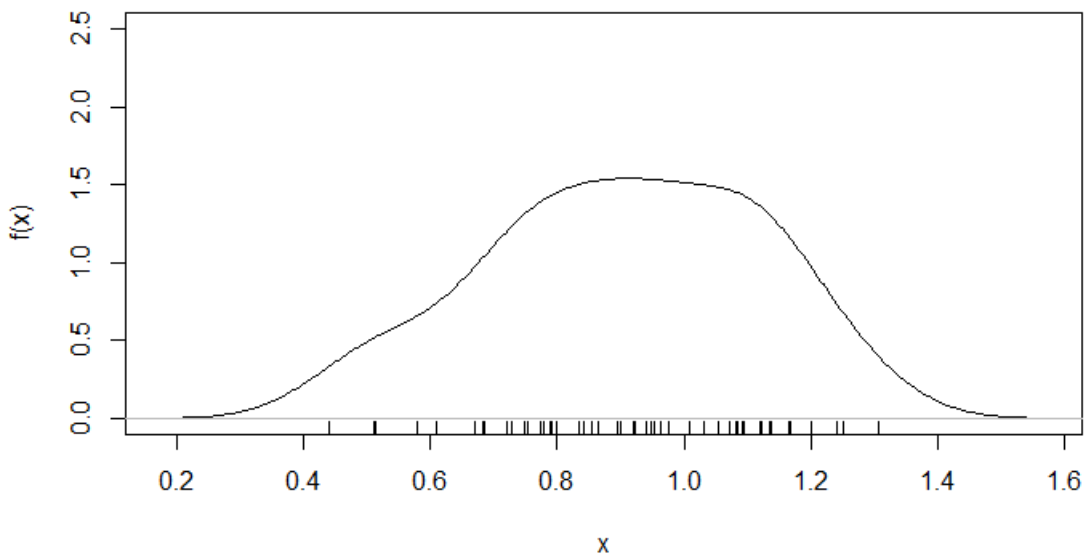


Figura 5.4: Densidade empírica da amostra aleatória gerada pela distribuição de Weibull.

5.1. Simulações

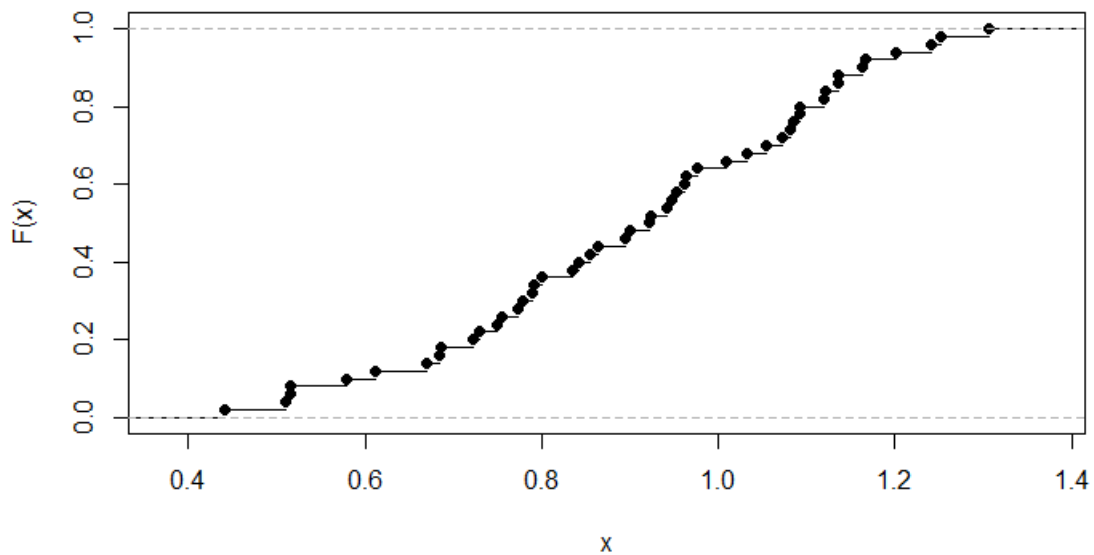


Figura 5.5: Função distribuição acumulada empírica da amostra aleatória gerada pela distribuição de Weibull.

