Universidade Estadual de Maringá
Departamento de Estatística
Programa de Pós-graduação em Bioestatística

Isabela Zara Cremoneze

Aplicações da Distribuição Gumbel e Generalizações na Análise de Dados Climatológicos

Maringá,
29 de agosto de 2016

Isabela Zara Cremoneze

Aplicações da Distribuição Gumbel e Generalizações na Análise de Dados Climatológicos

Dissertação apresentada à Universidade Estadual de Maringá, como parte das exigências do Programa de Pós-Graduação em Bioestatística, área de concentração em Estatística Aplicada, para a obtenção do título de Mestre em Bioestatística.

Universidade Estadual de Maringá - UEM

Orientador: Josmar Mazucheli

Maringá,
29 de agosto de 2016

Dados Internacionais de Catalogação na Publicação (CIP)
(Biblioteca Central - UEM, Maringá, PR, Brasil)

Cremoneze, Isabela Zara
C915a Aplicações da distribuição Gumbel e generalizações na análise de dados climatológicos / Isabela Zara Cremoneze. -- Maringá, 2016.
55 f. : il., (algumas color.), figs., tabs.

Orientador: Prof. Dr. Josmar Mazucheli.
Dissertação (mestrado) - Universidade Estadual de Maringá, Departamento de Estatística, Programa de Pós-Graduação em Bioestatística, 2016.

1. Climatologia - Análises de frequências. 2. Chuvas - Variação e comportamento. 3. Dados climatológicos. 4. Distribuição Gumbel. 5. Estimação de parâmetros. 6. Probabilidade - Seleção de modelos. 7. Climatologia - Valores extremos. 8. Verossimilhança - Inferência estatística. 9. Chuvas - Frequência da intensidade. I. Mazucheli, Josmar, orient. II. Universidade Estadual de Maringá. Departamento de Estatística. Programa de Pós-Graduação em Bioestatística. III. Título.

CDD 21.ed. 551.6

AMMA-003366

À minha mãe com muito amor e carinho

Resumo

A intensidade e a frequência com que fenômenos climáticos ocorrem, como por exemplo tempestades, nevascas, tornados, entre outros, influenciam diretamente a economia. Se torna interessante então estudar esses fenômenos em termos de suas variabilidades. A análise de frequências de variáveis climatológicas tem como objetivo analisar registros históricos de modo a estimar futuras probabilidades de ocorrência por meio de uma distribuição de probabilidade. Este trabalho, no contexto de análise de frequências, tem como objetivo comparar a distribuição Gumbel, que é uma distribuição de probabilidade tradicional na análise de dados climatológicos, com algumas de suas generalizações a saber: Gumbel transmutada, Gumbel exponenciada, Exponencial Gama, Gumbel Marshall-Olkin e Gumbel generalizada. A comparação é conduzida, primeiro, utilizando dados referentes ao recorde anual do nível do mar na cidade de Porto Pirie na Austrália e, segundo, utilizando dados históricos de precipitação obtidos em 33 estações meterológicas localizadas na região sul do Brasil referentes ao período de 1961 à 2015. Em ambas análises, a distribuição Gumbel dentre todas generalizações utilizadas obteve a melhor performance, em termos da qualidade do ajuste segundo os critérios de Informação de Akaike (AIC), Informação de Akaike corrigido (AICc) e de Informação Bayesiana (BIC). Nosso estudo evidenciou que as generalizações da distribuição Gumbel não necessariamente são melhores que a distribuição Gumbel de dois parâmetros. Em um segundo momento nesta dissertação considera-se o problema de analise do comprimento de dias chuvosos com o objetivo introduzir algumas distribuições alternativas às distribuições Geométrica, Poisson truncada em zero, Logaritmica, Binomial negativa truncada em zero, Binomial negativa deslocada da origem e distribuição Binomial negativa ponderada. A análise revelou que as distribuições Poisson-Lindley truncada no zero, Poisson-Lindley deslocada da origem e Poisson-Lindley ponderada muitas vezes superam, em termos da qualidade do ajuste, as distribuições tradicionais na analise do comprimento de dias chuvosos.

Palavras-chaves: Análises de frequências, Comprimento de dias chuvosos, Dados climatológicos, Distribuição Gumbel, Estimação, Seleção de modelos, Valores extremos, Verossimilhança.

Abstract

The economics results are directly influentiated by the intensity and the frequency of the climatic phenomenas such as, storms, snowdrifts, tornadoes, etc. Thus, the study of those phenomenas is interesting in terms of their variabilities. Aiming analyze historic data to estimate future probabilities of occurency, the frequency analysis of climatological variables is made by means of the probability distribution. In the frequency analysis context, this study aims to compare the Gumbel distribution, the traditional probability distribution in climatological data analysis, with some generalizations: Gumbel transmuted, Gumbel exponenciated, Exponential Gamma, Gumbel Marshall-Olkin and generalized Gumbel. The comparison was made, first, by using the annual record of the sea level dataset from the city of Port Pirie, in Australia. Secondly, we use the historical data of precipitation, obtained from 33 meteorological stations in South of Brazil, from 1961 to 2015. For both datasets, the Gumbel distribution presented the best results, in terms of goodness-of-fit, compared with those generalizations according to the Akaike Information (AIC), Akaike Information corrected (AIC) and Bayesian Information (BIC) criterias. It shows that the generalizations are not necessarily better than the Gumbel distribution of two parameters. Moreover, we analysed the length of wet spells days aiming insert some alternative distributions to the Geometric, Logarithms, Poisson and Negative Binomial truncated to zero, Negative Binomial shifted the origin and weighted Negative Binomial ones. The analysis showed that the Poisson-Lindley truncated to zero, Poisson-Lindley shifted the origin and weighted Poisson-Lindley distributions presents better results in terms of goodness-of-fit than the usual distributions in this field.

Key-words: Frequency analysis, Length of wet spells, Climatological data, Gumbel distribution, Parameters estimation, Model selection, Extreme values, Likelihood.

Sumário

Capítulo 1

Visão Geral

1.1 Introdução

O ajuste de uma distribuição de probabilidade a dados históricos de natureza climatológica é conhecido como análises de frequências (HARTMANN; MOALA; MENDONCA, 2011). Esta análise tem por objetivo relacionar a magnitude dos eventos históricos com a sua frequência de ocorrência por intermédio de uma distribuição de probabilidade, de modo a estimar futuras probabilidades de ocorrência (NAGHETTINI; PINTO, 2007).

No geral, a modelagem de dados de natureza climatólogica geralmente utiliza as distribuições Weibull, Gama, Log-Normal, Log-Logística, Gumbel e Fréchet (WMO, 2009). Além disso, uma vez que a distribuição de probabilidade tenha sido escolhida para a modelagem do comportamento dos dados, pode-se então calcular o tempo médio esperado para que um determinado evento hidrológico seja igualado ou superado pelo menos uma vez. Este valor é conhecido na literatura como período de retorno (MARTINS, 2000).

Uma variável climatólogica muito analisada é o máximo observado em uma dada escala de tempo, máximo decendial, mensal, anual etc. Seguramente neste tipo de variável a distribuição Gumbel é mais mais utilizada. Apesar disso, várias generalizações da mesma tem sido propóstas. Uma revisão na literatura revelou pelo menos 10 generalizações, generalizações estas apresentadas e discutidas nesta dissertação.

Nesse trabalho propomos avaliar a relevância desse grande número de generali-

zações em termos de qualidade de ajuste com relação a distribuição Gumbel tradicional.

Ainda no contexto de dados de natureza climatológica, em um segundo momento, consideramos a análise do comprimento das sequências de dias chuvosos que é tão importante quanto a análise do volume da precipitação. Algumas distribuições são consideradas nesse tipo de análise devido a natureza discreta dos dados, tais como a distribuição Geométrica, Poisson truncada em zero, Logaritmica, Binomial negativa truncada em zero, Binomial negativa deslocada da origem e distribuição Binomial negativa ponderada. Propomos algumas distribuições alternativas à essas distribuições tradicionais. Essas distribuições alternativas são modificações na distribuição Poisson-Lindley para as situações de truncamento em zero.

1.2 Justificativa

Nos últimos anos diversas generalizações da distribuição Gumbel foram propostas, a saber, a distribuição Gumbel exponenciada generalizada (CORDEIRO; ORTEGA; CUNHA, 2013), Kumaraswamy Gumbel (CORDEIRO; NADARAJAH; ORTEGA, 2012), Gumbel Transmutada (SHAW; BUCKLEY, 2007), Gumbel Exponenciada (NADARAJAH, 2006), Beta Gumbel (NADARAJAH; KOTZ, 2004), Exponencial Gama (GUMBEL, 1935), Exponencial Gama (OJO, 2001), Gumbel Marshall-Olkin (MARSHALL; OLKIN, 1997) e Gumbel Generalizada (DUBEY, 1969). Segundo Nadarajah (2006), a distribuição Gumbel é seguramente a distribuição mais utilizada na análise de extremos climatológicos. No entanto, essas generalizações não foram muito utilizadas neste tipo de análise. Recentemente Pinheiro (2014) utilizou algumas dessas generalizações e aplicou em dados climatológicos. Pautados nesse motivo, comparamos a qualidade do ajuste dessas generalizações com a distribuição Gumbel.

1.3 Objetivos

1.3.1 Objetivo Geral

Este trabalho tem por objetivo apresentar e descrever a distribuição Gumbel e suas generalizações supracitadas. Feito isso, ajustaremos as mesmas em dados reais e

avaliaremos se as generalizações/extensões têm performances superiores, em termos de ajuste e critérios de discriminação, em comparação a distribuição Gumbel. A questão que surge naturalmente é: Será que é necessária a existência de tantas generalizações?

Além de encontrar entre as distribuições Poisson-Lindley, alguma distribuição alternativa, também em termos de melhor ajuste, para a análise do comprimento das sequências de dias chuvosos, através de uma aplicação em dados reais.

1.3.2 Objetivos Específicos

- Utilizando dados reais e de literatura, temos como objetivo comparar o ajuste da distribuição gumbel com ajuste de algumas de suas generalizações, a saber: Gumbel transmutada, Gumbel Marshall-Olkin, Gumbel exponenciada, Exponencial-Gama com três parâmetros e Gumbel generalizada.

- Modelar o comprimento de dias chuvosos com alguns modelos discretos, a saber: Binomial negativa deslocada da origem, Binomial negativa ponderada, Poisson-Lindley truncada no zero,Poisson-Lindley deslocada da origem e Poisson-Lindley ponderada, além das distribuições tradionais: Geométrica, Poisson truncada em zero, Logaritmica, Binomial negativa truncada em zero, Binomial negativa deslocada da origem e distribuição Binomial negativa ponderada.

1.4 Organização do Trabalho

Este trabalho está organizado como segue: no Capítulo 2 são apresentadas as distribuição Gumbel e as generalizações recentemente consideradas na literatura, além de uma aplicação com dados de literatura que consistem dos recorde anual do nível do mar observados na cidade de Porto Pirie na Austrália. No Capítulo 3, é ilustrada a modelagem dos dados referentes a precipitação máxima mensal observadas nas estações meteorológicas na região sul do Brasil. E, por fim, no Capítulo 4 é apresentado o artigo que descreve o comprimento das sequências de dias chuvosos, também na região sul do Brasil, por meio de alguns modelos discretos.