Na estimativa dos parâmetros, utilizando o pacote *rootSolve* para o cálculo inicial das raízes da equação 5.17 (Soetaert, 2009), o Método dos Momentos estimou $\beta = 1.006043$, $k = 6.084726$ e $\xi = 0.06595146$; o Método da Máxima Verossimilhança estimou $\beta = 1.113549$, $k = 5.752357$ e $\xi = -0.1021109$; e o Método dos Momentos-L estimou $\beta = 1.162916$, $k = 6.030239$ e $\xi = 0.1498022$.

Na figura 5.6 foram desenhadas junto da densidade empírica da amostra aleatória a distribuição de Weibull teórica que gerou a amostra e as distribuições de Weibull teóricas com os parâmetros estimados pelos três métodos.

O código *R* que implementou esta análise pode ser visto na listagem 5.2.

Listagem 5.2: Simulação de uma amostra aleatória da distribuição de Weibull

1º Simulando uma amostra da distribuição de Weibull

```
> # Gerando uma semente
> set.seed(3000)
> # Definindo o tamanho da amostra
> n = 50
> # Definindo o parametro beta
> sc = 1.0
> # Definindo o parametro k
> sh = 5.0
> # Definindo o parametro csi
> thres = 0.0
> # Simulando a distribuição
> x.wei <- rweibull3(n=n,scale=sc,shape=sh,thres=thres)
```

5.1. Simulações

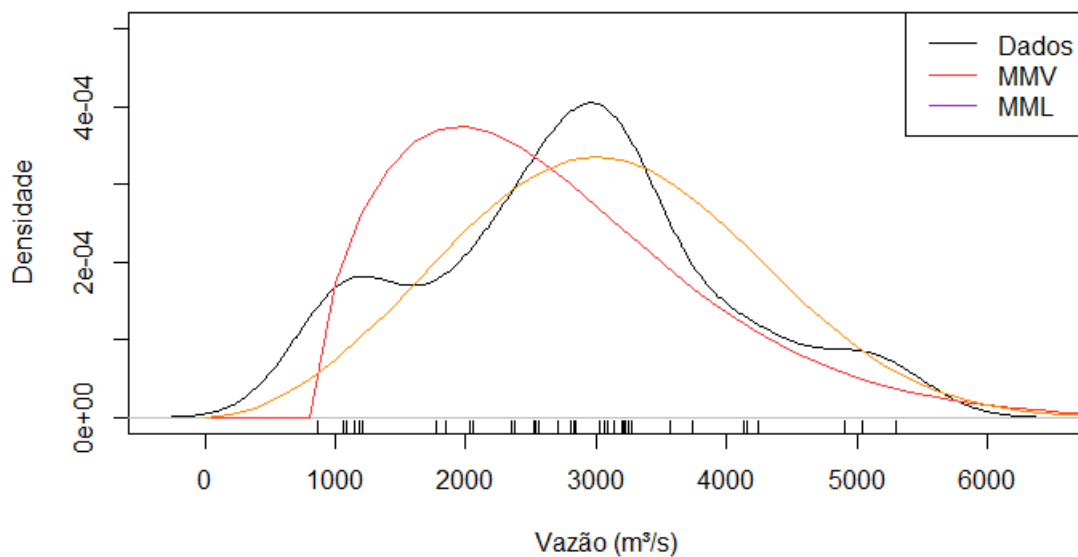


Figura 5.6: Comparativo das distribuições de Weibull teóricas dos parâmetros estimados pelo Método dos Momentos (MM), Método da Máxima Verossimilhança (MMV) e Método dos Momentos-L (MML) com a amostra aleatória e sua distribuição teórica.

```
> # Cálculo da média e variância amostrais
> mean <- mean(x.wei)
> var <- var(x.wei)
> sd <- sd(x.wei)
```

2º Método da Máxima Verossimilhança

```
> # Função de Log-verossimilhança
logl <- function(lambda,alpha,csi,dados,n) {
>   nll <- -n*(log(alpha)-alpha*log(lambda))-(alpha-1)*sum(log(dados-csi))+
  (1/(lambda^alpha))*sum((dados-csi)^alpha)
  return(nll)
}
> # Otimizador
> est <- mle(minuslogl=logl,start=list(lambda=2,alpha=1,csi=0),
  fixed=list(dados=x.wei, n=length(x.wei)), method="BFGS")
> # Parâmetros estimados pelo método
> MVsc <- getElement(coef(est),1)
> MVsh <- getElement(coef(est),2)
> MVth <- getElement(coef(est),3)
```