1.5 Conclusões

De acordo com a avaliação das performances das generalizações da distribuição Gumbel, no ajuste das séries climatológicas formadas pelos máximos mensais de precipitação pluvial, verificou-se que a distribuição Gumbel obteve melhor ajuste segundo os critérios de discriminação adotados, Informação de Akaike (AIC), Informação de Akaike corrigido (AICc) e Informação Bayesiana (BIC). É importante ressaltar que neste trabalho todas as generalizações utilizadas possuem um parâmetros de locação, um de escala e outro de forma.

A distribuição Gumbel padrão, com 2 parâmetros, obteve melhor performance comparada às suas generalizações. O bom ajuste da distribuição Gumbel evidenciou nenhum ganho ao se utilizar distribuições mais complexas.

A análise do comprimento das sequências de dias chuvosos revelou que as distribuições Poisson-Lindley truncada no zero, Poisson-Lindley deslocada da origem e Poisson-Lindley ponderada muitas vezes superam, em termos da qualidade do ajuste, as distribuições tradicionais.

Capítulo 2

A Distribuição Gumbel e Algumas de Suas Generalizações

A distribuição Gumbel foi originalmente proposta, em 1928, por Fisher e Tippett em teoria de Valores Extremos. Fisher e Tippett (1928), definiram três distribuições assintóticas de valores extremos, conhecidas como Gumbel, Fréchet e Weibull ou Valor Extremo tipo I, II e III, respectivamente.

Em 1945, Gumbel sugeriu que uma das formas dessa distribuição seria apropriada para a análise de frequência das cheias, desde que a série fosse anual, então essa distribuição recebeu o nome de Gumbel em sua homenagem (PINHEIRO, 2014).

Jenkinson (1955) mostrou que essas três distribuições podem ser representados em apenas uma forma paramétrica, denominada distribuição de Valores Extremos Generalizada.

Em 1945, Gumbel sugeriu que uma das formas dessa distribuição seria apropriada para a análise de frequência das cheias, desde que a série fosse anual, então essa distribuição recebeu o nome de Gumbel em sua homenagem (PINHEIRO, 2014).

Nos últimos anos a distribuição Gumbel assim como outras distribuições tem sido extensivamente generalizada. O desenvolvimento dessas extensões tem sido comum por oferecerem uma maior flexibilidade à distribuição e em consequência um ganho no agrupamento de modelos derivados da generalização (BRITO, 2014).

Neste capítulo será apresentado a distribuição Gumbel e algumas de suas gene-

ralizações assim como algumas das suas respectivas características tais como: média, variância, *p-ésimo* percentil, a expressão da log-verossimilhança e as expressões dos estimadores quando esses possuirem forma analítica. É importante destacar que a função de distribuição acumulada, da qual originará a distribuição generalizada, é denominada de distribuição base e representada por $G\left(x \mid \Theta\right)$.

2.1 A Distribuição Gumbel

Fisher e Tippett (1928) tomando, de vários conjuntos de dados, o maior valor de cada conjunto mostraram que a distribuição dos valores extremos é independente da distribuição original e se comporta como função limite (WATANABE, 2013). Gumbel (1945) sugeriu que essa distribuição de valores extremos seria apropriada para a análise de frequência de vazões máximas, desde que a série fosse anual, isto é, cada vazão da série de valores extremos fosse a maior vazão de uma amostra de 365 possibilidades (maior vazão do ano).

Diversos autores como Barbosa et al. (2014), Santos et al. (2014) e Liska et al. (2003), utilizaram essa distribuição na análise da temperatura mínima mensal, dados pluviométricos de precipitações máximas mensais e dados de máximos diários de velocidade do vento, respectivamente.

A função densidade de probabilidade da distribuição Gumbel é definida por:

$$f(x \mid \Theta) = \frac{1}{\sigma} \exp\left[-\frac{x-\mu}{\sigma} - \exp\left(-\frac{x-\mu}{\sigma}\right)\right], \tag{2.1}$$

em que $\Theta = (\mu, \sigma)$ onde $\mu \in \mathbb{R}$ é o parâmetro de locação e $\sigma > 0$ o parâmetro de escala.

A função de distribuição acumulada é escrita na forma:

$$F(x \mid \Theta) = \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]. \tag{2.2}$$

É importante destacar que (2.2) é obtida a partir da função de distribuição acumulada da distribuição de Valores Extremos Generalizada quando $\xi \to 0$ e dada por:

$$F(x \mid \Theta, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}. \tag{2.3}$$

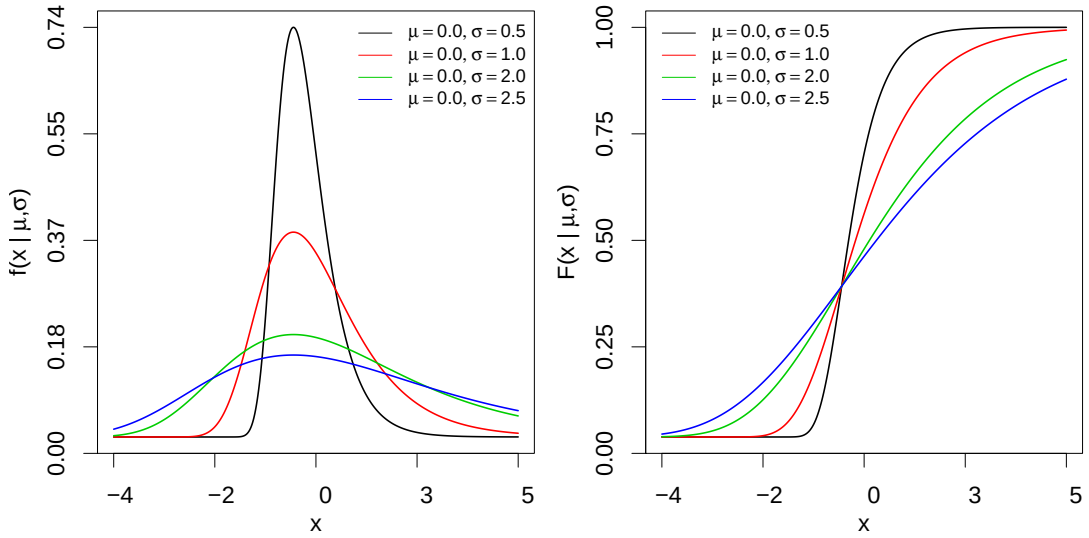

Figura 1 – Função densidade de probabilidade e função de distribuição acumulada para diferentes valores dos parâmetros da distribuição Gumbel.

Os comportamentos das funções densidade de probabilidade e da distribuição acumulada, considerando $\mu = 0$ e $\sigma = 0.5, 1.0, 2.0$ e 2.5 são ilustrados na Figura 1. Observa-se que o comportamento da função densidade de probabilidade é unimodal para todo μ e σ.

Partindo de (2.1), a média e a variância de uma variável aleatória contínua que segue uma distribuição Gumbel são dadas, respectivamente, por:

$$\mathbb{E}\left(X\right) = \mu + \gamma\sigma \tag{2.4}$$

e,

$$\mathbb{V}\left(X\right) = \frac{\pi^2}{6}\sigma^2, \tag{2.5}$$

em que γ é a constante de Euler, $\gamma \approx 0,5772$ (PINHEIRO, 2014).

O *p-ésimo* quantil da distribuição Gumbel é dado por:

$$x_p = \mu - \sigma \log\left[-\log(p)\right]. \tag{2.6}$$

em que $0 < p < 1$.

Tomando $\mathbf{x} = (x_1, \ldots, x_n)$ uma amostra aleatória proveniente da distribuição $Gumbel(\mu, \sigma)$. As estimativas de μ e σ podem ser encontradas a partir do método da

máxima verossimilhança. De (2.1) tem-se a função de verossimilhança expressa como:

$$L\left(\mu,\sigma \mid \mathbf{x}\right) = \frac{1}{\sigma^n}\exp\left[\sum_{i=1}^{n} -\frac{x_i-\mu}{\sigma} - \exp\left(-\frac{x_i-\mu}{\sigma}\right)\right], \tag{2.7}$$

cujo logarítimo fica escrito na forma:

$$\ell\left(\mu,\sigma \mid \mathbf{x}\right) = -n\log(\sigma) - \sum_{i=1}^{n}\frac{x_i-\mu}{\sigma} - \sum_{i=1}^{n}\exp\left(-\frac{x_i-\mu}{\sigma}\right). \tag{2.8}$$

As estimativas de máxima verossimilhança de μ e σ, $\widehat{\mu}$ e $\widehat{\sigma}$ respectivamente, são obtidas pela maximização da função de verossimilhança ou log-verossimilhança. A função dada em (2.8) é maximizada resolvendo, simultaneamente em μ e σ, o sistema de equações formado por:

$$U(\mu,\sigma \mid \mathbf{x}) = \begin{cases} \dfrac{\partial}{\partial\mu}\ell\left(\mu,\sigma \mid \mathbf{x}\right) = \dfrac{n}{\sigma} - \sum\limits_{i=1}^{n}\exp\left(-\dfrac{x_i-\mu}{\sigma}\right) \\ \\ \dfrac{\partial}{\partial\sigma}\ell\left(\mu,\sigma \mid \mathbf{x}\right) = -\dfrac{n}{\sigma} + \sum\limits_{i=1}^{n}\dfrac{x_i-\mu}{\sigma^2}\left\{1-\exp\left(-\dfrac{x_i-\mu}{\sigma}\right)\right\}. \end{cases} \tag{2.9}$$

O sistema de equações dado em (2.9) é não-linear. Logo, as estimativas de máxima verossimilhança $\widehat{\Theta} = (\widehat{\mu},\widehat{\sigma})$ de $\Theta = (\mu,\sigma)$, podem ser obtidas via métodos numéricos. Os intervalos de confiança para os parâmetros da distribuição Gumbel podem ser construídos a partir da normalidade assintótica dos estimadores de máxima verossimilhança.

2.2 Generalizações da Distribuição Gumbel

Nesta seção serão apresentadas as generalizações da distribuição Gumbel utilizadas nesta dissertação, a saber: a distribuição Gumbel transmutada, Gumbel Marshall-Olkin, Gumbel exponenciada, Exponencial-Gama com três parâmetros e Gumbel generalizada. Todas essas distribuições foram consideradas em Pinheiro (2014). Ainda em Pinheiro (2014) são apresentadas mais generalizações, a saber: Kumaraswamy Gumbel, Exponencial Gama, Logística generelizada tipo IV, Gumbel exponenciada generalizada, Beta Gumbel, Kummer beta generalizada Gumbel.

2.2.1 A Distribuição Gumbel Transmutada

Uma distribuição transmutada é a combinação de uma função de distribuição acumulada com sua inversa, resultando em uma distribuição transmutada mais flexível em relação a assimetria além disso o parâmetro de forma adicionado é flexivel e interpretável (ARYAL; TSOKOS, 2009). A função distribuição acumulada de uma variável aleatória X definida nos reais com distribuição transmutada dada por Shaw e Buckley (2007) é escrita na forma:

$$F(x \mid \Psi) = (1+\alpha)\, G(x \mid \Theta) - \alpha G(x \mid \Theta)^2, \tag{2.10}$$

em que $\Psi = (\mu, \sigma, \alpha)$.