3º Método dos Momentos

```
> cv = sd/mean
> # Cálculo dos momentos amostrais
> momentos <- all.moments(x.wei)
> p_mom <- momentos[1]
```

5.2. Dados reais

```
> min_k = 0.1
> max_k = 1000
> # Equação para se obter k
> achak <- function(k) {
  ((sqrt(gamma(1+(2/k)))-(gamma(1+(1/k)))^2)/gamma(1+(1/k)))/cv)-1
}
> # Encontrando as raízes da equação
> MMsh <- uniroot.all(achak,c(min_k,max_k))
> # Cálculo de beta
> MMsc = (mean/gamma(1+(1/k)))^k
> # Cálculo de csi
> MMth = p_mom-MMsc*gamma(1+(1/MMsh))
```

4º Método dos Momentos-L

```
> # Cálculo dos momentos-L amostrais
> lmoments <- lmoms(x.wei)
> # Estimando os parâmetros a partir dos momentos-L
> para.lm <- lmom2par(lmoments, type="wei")
> MLMsh <- para.lm[[2]][[3]]
> MLMsc <- para.lm[[2]][[2]]
> MLMth <- para.lm[[2]][[1]]
```

5.2 Dados reais

5.2.1 Caso 3: Dados de cota máxima mensais

O primeiro conjunto de dados é de dados de cota máxima do Rio Tibagi da estação fluviométrica Barra do Mandacaia, em Telêmaco Borba, Paraná (Código 64482600). Esta é uma série histórica com 668 dados mensais de cota máxima, no período de 1987 a 2009, cuja série temporal, histograma e densidade empírica podem ser vistos, respectivamente, nas figuras 5.7, 5.8 e 5.9.

Para estes dados, optou-se por fazer um ajuste pela distribuição gama. Na estimativa dos parâmetros, o Método dos Momentos estimou $\beta = 0.1105638$ e $k = 23.35574$; o Método da Máxima Verossimilhança estimou $\beta = 0.1225037$ e $k = 25.8776096$; e o Método dos Momentos-L estimou $\beta = 0.1214153$ e $k = 25.64802$. Na figura 5.10, foram desenhadas as curvas teóricas geradas com os parâmetros estimados pelos três métodos em comparação com os dados de cota da estação.

5.2.2 Caso 4: Dados de vazão máxima anuais

O segundo conjunto de dados é de dados de vazão do Rio Paraguai da estação fluviométrica São Francisco, em Corumbá, Mato Grosso do Sul (Código 66810000). É uma série histórica com 442 registros mensais de vazão, no período de 1967 a 2007, e que pode ser vista na figura 5.11. Para esta análise, foi utilizada a vazão máxima de cada ano. Anos com mais de três meses de dados mensais não registrados foram desconsiderados, e foram usados, então, a vazão máxima de 36 anos. O histograma e a densidade empírica utilizados podem ser vistos nas figuras 5.12 e 5.13.

Para estes dados, fez-se um ajuste pela distribuição de Weibull. Na estimativa dos parâmetros, o Método da Máxima Verossimilhança estimou $\beta = 1.62107$, $k = 2044.227$

5.2. Dados reais

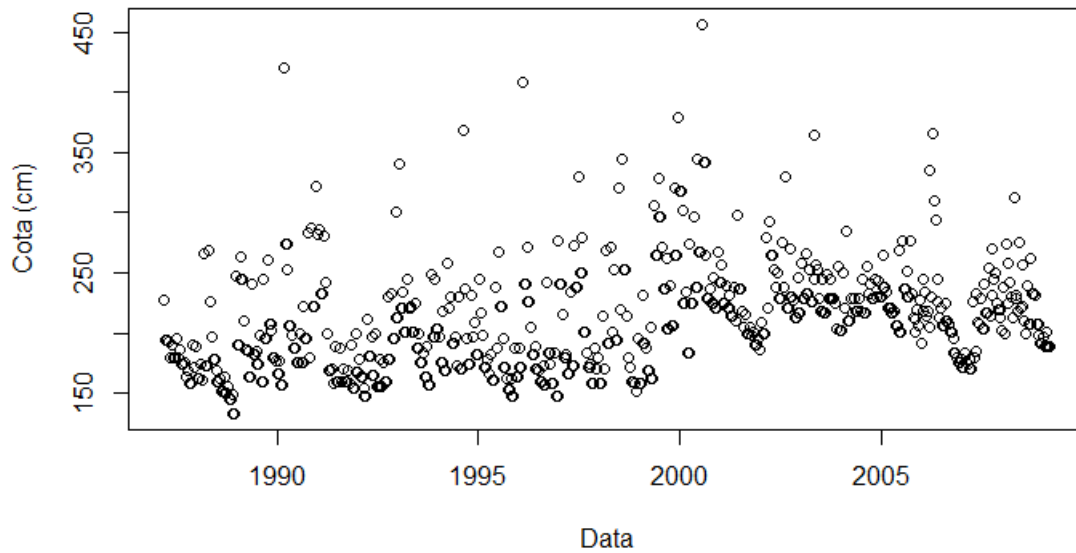


Figura 5.7: Série histórica dos dados mensais de cota máxima da estação fluviométrica Barra do Mandacaia, em Telêmaco Borba (PR).

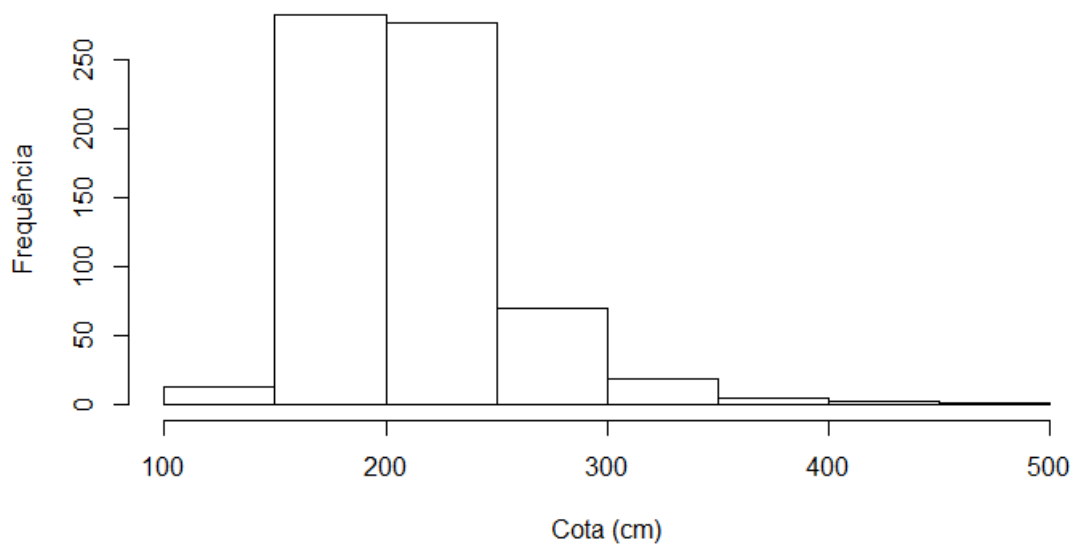


Figura 5.8: Histograma dos dados mensais de cota máxima da estação fluviométrica Barra do Mandacaia, em Telêmaco Borba (PR).