De (2.10) a função densidade de probabilidade é dada por:

$$f(x \mid \Psi) = g(x \mid \Theta)\left[1 + \alpha - 2\alpha G(x \mid \Theta)\right]. \tag{2.11}$$

Assim quando substituimos a distribuição base pela distribuição Gumbel, obtem-se a expressão da função distribuição acumulada da distribuição Gumbel transmutada:

$$F\left(x \mid \Psi\right) = (1+\alpha)\exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right] - \alpha\exp\left[-2\exp\left(-\frac{x-\mu}{\sigma}\right)\right], \tag{2.12}$$

e a função densidade expressa por:

$$\begin{aligned} f\left(x \mid \Psi\right) &= \frac{1}{\sigma}\exp\left[-\frac{x-\mu}{\sigma} - \exp\left(-\frac{x-\mu}{\sigma}\right)\right] \\ &\times \left\{1 + \alpha - 2\alpha\exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\}, \end{aligned} \tag{2.13}$$

em que $\mu \in \mathbb{R}$ é o parâmetro de locação, $\sigma > 0$ de escala e $\mid \alpha \mid \leq 1$ de forma, denotada por $TGum(\mu, \sigma, \alpha)$. É fácil ver que quando $\alpha = 0$ temos como caso particular a distribuição Gumbel.

Na Figura 2 é apresentada a influência do parâmetro α nos comportamentos da função densidade de probabilidade e da distribuição acumulada Gumbel transmutada.

A média e a variância de uma variável aleatória que segue uma distribuição Gumbel transmutada é escrita na forma:

$$\mathbb{E}\left(X\right) = (\mu + \gamma\sigma) - \alpha\sigma\log 2 \tag{2.14}$$

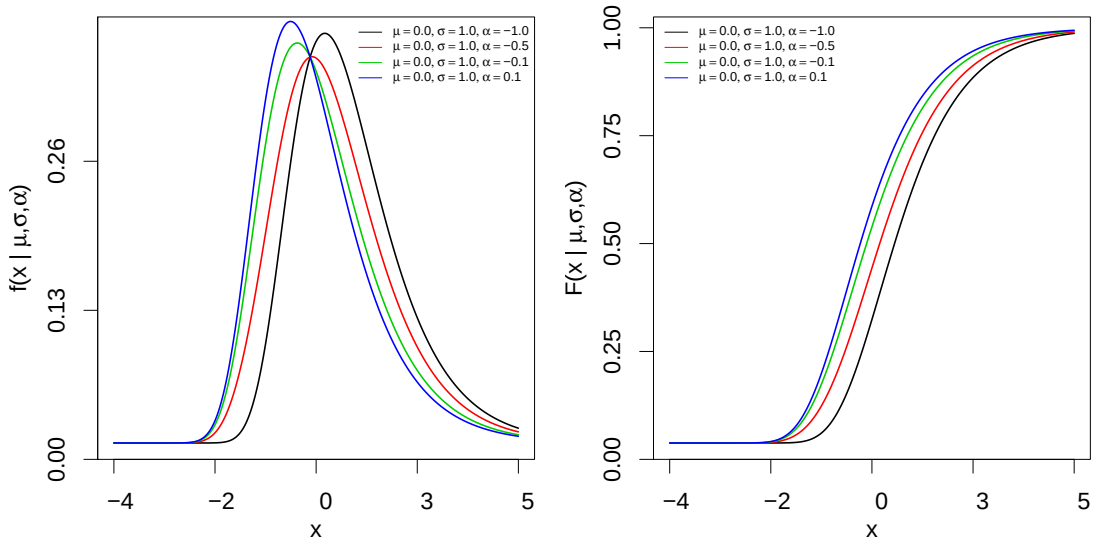

Figura 2 – Função densidade de probabilidade e função de distribuição acumulada para diferentes valores dos parâmetros da distribuição Gumbel transmutada

e,

$$\mathbb{V}(X) = \sigma^2 \left[\frac{\pi^2}{6} - \alpha (1+\alpha) (\log 2)^2 \right], \tag{2.15}$$

em que γ corresponde a constante de Euler, $\gamma \approx 0,5772$.

O *p-ésimo* percentil da distribuição Gumbel transmutada dado por:

$$x_p = \mu - \sigma \log \left[-\log(p) \right]. \tag{2.16}$$

Seja $\mathbf{x} = (x_1, \ldots, x_n)$ uma amostra aleatória com distribuição $TGum(\mu, \sigma, \alpha)$. A função de verossimilhança é expressa como:

$$\begin{aligned} \ell(\Psi \mid \mathbf{x}) &= -n \log(\sigma) - \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma} - \sum_{i=1}^{n} \exp\left(-\frac{x_i - \mu}{\sigma} \right) \\ &+ \log \sum_{i=1}^{n} \left\{ 1 + \alpha - 2\alpha \exp\left[-\exp\left(-\frac{x_i - \mu}{\sigma} \right) \right] \right\}. \end{aligned} \tag{2.17}$$

As estimativas de máxima verossimilhança de Ψ são obtidas resolvendo numericamente e simultaneamente o sistema formado pelas equações:

$$\left(\frac{\partial}{\partial \mu} \ell(\Psi \mid \mathbf{x}), \frac{\partial}{\partial \sigma} \ell(\Psi \mid \mathbf{x}), \frac{\partial}{\partial \alpha} \ell(\Psi \mid \mathbf{x}) \right) = 0 \tag{2.18}$$

2.2.2 A Distribuição Gumbel Marshall-Olkin

A classe de distribuições Marshall-Olkin foi proposta por Marshall e Olkin (1997), que introduziram um método de adicionar um novo parâmetro a uma distribuição base. A classe de distribuições Marshall-Olkin tem função distribuição acumulada expressa por:

$$F\left(x \mid \Psi\right) = 1 - \frac{\alpha\left[1 - G(x \mid \Theta)\right]}{1 - (1-\alpha)\left[1 - G(x \mid \Theta)\right]} \tag{2.19}$$

em que $\Psi = (\mu, \sigma, \alpha)$, $\mu \in \mathbb{R}$ é o parâmetro de locação, $\sigma > 0$ o parâmetro de dispersão e $\alpha > 0$ o parâmetro de forma.

De (2.19) a função densidade de probabilidade é expressa na forma:

$$f\left(x \mid \Psi\right) = \frac{\alpha g(x \mid \Theta)}{\left(1 - (1-\alpha)\left[1 - G(x \mid \Theta)\right]\right)^2}. \tag{2.20}$$

Portanto a distribuição Gumbel Marshall-Olkin possui a função de distribuição acumulada na forma:

$$F\left(x \mid \Psi\right) = \frac{1 - \alpha\left(1 - \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right].\right)}{\left(1 - (1-\alpha)(1 - \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right])\right)} \tag{2.21}$$

e a função densidade de probabilidade:

$$f\left(x \mid \Psi\right) = \frac{\frac{\alpha}{\sigma}\exp\left[-\frac{x-\mu}{\sigma} - \exp\left(-\frac{x-\mu}{\sigma}\right)\right]}{\sigma\left(1 - (1-\alpha)\left[1 - \exp\left(-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right)\right]\right)^2} \tag{2.22}$$

Quando $\alpha = 1$ a distribuição Gumbel Marshall-Olkin coincide com a distribuição Gumbel.

A Figura 3 apresenta a influência do parâmetro α nos comportamentos da função densidade de probabilidade e da função distribuição Gumbel Marshall-Olkin.

A média e variância de uma variável aleatória que segue uma distribuição Gumbel Marsharll-Olkin não podem ser escritos de forma explicita.

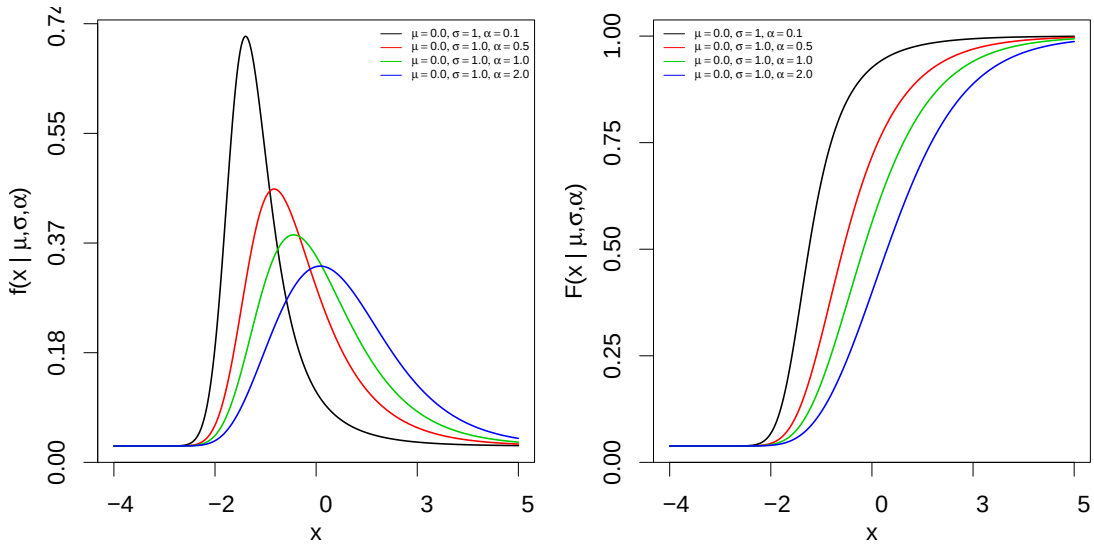

Figura 3 – Função densidade de probabilidade e função de distribuição acumulada para diferentes valores dos parâmetros da distribuição Gumbel Marshall-Olkin.

Seja $\mathbf{x} = (x_1, \ldots, x_n)$ uma amostra aleatória com distribuição $MHGumbel(\mu, \sigma, \alpha)$. A função de verossimilhança é expressa como:

$$\begin{aligned} \ell(\Psi \mid \mathbf{x}) &= n\log(\alpha) + \sum_{i=1}^{n}\left(-\frac{x_i-\mu}{\sigma} - \exp\left(-\frac{x_i-\mu}{\sigma}\right)\right) \\ &- n\log(\sigma) + \sum_{i=1}^{n}\log\left(\left(1-(1-\alpha)\left[1-\exp\left(-\exp\left(-\frac{x_i-\mu}{\sigma}\right)\right)\right]\right)^2\right). \end{aligned} \tag{2.23}$$

Os estimadores de máxima verossimilhança de Ψ são obtidos resolvendo-se o sistema formado pelas equações:

$$\left(\frac{\partial}{\partial\mu}\ell(\Psi \mid \mathbf{x}), \frac{\partial}{\partial\sigma}\ell(\Psi \mid \mathbf{x}), \frac{\partial}{\partial\alpha}\ell(\Psi \mid \mathbf{x})\right) = 0. \tag{2.24}$$

2.2.3 A Distribuição Gumbel Exponenciada

Segundo Pascoa (2012), a classe de distribuições estendidas é uma variação das classes das distribuições exponenciadas. Também chamada de distribuição Gumbel estendida a distribuição Gumbel exponenciada foi proposta por Nadarajah (2006), inspirada na distribuição exponencial exponenciada. Sua função de distribuição acumulada é obtida exponenciando a função de sobrevivência a um parâmetro α, ao invés da função distribuição.