5.2. Dados reais

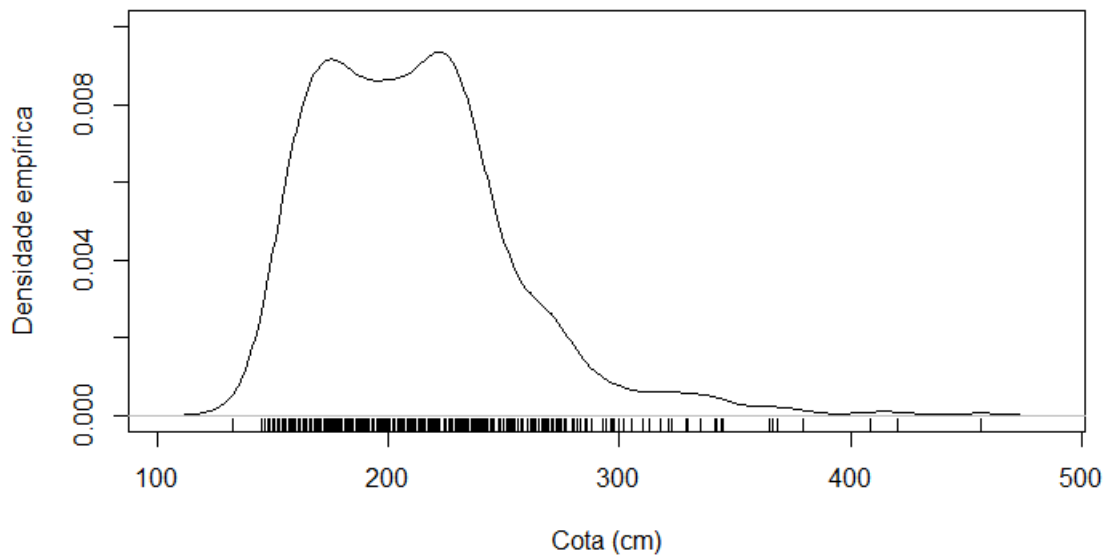


Figura 5.9: Densidade empírica dos dados de cota máxima da estação fluviométrica Barra do Mandacaia, em Telêmaco Borba (PR).

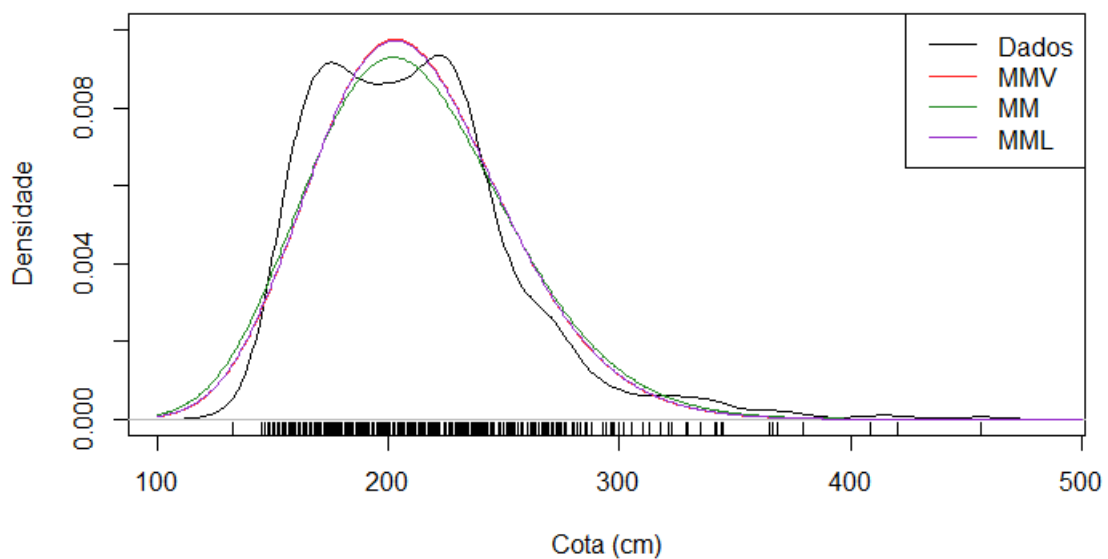


Figura 5.10: Comparativo das distribuições gama teóricas dos parâmetros estimados pelo Método dos Momentos (MM), Método da Máxima Verossimilhança (MMV) e Método dos Momentos-L (MML) com os dados de cota máxima.