Considerando $(\alpha > 0)$, tem-se a expressão:

$$F(x \mid \Psi) = 1 - [1 - G(x \mid \Theta)]^{\alpha}, \tag{2.25}$$

em que $\Psi = (\mu, \sigma, \alpha)$. Logo a função densidade de probabilidade pode ser obtida derivando (2.25), ou seja:

$$f(x \mid \Psi) = \alpha g(x \mid \Theta)[1 - G(x \mid \Theta)]^{\alpha - 1}. \tag{2.26}$$

Substituindo então, a distribuição base pela função distribuição Gumbel (2.1) obtem-se a função distribuição acumulada Gumbel Exponenciada para uma variável aleatória X definida na reta dos reais, expressa por:

$$F(x \mid \Psi) = 1 - \left\{1 - \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\}^{\alpha}, \tag{2.27}$$

e sua função densidade de probabilidade, por:

$$\begin{aligned} f(x \mid \Psi) &= \frac{\alpha}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right)\exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right] \\ &\times \left\{1 - \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\}^{\alpha-1}, \end{aligned} \tag{2.28}$$

com o parâmetro de locação $\mu \in \mathbb{R}$, de escala $\sigma > 0$ e de forma $\alpha > 0$, denotada por $EstGumbel(\mu, \sigma, \alpha)$. Note que para $\alpha = 1$ obtem-se a função densidade de probabilidade Gumbel.

A Figura 4 ilustra a influência do parâmetro α na densidade e na função de distribuição acumulada.

As expressões para a média e variância não possuem forma analítica. A partir de (2.28), é possível calcular o quantil x_p:

$$x_p = \mu - \sigma \log\left\{-\log\left[1 - (1-p)^{1/\alpha}\right]\right\}. \tag{2.29}$$

Considerando uma amostra aleatória $\mathbf{x} = (x_1, \ldots, x_n)$ da distribuição $ExpGumbel(\mu, \sigma, \alpha)$ a função log-verossimilhança é expressa como:

$$\begin{aligned} \ell(\Psi \mid \mathbf{x}) &= n\log(\alpha) - n\log(\sigma) - \sum_{i=1}^{n}\frac{x_i - \mu}{\sigma} - \sum_{i=1}^{n}\exp\left(-\frac{x_i-\mu}{\sigma}\right) \\ &+ (\alpha - 1)\log\sum_{i=1}^{n}\left\{1 - \exp\left[-\exp\left(-\frac{x_i-\mu}{\sigma}\right)\right]\right\}. \end{aligned} \tag{2.30}$$

As estimativas de máxima verossimilhança de Ψ são obtidas resolvendo numericamente o sistema de equações gerado de:

$$\left(\frac{\partial}{\partial\mu}\ell(\Psi \mid \mathbf{x}), \frac{\partial}{\partial\sigma}\ell(\Psi \mid \mathbf{x}), \frac{\partial}{\partial\alpha}\ell(\Psi \mid \mathbf{x})\right) = 0. \tag{2.31}$$

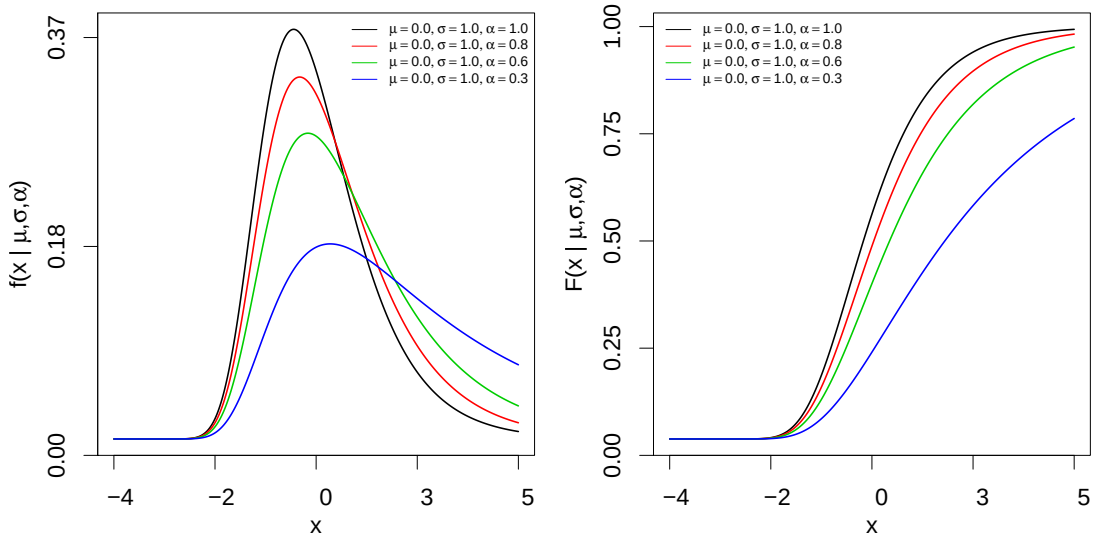

Figura 4 – Função densidade de probabilidade e função de distribuição acumulada para diferentes valores dos parâmetros da distribuição Gumbel exponenciada.

2.2.4 A Distribuição Exponencial Gama com Três Parâmetros

Proposta por Ojo (2001), a distribuição exponencial Gama com três parâmetros foi a primeira generalização da distribuição Gumbel.

Uma variável aleatória X, definida nos reais, tem a função distribuição acumulada expressa por:

$$F\left(x \mid \Psi\right) = \frac{1}{\Gamma\left(\alpha\right)}\Gamma\left[\alpha, \exp\left(-\frac{x-\mu}{\sigma}\right)\right], \tag{2.32}$$

onde $\Psi = (\mu, \sigma, \alpha)$ e sua função densidade de probabilidade, por:

$$f\left(x \mid \Psi\right) = \frac{1}{\Gamma\left(\alpha\right)}\frac{1}{\sigma}\exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\exp\left(-\alpha\frac{x-\mu}{\sigma}\right), \tag{2.33}$$

com o parâmetro de locação $\mu \in \mathbb{R}$, de dispersão $\sigma > 0$ e de forma $\alpha > 0$ e a expressão $\Gamma\left(w, t\right) = \int_x^\infty t^{w-1}\exp(-t)dt$ é a função gama incompleta. Quando $\alpha = 1$ a função é um caso particular de uma distribuição Gumbel.

A Figura 5 apresenta a influência do parâmetro α nos comportamentos da função densidade de probabilidade e da função distribuição da exponencial Gama com 3 parâmetros.

A média e a variância de uma variável aleatória que segue uma distribuição expo-

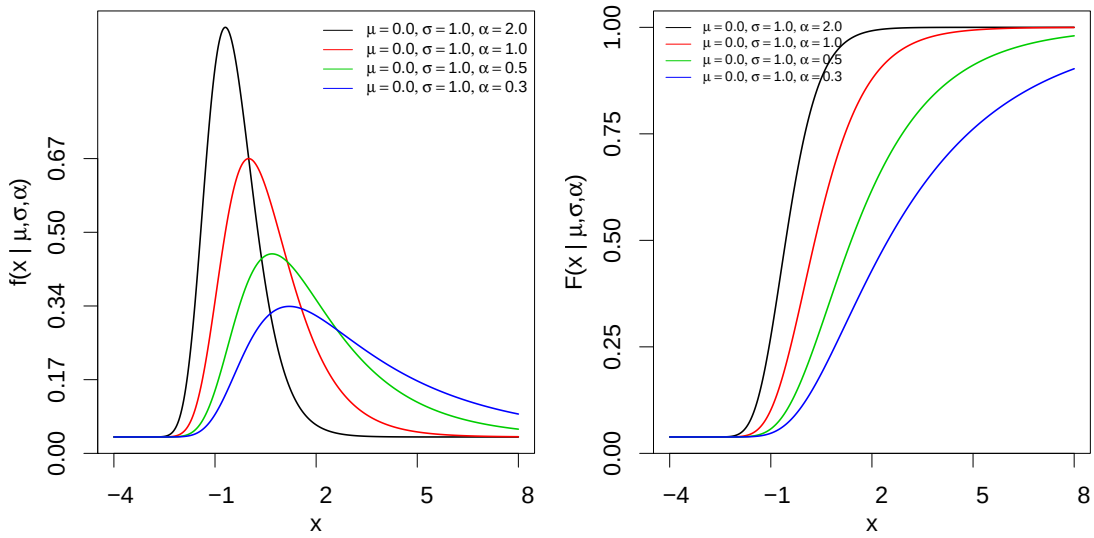

Figura 5 – Função densidade de probabilidade e função de distribuição acumulada para diferentes valores dos parâmetros da distribuição exponencial Gama com 3 parâmetros.

nencial Gama com 3 parâmetros é escrita na forma (PINHEIRO, 2014):

$$\mathbb{E}\left(X\right) = \mu - \sigma\psi\left(\alpha\right) \tag{2.34}$$

e,

$$\mathbb{V}\left(X\right) = \sigma^2\psi'\left(\alpha\right), \tag{2.35}$$

em que $\psi(\theta) = \frac{\partial}{\partial\theta}\log\left[\Gamma\left(\theta\right)\right]$. O *p-ésimo* percentil x_p da distribuição exponencial Gama com três parâmetros não pode ser obtido analiticamente.

Considerando $\mathbf{x} = (x_1, \ldots, x_n)$ uma amostra aleatória independente com distribuição exponencial Gama com 3 parâmetros. Obtem-se as estimativas de seus parâmetros através do logaritmo da função de verossimilhança, que é expressa por:

$$\ell\left(\Psi \mid \mathbf{x}\right) = -n\log\left[\Gamma(\alpha)\right] - n\log(\sigma) - \exp\left(-\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma}\right) - \sum_{i=1}^{n}\alpha\frac{x_i-\mu}{\sigma}. \tag{2.36}$$

As estimativas de máxima verossimilhança de Ψ são obtidas resolvendo numericamente o sistema de equações gerado de:

$$\left(\frac{\partial}{\partial\mu}\ell\left(\Psi \mid \mathbf{x}\right), \frac{\partial}{\partial\sigma}\ell\left(\Psi \mid \mathbf{x}\right), \frac{\partial}{\partial\alpha}\ell\left(\Psi \mid \mathbf{x}\right)\right) = 0. \tag{2.37}$$

2.2.5 A Distribuição Gumbel Generalizada

A distribuição Gumbel generalizada, foi proposta por Dubey (1969) como uma generalização da distribuição valor extremo máximo, que adicionou um terceiro parâmetro na distribuição e supõe que esse parâmetro tem distribuição gama (α, β), onde a função distribuição acumulada é dada por:

$$F\left(x \mid \Psi\right) = \left[1 + \frac{\sigma}{\beta} \exp\left(-\frac{x-\mu}{\sigma}\right)\right]^{-\alpha}, \tag{2.38}$$

onde $\Psi = (\mu, \sigma, \alpha)$ em que $\mu \in \mathbb{R}$ é o parâmetro de locação, $\sigma > 0$ o parâmetro de dispersão, $\alpha > 0$ e $\beta > 0$. Tendo,

$$F\left(x \mid \Psi\right) = \left[1 + \frac{1}{\alpha} \exp\left(-\frac{x-\mu^*}{\sigma}\right)\right]^{-\alpha}. \tag{2.39}$$

Tomando $\mu^* = \mu + \sigma \log\left(\sigma\alpha\beta^{-1}\right) \in \mathbb{R}$, $\sigma > 0$ o parâmetro de dispersão, $\alpha > 0$ e $\beta > 0$ a distribuição acima é equivalente a distribuição Gumbel generalizada, μ, σ, e α. Sua função densidade de probabilidade é expressa por:

$$f\left(x \mid \Psi\right) = \frac{1}{\sigma}\left[1 + \frac{1}{\alpha} \exp\left(-\frac{x-\mu}{\sigma}\right)\right]^{-\alpha-1} \exp\left(-\frac{x-\mu}{\sigma}\right), \tag{2.40}$$

e sua função distribuição acumulada, por:

$$F\left(x \mid \Psi\right) = \left[1 + \frac{1}{\alpha} \exp\left(-\frac{x-\mu}{\sigma}\right)\right]^{-\alpha}, \tag{2.41}$$

o parâmetro de locação $\mu \in \mathbb{R}$, o parâmetro de dispersão $\sigma > 0$ e $\alpha > 0$. A função distribuição Gumbel generalizada tem como caso particular a distribuição Gumbel quando $\alpha \to \infty$.

A Figura 6 apresenta a influência do parâmetro α nos comportamentos da função densidade de probabilidade e da função distribuição Gumbel generalizada.

A média e a variância de uma variável aleatória que segue uma distribuição Gumbel generalizada é escrita na forma (PINHEIRO, 2014):

$$\mathbb{E}\left(X\right) = \mu - \sigma(-\gamma + \log(\alpha) - \psi(\alpha)) \tag{2.42}$$

e,

$$\mathbb{V}\left(X\right) = \sigma^2\left[\frac{\pi^2}{6} + \psi(\alpha)\right], \tag{2.43}$$

onde γ é a constante de Euler, $\gamma \approx 0,5772$, e $\psi(\theta) = \frac{\partial}{\partial\theta}\log\left[\Gamma\left(\theta\right)\right]$.

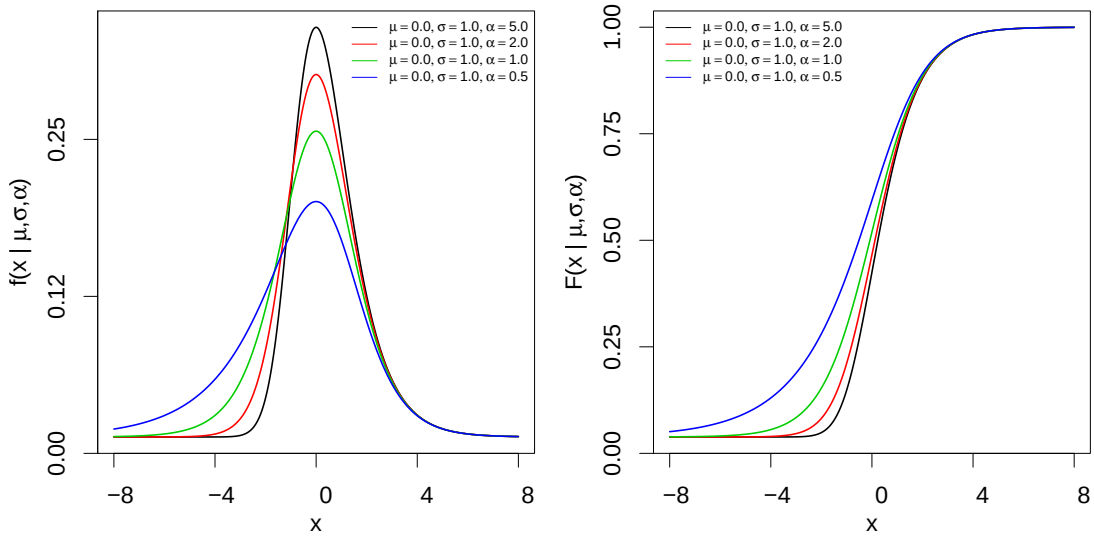

Figura 6 – Função densidade de probabilidade e função de distribuição acumulada para diferentes valores dos parâmetros da distribuição Gumbel generalizada.

O *p-ésimo* da distribuição Gumbel generalizada é dado por:

$$x_p = \mu - \sigma \log \left[\alpha \left(p^{-1/\alpha} - 1 \right) \right]. \tag{2.44}$$

Seja $\mathbf{x} = (x_1, \ldots, x_n)$ uma amostra aleatória com distribuição $GGum(\mu, \sigma, \alpha)$. A função de verossimilhança é expressa como:

$$\ell\left(\Psi \mid \mathbf{x}\right) = -n \log(\sigma) + (\alpha - 1) \log(1 + \frac{1}{\alpha} \exp\left(- \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma} \right) - \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma}. \tag{2.45}$$

As estimativas de máxima verossimilhança de Ψ são obtidos resolvendo numericamente e simultaneamente o sistema de equações gerado de:

$$\left(\frac{\partial}{\partial \mu} \ell\left(\Psi \mid \mathbf{x}\right), \frac{\partial}{\partial \sigma} \ell\left(\Psi \mid \mathbf{x}\right), \frac{\partial}{\partial \alpha} \ell\left(\Psi \mid \mathbf{x}\right) \right) = 0 \tag{2.46}$$

2.2.6 Demais Generalizações

Encontra-se ainda na literatura outras generalizações da distribuição Gumbel, tais como: Gumbel exponenciada generalizada (CORDEIRO; ORTEGA; CUNHA, 2013), Gumbel estendida , kumaraswamy Gumbel (CORDEIRO; NADARAJAH; ORTEGA, 2012), beta Gumbel (NADARAJAH; KOTZ, 2004), exponencial gama (ADEYEMI; OJO, 2003). Entretanto essas distribuições não foram consideradas na comparação pois seus parâmetros são não identificaveis.

2.3 Resumo das Distribuições Candidatas

Na Tabela 1 é apresentado um resumo das funções densidades das distribuições generalizadas e seus casos particulares em referência a distribuição Gumbel.

Tabela 1 – Função densidade de probabilidade da distribuição Gumbel e algumas de suas generalizações.

Distribuição	Função densidade	Caso Particular
Gumbel	$f(x \mid \Theta) = \frac{1}{\sigma} e^{\left(-u-e^{-u}\right)},$	-
Gumbel Transmutada	$f\left(x \mid \Psi\right) = \frac{1}{\sigma} e^{\left(-u-e^{-u}\right)} \left[1+\alpha-2\alpha e^{\left(-e^{-u}\right)}\right]$	$\alpha = 0$
Gumbel Exponenciada	$f\left(x \mid \Psi\right) = \frac{\alpha}{\sigma} e^{\left(-u-e^{-u}\right)} \left[e^{\left(-e^{-u}\right)}\right]^{\alpha-1}$	$\alpha = 1$
Exponencial Gama com 3 parâmetros	$f\left(x \mid \Psi\right) = \frac{1}{\Gamma(\alpha)} \frac{1}{\sigma} \exp\left[-e^{-u}\right] e^{-\alpha u}$	$\alpha = 1$
Marshall-Olkin Gumbel	$f\left(x \mid \Psi\right) = \dfrac{\alpha \frac{1}{\sigma} e^{\left(-u-e^{-u}\right)}}{\left(1-(1-\alpha)\left[1-e^{-e^{-u}}\right]\right)^2}$	$\alpha = 1$
Gambel Generalizada	$f\left(x \mid \Psi\right) = \frac{1}{\sigma} \left[1+\frac{1}{\alpha} e^{-u}\right]^{-\alpha-1} e^{-u}$	$\alpha \to \infty$

em que $u = \left(\dfrac{x-\mu}{\sigma}\right)$.

2.4 Aplicação em Dados de Literatura

Nesta aplicação consideramos os dados da literatura que descrevem os recordes máximos anuais do nível do mar da cidade de Porto Pirie, no Sul da Austrália no período de 1923 a 1987. Podemos observar a série na Tabela 2, onde são dispostas as 65 observações referentes ao nível máximo anual do mar, ou seja, a maior altitude média da superfície do mar medida em metros para cada ano do período. Esses dados estão disponíveis na biblioteca *ismev* (HEFFERNAN; STEPHENSON, 2012) do software R. Coles et al. (2001), também utilizaram os dados em uma aplicação ajustando a distribuição Valor Extremo generalizada. Davison et al. (2009), por outro lado, também analisaram esses dados por meio da distribuição

Valor Extremo Generalizada porém utilizaram métodos bayesianos.

Tabela 2 – Recordes máximos anuais do nível do mar da cidade de Porto Pirie nos anos de 1923 a 1987.

Ano	Nível do Mar	Ano	Nível do Mar	Ano	Nível do Mar	Ano	Nível do Mar	Ano	Nível do Mar
1923	4.03	1936	3.96	1949	4.06	1962	4.11	1975	3.91
1924	3.83	1937	3.85	1950	3.71	1963	4.24	1976	3.72
1925	3.65	1938	3.93	1951	3.96	1964	3.96	1977	4.00
1926	3.88	1939	3.75	1952	4.06	1965	4.21	1978	3.66
1927	4.01	1940	3.63	1953	4.55	1966	3.74	1979	3.62
1928	4.08	1941	3.57	1954	3.79	1967	3.85	1980	4.33
1929	4.18	1942	4.25	1955	3.89	1968	3.88	1981	4.55
1930	3.80	1943	3.97	1956	4.11	1969	3.66	1982	3.75
1931	4.36	1944	4.05	1957	3.85	1970	4.11	1983	4.08
1932	3.96	1945	4.24	1958	3.86	1971	3.71	1984	3.90
1933	3.98	1946	4.22	1959	3.86	1972	4.18	1985	3.88
1934	4.69	1947	3.73	1960	4.21	1973	3.90	1986	3.94
1935	3.85	1948	4.37	1961	4.01	1974	3.78	1987	4.33

Antes da escolha e ajuste de uma distribuição de probabilidade, foi avaliado a validade das suposições de que a série é independente e identicamente distribuída. Esta suposição é necessária, por exemplo, para estimar os parâmetros das distribuições candidatas pelo método da máxima verossimilhança. Nesta aplicação as suposições são avalidas a partir dos valores das estatísticas dos testes de Bartels (BARTELS, 1982) e Wald-Wolfwitz (WALD; WOLFOWITZ, 1943), onde os valores-p foram, respectivamente, $0,91$ e $0,11$. Esses valores-p não rejeitam a hipótese de independencia e aleatoriedade.

Um resumo das estimativas de máxima verossimilhança de cada parâmetro das distribuições consideradas para a análise são expostos na Tabela 3. Nesta tabela, também são apresentados os erros padrão e os intervalos de confiança de 95%. Pode-se então observar que o intervalo de confiança para o parâmetro de forma, α, possuem o valor que as fazem coincidir com a distribuição Gumbel em todas as ditribuições, à exceção da distribuição Gumbel exponenciada. Ou seja, existe evidências que a distribuição Gumbel possui uma boa perfomance no ajuste dos dados analisados.

Tabela 3 – Estimativas de máxima verossimilhança, erros padrão e intervalo de confiança para as distribuições candidatas.