5.2. Dados reais

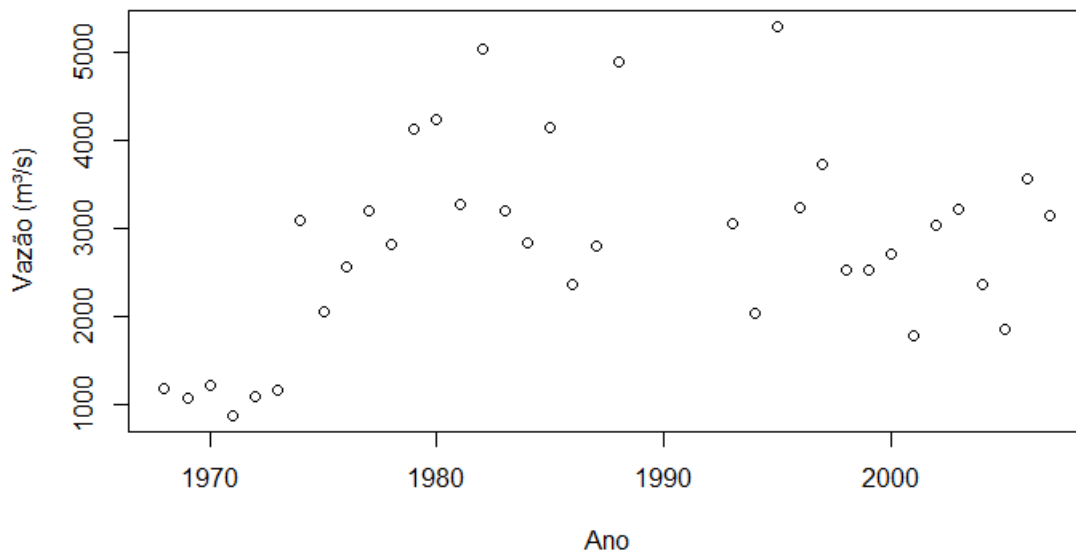


Figura 5.11: Série histórica dos dados de vazão da estação fluviométrica São Francisco, em Corumbá (MS).

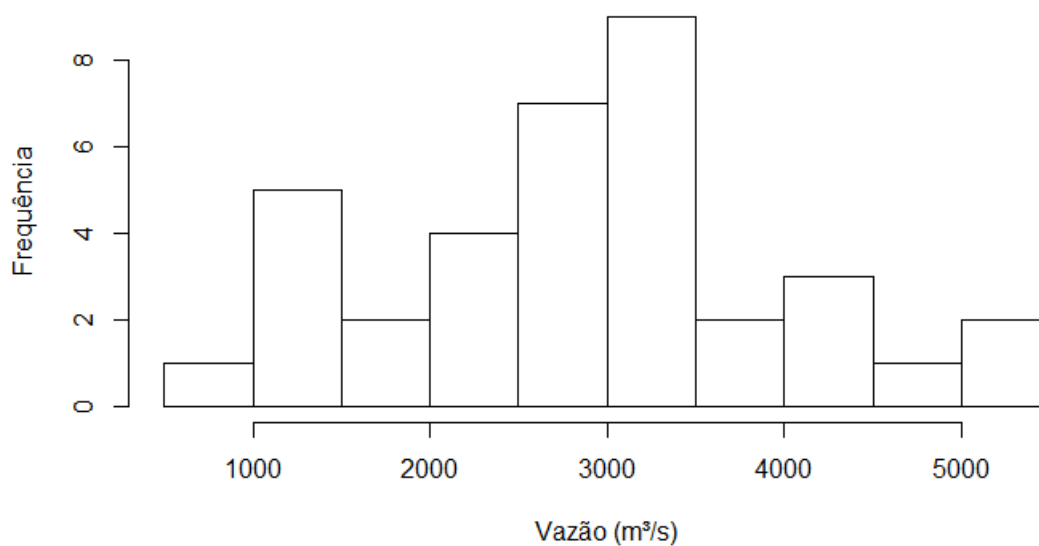


Figura 5.12: Histograma dos dados de vazão máxima da estação fluviométrica São Francisco, em Corumbá (MS).