				Intervalo de Confiança 95%	
Distribuição	Parâmetro	Estimativa	Erro Padrão	Limite Inferior	Limite Superior
1	μ	3.8694	0.0255	3.8195	3.9194
	σ	0.1949	0.0188	0.1579	0.2318
2	μ	3.8407	0.0751	3.6935	3.9879
	σ	0.1864	0.0233	0.1407	0.2321
	α	-0.2554	0.6283	-1.4867	0.9760
3	μ	3.8090	0.1034	3.6064	4,0117
	σ	0.1790	0.0358	0.1090	0.2491
	α	1.7594	1.7464	0.6635	5.1823
4	μ	3.7838	0.0233	3.7381	3.8294
	σ	0.1949	0.0188	0.1579	0.2318
	α	0.4395	0.0045	0.4306	0.4484
5	μ	3.9402	0.2530	3.4443	4.4362
	σ	0.2274	0.1125	0.0069	0.4479
	α	1.2926	1.0555	-0.7761	3.3613
6	μ	3.8707	0.0280	3.8158	3.9256
	σ	0.1940	0.0206	0.1536	0.2343
	α	116.7854	1,0940	114,6412	118,9296

1: Gumbel; 2: Gumbel Transmutada; 3:Marshall-Olkin Gumbel; 4: Gumbel Exponenciada; 5: Exponencial Gama com 3 parâmetros; 6: Gumbel Generalizada

A avaliação dos ajustes dos parâmetros foi realizada por meio das estatísticas Kolmogorov-Smirnov (KS), Anderson-Darling (AD) e Cramér-von Mises (CvM) e é exibida na Tabela 4. Segundo esses critérios pode-se observar que todas as distribuições candidatas ajustaram a série de dados, considerando o nível de significância 5%.

Tabela 4 – Valores-p das estatísticas usadas como critério de ajuste.

Distribuição	KS	AD	CvM
Gumbel	0,8894	0,9967	0,9923
Gumbel Transmutada	0,9979	0,9967	0,9952
Marshall-Olkin Gumbel	0,9657	0,9983	0,9958
Gumbel Exponenciada	0,4678	0,1934	0,2089
Gama Exponenciada com 3 parâmetros	0,9476	0,9982	0,9957
Gumbel Generalizada	0,9892	0,9990	0,9928

Para decidir pela distribuição que melhor se ajusta a série de dados, foram utilizados critérios de discriminação: -2ℓ, critério de informação Akaike (AIC), critério de informação

Akaike corrigido (AICc) e critério de informação Bayesiano (BIC), para todos os critérios decide-se em favor da distribuição que apresenta o menor valor da estatística (AKAIKE, 1974; BOZDOGAN, 1987; SCHWARZ, 1978). Os resultados obtidos podem ser observados na Tabela 5, cada critério teve as estatísticas ranqueadas do menor para o maior valor, esse ranque esta na forma de um número sobrescrito em cada estatística. Portanto a coluna R, é o resultado da soma dos ranques para cada distribuição, assim a distribuição que possuir o menor ranque em R é a distribuição que melhor se ajusta a série.

Tabela 5 – Valores das estatísticas usadas como critério de discriminação.

Distribuição	-2ℓ	AIC	AICc	BIC	R
Gumbel	4,2177^{2}	-4,4354^{1}	-4,2418^{1}	-0,0866^{1}	07^{1}
Gumbel Transmutada	4,2882^{4}	-2,5764^{2}	-2,1829^{5}	3,9468^{3}	14^{2}
Marshall-Olkin Gumbel	4,3662^{5}	-2,5764^{2}	-2,3390^{6}	3,7907^{2}	15^{3}
Gumbel Exponenciada	4,2177^{2}	-2,4354^{4}	-2,0419^{3}	4,0878^{5}	14^{2}
Gama Exponenciada com 3 parâmetros	4,2718^{3}	-2,5437^{3}	-2,1503^{4}	3,9795^{4}	15^{3}
Gumbel Generalizada	4,2128^{1}	-2,4256^{5}	-2,0321^{2}	4,0975^{6}	14^{2}

Finalmente, pode-se então concluir que dentre todas as distribuições utilizadas para a análise, a distribuição Gumbel foi a que teve a melhor perfomance segundo os critérios adotados $(R = 7)$. Observa-se então que a distribuição Gumbel, que possui apenas dois parâmetros, evidenciou que suas possiveis generalizações não colaboraram em nada no ajuste.

Capítulo 3

Modelando a Precipitação Máxima Mensal da Região Sul do Brasil

3.1 Introdução

A previsão do tempo sempre foi um elemento de grande importância para a sociedade, pois várias atividades econômicas dependem das condições climáticas. Portanto o conhecimento da distribuição de variáveis climáticas são extremamente importantes, uma vez que a intensidade, a frequência e a duração dos fenômenos climáticos influenciam diretamente a economia.

O comportamento de dados climatológicos pode estar associado a diversos fatores ambientais, que em geral são difíceis de serem representados matematicamente. Portanto, conhecer previamente o comportamento e a distribuição da precipitação pluvial, contribui para estimar a probabilidade de ocorrência de eventos extremos que são característicos de uma determinada região e assim possivelmente minimizar suas consequências (BISPO, 2008).

A modelagem de dados de natureza climatólogica geralmente utiliza as distribuições Weibull, Gama, Log-Normal, Log-Logística, Gumbel e Fréchet (WMO, 2009). Dentre alguns trabalhos que utilizam estas distribuições pode-se citar: Vieira et al. (2010) estudou o comportamento da precipitação pluvial mensal no período da estação chuvosa na região de Diamantina/MG, por meio da distribuição Weibull; Aquino et al. (2012) analisaram séries quinzenais de precipitação pluvial de uma estação localizada na cidade de Pelotas/RS, as distribuições Gama e Log-Normal foram ajustadas afim de estimar a precipitação mensal mínima; Silva et al. (2007) ajustaram as distribuições Gama, Weibull, Normal, Log-Normal e

Exponencial, a séries diárias do município de Santa Maria/RS, com o objetivo de analisar a quantidade de precipitação e o número de dias com chuva utilizando as distribuições Gama e Weibull. Hartmann, Moala e Mendonça (2011) analisaram dados de precipitação pluvial máxima mensal, da estação meteorológica de Presidente Prudente/SP ajustando a distribuição Gumbel com o objetivo de estimar a precipitação máxima esperada para diferentes níveis de probabilidade e verificar o grau de ajuste dos dados ao modelo proposto; Santos et al. (2014) analisaram as 48 séries históricas dos valores máximos diários de precipitação dos registros de uma estação meteorológica em Mossoró/RN, também com o objetivo de estimar a precipitação máxima diária provável para diferentes períodos de retorno.

3.2 Os Dados

Essa aplicação consideraou séries históricas de 33 estações meteorológicas localizadas na região Sul do Brasil: 8 estações no estado do Paraná, 7 estações em Santa Catarina e 18 estações no Rio Grande do Sul, disponíveis no "Banco de dados meteriológicos para educação e pesquisa" do Instituto Nacional de Meteriologia (<http://www.inmet.gov.br>). As séries analisadas consistem na observação do valor máximo de precipitação pluvial mensal segundo o total observado em um período de 24 horas, ou seja, foram observados o volume total de cada dia e registrado a precipitação máxima mensal no período de 1961 à 2015.

Uma análise descritiva para cada série é apresentada na Tabela 6. A saber, código da estação (Estação), tamanho das séries (n), média aritmética das precipitações (Média), Desvio Padrão (Desvio Padrão), a menor quantidade constatada (Mínimo), a maior quantidade constatada (Máximo), valor mediano de precipitação (Mediana), coeficiente de assimetria (Assimetria) e o coeficiente de curtose (Curtose). Com base no coeficiente de assimetria todas as séries são consideradas assimetricas positivas, ou assimetrica à direita, devido ao coeficiente ter valor superior a zero. O coeficiente de curtose também mostra que todas as séries são platicurticas, ou seja, a distribuição apresenta uma curva de frequência achatada em sua parte superior, pois todos seus coeficientes resultaram em um valor superior a $0,263$.

3.3 Resultados e Discussões

Para avaliar a validade das suposições de que as séries são independentes e identicamente distribuídas foi realizado os testes de Wald-Wolfwitz e Bartels, ambos os testes

Tabela 6 – Medidas descritivas para cada série

Estação	n	Média	Desvio Padrão	Mínimo	Máximo	Mediana	Assimetria	Curtose
83766	521	42.2923	26.4417	0.2000	231.4000	38.4000	1.6299	6.5427
83767	434	44.0134	25.6829	0.2000	151.5000	41.0000	0.9352	1.2835
83783	545	44.9716	26.9501	0.3000	191.9000	40.9000	1.3276	3.4226
83811	384	42.8281	23.6585	0.4000	184.0000	38.8000	1.2575	3.7500
83813	502	39.1243	21.5377	0.5000	153.6000	35.5500	1.1070	2.2810
83836	510	40.7525	22.6035	0.1000	175.0000	37.5500	1.3891	4.1922
83842	569	39.6125	21.7897	0.7000	146.2000	36.2000	1.0156	1.6620
83844	544	49.9987	34.1806	0.5000	295.8000	41.9500	1.8753	6.4080
83881	476	50.5435	27.2245	4.0000	200.0000	45.0000	1.2840	2.8473
83907	503	54.0853	29.9222	4.5000	189.8000	49.3000	1.1740	2.1880
83912	454	48.5936	25.4340	3.8000	142.8000	43.7500	0.9872	1.1624
83914	501	49.0242	25.9789	3.3000	174.0000	43.6000	1.2299	2.3626
83916	390	45.7479	23.8689	1.3000	166.0000	41.4500	1.1278	2.4251
83919	434	42.9048	22.6022	4.6000	151.9000	39.3500	1.3049	2.6612
83927	454	48.4097	31.9917	0.9000	183.2000	40.4500	1.2590	1.6430
83936	508	48.5618	26.2503	2.4000	183.9000	43.5000	1.3353	3.2279
83941	409	43.2254	22.5927	5.6000	126.0000	38.8000	1.1162	1.4203
83942	461	44.6334	22.0592	4.1000	136.7000	41.2000	1.0241	1.0631
83948	476	40.0475	27.4518	2.7000	257.3000	34.6000	2.5655	12.5528
83953	267	45.2925	26.9610	4.0000	163.8000	41.0000	1.1073	1.4065
83964	483	45.8133	24.7482	0.9000	184.0000	42.3000	1.2689	2.8863
83967	566	38.0461	20.3120	0.3000	149.6000	34.9500	1.1015	2.3529
83980	518	45.0898	25.9901	0.8000	181.0000	40.3500	1.1326	2.0273
83985	328	41.6369	24.9461	1.6000	216.8000	36.5000	1.8755	8.0297
83995	460	40.6289	26.0424	1.4000	194.0000	35.2000	1.7179	5.2687
83997	483	40.2232	26.5896	1.0000	167.6000	34.8000	1.5427	3.7094

apresentaram valores-p superiores a 0.05, portanto não rejeitam a hipótese de independencia e aleatoriedade.

Seguindo o mesmo processo da aplicação em dados de literatura, a avaliação dos ajustes das distribuições foi realizada por meio das estatísticas Kolmogorov-Smirnov (KS), Anderson-Darling (AD) e Cramér-von Mises (CvM). Uma ilustração da distribuição dos valores-p para cada estatística pode ser observada nas Figuras 7, 8 e 9. É importante ressaltar que todas as séries foram ajustadas com o nível de significância de 5%.

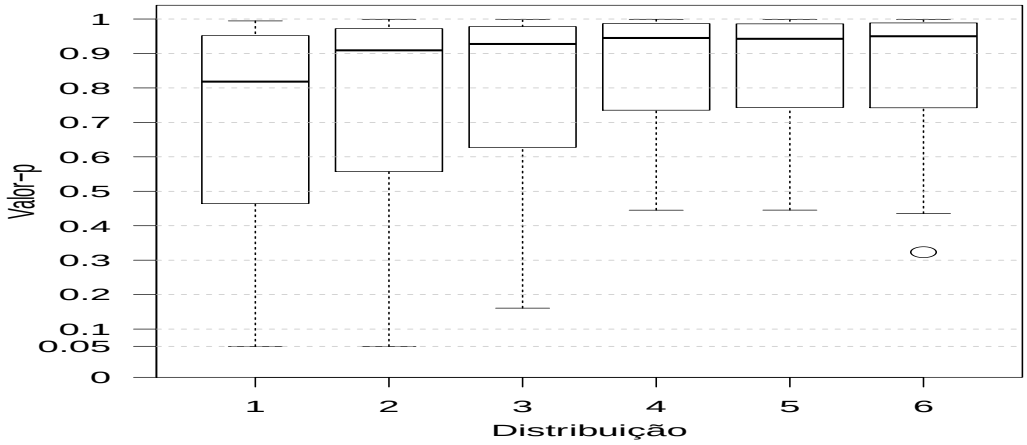

Figura 7 – Distribuição dos valores-p segundo a estatística de Anderson Darling, (1: Gumbel; 2: Gumbel Exponenciada; 3: Marshall-Olkin Gumbel; 4: Gumbel Transmutada; 5: Gumbel Generalizada; 6: Exponencial Gama com 3 parâmetros).

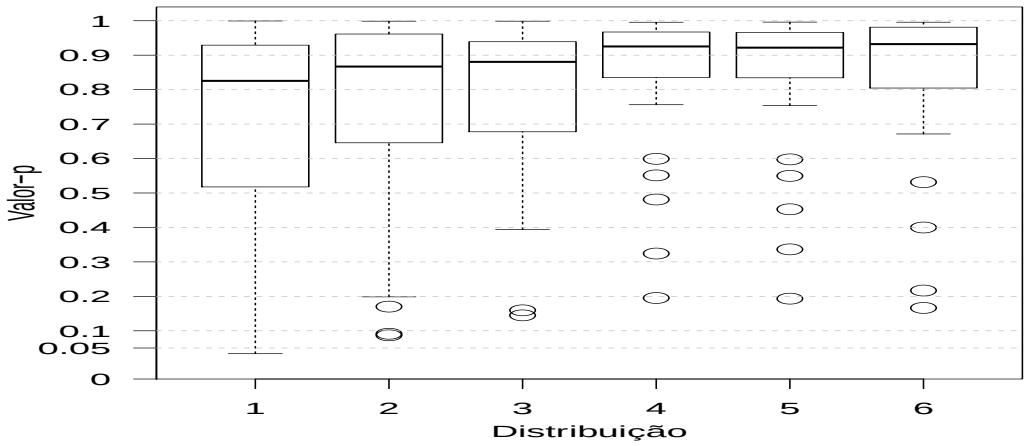

Figura 8 – Distribuição dos valores-p segundo a estatística de Kolmogorov Smirnov, (1: Gumbel; 2: Gumbel Exponenciada; 3: Marshall-Olkin Gumbel; 4: Gumbel Transmutada; 5: Gumbel Generalizada; 6: Exponencial Gama com 3 parâmetros).

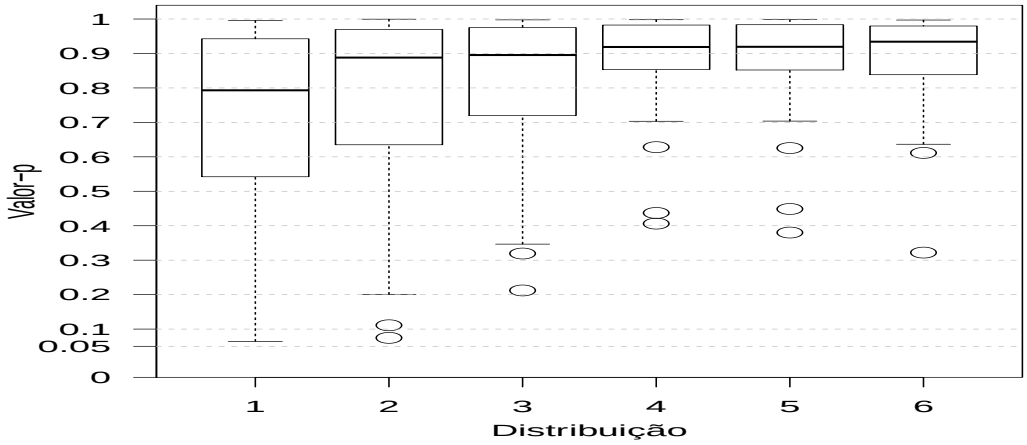

Figura 9 – Distribuição dos valores-p segundo a estatística de Cramér von Mises, (1: Gumbel; 2: Gumbel Exponenciada; 3: Marshall-Olkin Gumbel; 4: Gumbel Transmutada; 5: Gumbel Generalizada; 6: Exponencial Gama com 3 parâmetros).

As Tabelas 7, 8 e 9 exibem o valor das estatísticas dos respectivos critérios: critério de informação Akaike (AIC), critério de informação Akaike corrigido (AICc) e Critério de informação Baysiano (BIC), para todos os critérios decide-se em favor da distribuição que apresenta o menor valor da estatística. Na Tabela 10, pode-se observar o resumo das somas dos ranks. A distribuição com o menor rank é considerada a distribuição com melhor performance, ou seja, a distribuição Gumbel obteve menor soma dos ranks segundo os três critérios adotados.

Tabela 10 – Resumo das somas dos Ranks

Critérios	1	2	3	4	5	6
AIC	63	131	113	74	90	75
$AICc$	63	131	113	74	90	75
BIC	44	133	115	79	95	80

1: Gumbel; 2: Gumbel Exponenciada; 3: Marshall Olkin Gumbel; 4: Gumbel Transmutada; 5: Gumbel Transmutada;6: Exponencial Gama com 3 parâmetros

Finalmente, pode-se então concluir que dentre todas as distribuições utilizadas para a análise, a distribuição Gumbel foi a que teve a melhor perfomance segundo os critérios

Tabela 7 – Valores da estatística do critério de ajuste AIC.

Estação	1	2	3	4	5	6
83766	4771.481^{1}	4773.346^{3}	4773.311^{2}	4773.361^{4}	4773.372^{5}	4773.441^{6}
83767	3988.878^{1}	3990.492^{3}	3990.046^{2}	3990.870^{6}	3990.853^{5}	3990.673^{4}
83783	5023.997^{1}	5025.989^{6}	5025.989^{5}	5025.793^{2}	5025.801^{3}	5025.912^{4}
83811	3457.569^{1}	3459.569^{6}	3458.570^{3}	3458.709^{4}	3458.779^{5}	3457.957^{2}
83813	4424.464^{1}	4426.357^{3}	4426.395^{6}	4426.380^{4}	4426.385^{5}	4426.353^{2}
83836	4526.661^{1}	4526.974^{2}	4528.309^{4}	4528.559^{5}	4528.576^{6}	4528.142^{3}
83842	5035.443^{1}	5036.816^{2}	5037.013^{3}	5037.228^{5}	5037.230^{6}	5037.143^{4}
83844	5179.928^{6}	5177.491^{5}	5166.856^{4}	5149.744^{1}	5149.928^{2}	5153.685^{3}
83881	4388.317^{4}	4389.886^{6}	4389.320^{5}	4387.001^{2}	4387.144^{3}	4386.499^{1}
83907	4745.704^{1}	4747.582^{6}	4747.464^{5}	4746.476^{3}	4746.548^{4}	4746.164^{2}
83912	4151.143^{1}	4153.138^{6}	4153.134^{5}	4152.743^{3}	4152.790^{4}	4152.378^{2}
83914	4577.010^{1}	4578.442^{6}	4577.873^{5}	4577.316^{2}	4577.324^{3}	4577.452^{4}
83916	3513.515^{1}	3515.515^{6}	3515.514^{5}	3515.284^{3}	3515.329^{4}	3514.522^{2}
83919	3836.543^{4}	3837.970^{6}	3837.486^{5}	3836.087^{1}	3836.141^{2}	3836.440^{3}
83927	4305.738^{5}	4307.738^{6}	4299.488^{4}	4285.065^{2}	4285.111^{3}	4283.093^{1}
83936	4648.004^{1}	4649.486^{6}	4649.261^{4}	4648.948^{3}	4648.940^{2}	4649.305^{5}
83941	3623.687^{4}	3625.138^{6}	3624.353^{5}	3622.334^{2}	3622.433^{3}	3621.969^{1}
83942	4071.281^{4}	4072.625^{6}	4071.559^{5}	4070.992^{2}	4071.007^{3}	4070.177^{1}
83948	4285.556^{6}	4281.454^{5}	4271.621^{3}	4270.209^{2}	4270.201^{1}	4278.436^{4}
83953	2459.770^{4}	2461.770^{6}	2460.579^{5}	2458.110^{2}	2458.241^{3}	2457.648^{1}
83964	4372.107^{1}	4373.486^{2}	4374.046^{5}	4374.006^{4}	4373.992^{3}	4374.105^{6}
83967	4918.926^{1}	4920.906^{6}	4920.884^{5}	4920.750^{3}	4920.796^{4}	4920.043^{2}
83980	4744.682^{1}	4746.682^{6}	4746.588^{5}	4745.294^{3}	4745.416^{4}	4744.709^{2}
83985	2948.167^{1}	2949.391^{4}	2949.630^{5}	2949.047^{3}	2949.044^{2}	2949.695^{6}
83995	4155.950^{5}	4156.195^{6}	4152.977^{4}	4145.173^{1}	4145.480^{2}	4145.952^{3}
83997	4390.542^{5}	4391.045^{6}	4387.687^{4}	4371.964^{2}	4372.676^{3}	4369.976^{1}
Total	63	131	113	74	90	75

1: Gumbel; 2: Gumbel Exponenciada; 3: Marshall Olkin Gumbel; 4: Gumbel Transmutada; 5: Gumbel Generalizada; 6: Exponencial Gama com 3 parâmetros

Tabela 8 – Valores da estatística do critério de ajuste AIC corrigido.