5.2. Dados reais

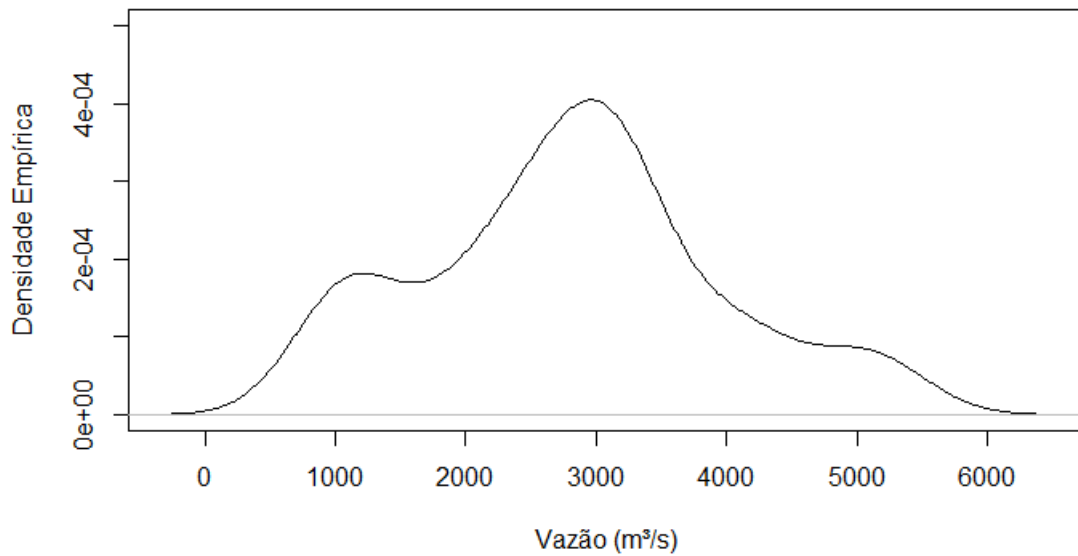


Figura 5.13: Densidade dos dados de vazão máxima da estação fluviométrica São Francisco, em Corumbá (MS).

e $\xi = 820.5561$, e o Método dos Momentos-L estimou $\beta = 2.947467$, $k = 3459.937$ e $\xi = 273.736$. O Método dos Momentos não estimou valores razoáveis para os parâmetros da distribuição, e as razões disso serão investigadas no projeto. Na figura 5.14, foram desenhadas as curvas teóricas geradas com os parâmetros estimados pelos dois métodos em comparação com os dados de cota da estação.

Observa-se que o Método dos Momentos-L obteve, em geral, estimadores similares aos do Método da Máxima Verossimilhança, considerado o mais eficaz em ajustes de curvas a dados em geral. No projeto, serão definidos métodos para se comparar a performance tanto dos ajustes feitos, quanto da distribuição de probabilidade utilizada no ajuste.

5.2. Dados reais

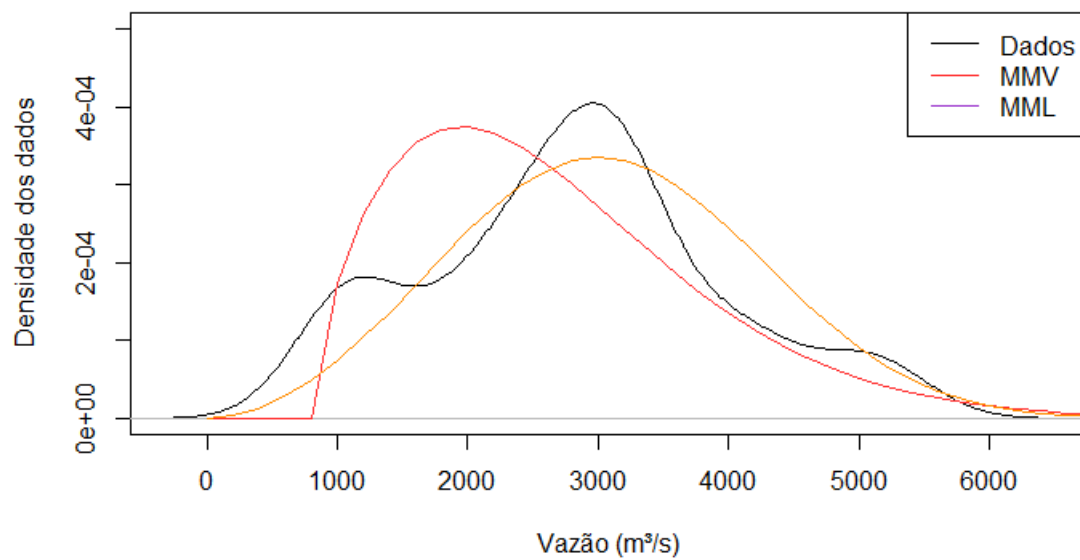


Figura 5.14: Comparativo das distribuições de Weibull teóricas dos parâmetros estimados pelo Método da Máxima Verossimilhança (MMV) e Método dos Momentos-L (MML) com os dados de vazão máxima.

6. Cronograma

O cronograma do projeto, definido para o 2º semestre do ano letivo, pode ser visto na tabela 6.1.

Tabela 6.1: Cronograma do projeto

| Atividade | Ago | Set | Out | Nov | Dez |
|--|------------|------------|------------|------------|------------|
| Revisão dos métodos para todas as distribuições usadas na hidrologia | X | | | | |
| Definição da metodologia para avaliação da performance dos métodos | | X | | | |
| Estudo de simulações | | X | X | | |
| Emprego dos métodos em dados reais | | | X | X | |
| Finalização do trabalho e construção do texto | | | | X | X |
| Defesa | | | | | X |

Referências Bibliográficas

- ANA (2005). *Hidroweb: sistemas de informações hidrológicas*. Disponível em: <http://hidroweb.ana.gov.br/HidroWeb>.
- Asquith, W. (2015). *lmomco—L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions*. R package version 2.1.4.
- Aucoin, F. (2015). *FDist: Distributions that are Sometimes Used in Hydrology*. R package version 2.1.
- Chow, V. T., Maidment, D. R., e Mays, L. W. (1988). *Applied Hydrology*. McGraw-Hill.
- Cousineau, D. (2009). Fitting the three-parameter weibull distribution: review and evaluation of existing and new methods. *IEEE Transactions on Dielectrics and Electrical Insulation*, 16(2):281–288.
- Delicado, P. e Goría, M. N. (2008). A small sample comparison of maximum likelihood, moments and l-moments methods for the asymmetric exponential power distribution. *Computational Statistics and Data Analysis*, 52(3):1661–1673.
- Delignette-Muller, M. L. e Dutang, C. (2015). *fitdistrplus: An R package for fitting distributions*. *Journal of Statistical Software*, 64(4):1–34.
- Fisher, R. A. (1912). On an absolute criterion for fitting frequency curves. *Messenger of Mathematics*, 41:155–160.
- Fisher, R. A. (1921). On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron*, 1:3–32.
- Fisher, R. A. (1937). Professor karl pearson and the method of moments. *Annals of Eugenics*, 7(4):303–318.
- Goda, Y., Kudaka, M., e Kawai, H. (2010). Incorporation of weibull distribution in l-moments method for regional frequency analysis of peaks-over-threshold wave heights. Em *Proceedings of the International Conference on Coastal Engineering*.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., e Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5):1049–1054.
- Hald, A. (1999). On the history of maximum likelihood in relation to inverse probability and least squares. *Statistical Science*, 14(2):214–222.
- Hosking, J. R. M. (1986). The theory of probability weighted moments. Technical report, IBM Research.

Referências Bibliográficas

- Hosking, J. R. M. (1989). Some theoretical results concerning l-moments. Technical report, IBM Research.
- Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society*, 52(1):105–124.
- Hosking, J. R. M. e Wallis, J. R. (1997). *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press.
- Komsta, L. e Novomestky, F. (2015). *moments: Moments, cumulants, skewness, kurtosis and related tests*. R package version 0.14.
- Landwehr, J. R., Matalas, N. C., e Wallis, J. R. (1979). Probability-weighted moments compared with some traditional techniques in estimating gumbel parameters and quantiles. *Water Resources Research*, 14:902–920.
- Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society A*, 185:71–110.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Serrano, S. E. (1997). *Hydrology for engineers, geologists and environmental professionals*. HydroScience Inc.
- Soetaert, K. (2009). *rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations*. R package version 1.6.
- Venables, W. N. e Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York, fourth edition. ISBN 0-387-95457-0.
- Öztekin, T. (2005). Comparison of parameter estimation methods for the three-parameter generalized pareto distribution. *Turkish Journal of Agriculture and Forestry*, 29:419–428.