Estação	1	2	3	4	5	6
83766	4771.504^{1}	4773.393^{3}	4773.358^{2}	4773.408^{4}	4773.418^{5}	4773.487^{6}
83767	3988.906^{1}	3990.548^{3}	3990.102^{2}	3990.926^{6}	3990.909^{5}	3990.729^{4}
83783	5024.019^{1}	5026.034^{6}	5026.034^{5}	5025.838^{2}	5025.845^{3}	5025.956^{4}
83811	3457.601^{1}	3459.632^{6}	3458.634^{3}	3458.772^{4}	3458.842^{5}	3458.020^{2}
83813	4424.488^{1}	4426.405^{3}	4426.443^{6}	4426.428^{4}	4426.433^{5}	4426.401^{2}
83836	4526.685^{1}	4527.021^{2}	4528.356^{4}	4528.606^{5}	4528.624^{6}	4528.189^{3}
83842	5035.464^{1}	5036.859^{2}	5037.056^{3}	5037.270^{5}	5037.273^{6}	5037.186^{4}
83844	5179.950^{6}	5177.535^{5}	5166.900^{4}	5149.789^{1}	5149.972^{2}	5153.729^{3}
83881	4388.342^{4}	4389.937^{6}	4389.371^{5}	4387.052^{2}	4387.195^{3}	4386.550^{1}
83907	4745.728^{1}	4747.630^{6}	4747.513^{5}	4746.524^{3}	4746.596^{4}	4746.212^{2}
83912	4151.170^{1}	4153.191^{6}	4153.187^{5}	4152.796^{3}	4152.843^{4}	4152.432^{2}
83914	4577.034^{1}	4578.490^{6}	4577.921^{5}	4577.365^{2}	4577.372^{3}	4577.500^{4}
83916	3513.546^{1}	3515.577^{6}	3515.576^{5}	3515.347^{3}	3515.391^{4}	3514.584^{2}
83919	3836.571^{4}	3838.026^{6}	3837.542^{5}	3836.143^{1}	3836.197^{2}	3836.496^{3}
83927	4305.765^{5}	4307.791^{6}	4299.541^{4}	4285.118^{2}	4285.164^{3}	4283.146^{1}
83936	4648.028^{1}	4649.533^{6}	4649.309^{4}	4648.996^{3}	4648.987^{2}	4649.353^{5}
83941	3623.717^{4}	3625.198^{6}	3624.413^{5}	3622.394^{2}	3622.492^{3}	3622.028^{1}
83942	4071.307^{4}	4072.677^{6}	4071.611^{5}	4071.045^{2}	4071.060^{3}	4070.229^{1}
83948	4285.581^{6}	4281.505^{5}	4271.672^{3}	4270.260^{2}	4270.252^{1}	4278.487^{4}
83953	2459.816^{4}	2461.862^{6}	2460.670^{5}	2458.201^{2}	2458.332^{3}	2457.739^{1}
83964	4372.132^{1}	4373.536^{2}	4374.096^{5}	4374.056^{4}	4374.042^{3}	4374.155^{6}
83967	4918.947^{1}	4920.949^{6}	4920.927^{5}	4920.793^{3}	4920.838^{4}	4920.086^{2}
83980	4744.706^{1}	4746.729^{6}	4746.634^{5}	4745.340^{3}	4745.463^{4}	4744.756^{2}
83985	2948.204^{1}	2949.465^{4}	2949.704^{5}	2949.121^{3}	2949.118^{2}	2949.769^{6}
83995	4155.977^{5}	4156.248^{6}	4153.030^{4}	4145.226^{1}	4145.533^{2}	4146.005^{3}
83997	4390.567^{5}	4391.095^{6}	4387.737^{4}	4372.014^{2}	4372.726^{3}	4370.026^{1}
Total	63	131	113	74	90	75

1: Gumbel; 2: Gumbel Exponenciada; 3: Marshall Olkin Gumbel; 4: Gumbel Transmutada; 5: Gumbel Generalizada; 6: Exponencial Gama com 3 parâmetros

Tabela 9 – Valores da estatística do critério de ajuste BIC.

Estação	1	2	3	4	5	6
83766	4779.992^{1}	4786.114^{3}	4786.079^{2}	4786.129^{4}	4786.139^{5}	4786.208^{6}
83767	3997.024^{1}	4002.712^{3}	4002.266^{2}	4003.090^{6}	4003.072^{5}	4002.892^{4}
83783	5032.599^{1}	5038.892^{6}	5038.892^{5}	5038.696^{2}	5038.703^{3}	5038.814^{4}
83811	3465.470^{1}	3471.421^{6}	3470.422^{3}	3470.561^{4}	3470.631^{5}	3469.809^{2}
83813	4432.901^{1}	4439.012^{3}	4439.051^{6}	4439.035^{4}	4439.041^{5}	4439.009^{2}
83836	4535.130^{1}	4539.677^{2}	4541.012^{4}	4541.262^{5}	4541.280^{6}	4540.845^{3}
83842	5044.131^{1}	5049.848^{2}	5050.045^{3}	5050.260^{5}	5050.262^{6}	5050.175^{4}
83844	5188.526^{5}	5190.387^{6}	5179.753^{4}	5162.641^{1}	5162.825^{2}	5166.581^{3}
83881	4396.648^{1}	4402.382^{6}	4401.816^{5}	4399.498^{3}	4399.640^{4}	4398.995^{2}
83907	4754.145^{1}	4760.244^{6}	4760.126^{5}	4759.137^{3}	4759.210^{4}	4758.825^{2}
83912	4159.379^{1}	4165.492^{6}	4165.488^{5}	4165.097^{3}	4165.144^{4}	4164.733^{2}
83914	4585.443^{1}	4591.092^{6}	4590.522^{5}	4589.966^{2}	4589.974^{3}	4590.102^{4}
83916	3521.447^{1}	3527.413^{6}	3527.413^{5}	3527.183^{3}	3527.227^{4}	3526.421^{2}
83919	3844.689^{1}	3850.189^{6}	3849.705^{5}	3848.306^{2}	3848.360^{3}	3848.660^{4}
83927	4313.974^{5}	4320.092^{6}	4311.842^{4}	4297.419^{2}	4297.465^{3}	4295.447^{1}
83936	4656.465^{1}	4662.177^{6}	4661.953^{4}	4661.640^{3}	4661.631^{2}	4661.996^{5}
83941	3631.714^{1}	3637.179^{6}	3636.394^{5}	3634.375^{3}	3634.474^{4}	3634.010^{2}
83942	4079.548^{1}	4085.025^{6}	4083.959^{5}	4083.393^{3}	4083.407^{4}	4082.577^{2}
83948	4293.887^{5}	4293.950^{6}	4284.117^{3}	4282.705^{2}	4282.698^{1}	4290.932^{4}
83953	2466.945^{1}	2472.532^{6}	2471.340^{5}	2468.872^{3}	2469.002^{4}	2468.410^{2}
83964	4380.467^{1}	4386.026^{2}	4386.586^{5}	4386.546^{4}	4386.532^{3}	4386.645^{6}
83967	4927.603^{1}	4933.922^{6}	4933.900^{5}	4933.766^{3}	4933.811^{4}	4933.059^{2}
83980	4753.182^{1}	4759.432^{6}	4759.337^{5}	4758.044^{3}	4758.166^{4}	4757.459^{2}
83985	2955.753^{1}	2960.770^{4}	2961.009^{5}	2960.426^{3}	2960.423^{2}	2961.074^{6}
83995	4164.213^{4}	4168.589^{6}	4165.371^{5}	4157.567^{1}	4157.874^{2}	4158.346^{3}
83997	4398.902^{4}	4403.585^{6}	4400.227^{5}	4384.504^{2}	4385.216^{3}	4382.516^{1}
Total	44	133	115	79	95	80

1: Gumbel; 2: Gumbel Exponenciada; 3: Marshall Olkin Gumbel; 4: Gumbel Transmutada; 5: Gumbel Generalizada; 6: Exponencial Gama com 3 parâmetros

adotados. Evidenciando novamente que as generalizações analisadas da distribuição Gumbel não colaboraram na performance do ajuste neste trabalho.

3.3.1 Período de Retorno

Uma vez que distribuição Gumbel forneceu um melhor ajuste para as séries, pode-se então calcular o período de retorno por meio de sua função de probabilidade. Entede-se por período de retorno, o intervalo médio de tempo que separam um evento de dimensão conhecida de outro evento com dimensão igual ou superior. E pode ser obtido pela expressão:

$$\tau = \frac{1}{1 - F(x)} \tag{3.1}$$

Na Tabela 11 é apresentada a precipitação máxima mensal esperadas para cada estação para os próximos 5, 10, 15, 20, 25 e 30 anos. Para cada ano então é esperado a quantidade descrita na Tabela 11, ou uma quantidade superior.

Tabela 11 – Precipitação máxima mensal (mm) esperadas para cada estação para os próximos 5, 10, 15, 20, 25 e 30 anos.

	Período de Retorno					
Estação	5	10	15	20	25	30
83766	60.3620	75.5967	84.1921	90.2103	94.8459	98.6171
83767	63.3706	79.0028	87.8223	93.9974	98.7540	102.6235
83783	63.6948	79.2745	88.0644	94.2188	98.9594	102.8159
83811	60.5755	74.6789	82.6358	88.2071	92.4985	95.9896
83813	54.9086	67.6316	74.8098	79.8358	83.7072	86.8566
83836	56.8464	69.9294	77.3107	82.4789	86.4598	89.6983
83842	55.8622	68.8677	76.2054	81.3430	85.3003	88.5197
83844	65.2829	81.1462	90.0962	96.3628	101.1897	105.1165
83872	60.2271	74.0273	81.8133	87.2648	91.4640	94.8801
83881	67.8174	82.7843	91.2284	97.1408	101.6949	105.3998
83883	71.2887	86.2523	94.6946	100.6057	105.1588	108.8628
83887	69.2744	84.1808	92.5909	98.4794	103.0151	106.7050
83891	55.9745	68.4709	75.5212	80.4577	84.2601	87.3535
83897	55.8215	69.2784	76.8706	82.1865	86.2812	89.6123
83907	74.2022	91.1412	100.6981	107.3896	112.5438	116.7369
83912	66.5415	81.3397	89.6887	95.5345	100.0373	103.7004
83914	66.1102	80.6267	88.8167	94.5511	98.9682	102.5616
83916	62.5414	76.3793	84.1865	89.6530	93.8635	97.2890
83919	57.2741	69.6795	76.6786	81.5791	85.3538	88.4246
83920	55.5911	67.6590	74.4676	79.2348	82.9069	85.8941
83923	55.7922	69.2073	76.7759	82.0753	86.1572	89.4780
83927	64.2663	80.1122	89.0523	95.3120	100.1335	104.0560
83936	65.9500	80.6276	88.9086	94.7068	99.1729	102.8062
83941	59.4792	73.0704	80.7384	86.1074	90.2429	93.6073
83942	59.2415	71.6784	78.6952	83.6081	87.3924	90.4710
83948	52.7435	65.2808	72.3542	77.3068	81.1217	84.2252
83953	61.8142	76.4800	84.7544	90.5478	95.0103	98.6407
83964	62.9623	77.1158	85.1010	90.6921	94.9987	98.5023
83967	52.5068	64.3609	71.0488	75.7315	79.3385	82.2728
83980	62.6276	77.3906	85.7197	91.5516	96.0436	99.6980
83985	57.9806	72.1413	80.1307	85.7246	90.0334	93.5388
83995	54.4044	67.4297	74.7784	79.9238	83.8871	87.1114
83997	53.6337	66.7839	74.2032	79.3979	83.3993	86.6545

Capítulo 4

Modelando o Comprimento dos Dias Chuvosos Considerando Distribuições Poisson-Lindley

4.1 Introdução

Tão importante quanto a análise do volume da precipitação pluviométrica acumulado em um determinado período de tempo é a análise do comprimento das sequências de dias chuvosos. Um conjunto de dias chuvosos consecutivos constitui um comprimento de dias chuvosos. O comprimento da sequencia de dias chuvosos pode ser definido como uma sequência de dias chuvosos incluindo dias com valor maior que o limiar de precipitação. Um dia é considerado chuvoso se a precipitação acumulada em um período de 24 horas for simplesmente mensurável ou superior a algum limiar fixado a priori, por exemplo 0.1, 1.0, 5.0 ou 10.0 milímetros Shaw et al. (2010).

O valor do limiar usado para definir um dia como chuvoso varia bastante na literatura. Por exemplo, a fim de detectar seus impactos no comprimento das sequências de dias chuvosos, em (TOLIKA; MAHERAS, 2005) são considerados limiares de 0.1 e 1.0 milímetro/dia já em (DASH et al., 2011) é adotado um limiar de 2.5 mil ímetros/dia.

Em (CINDRIC; PASARIC; GAJIC-CAPKA, 2010) são considerados os valores 0.1, 1.0, 5.0 e 10.0 milímetros/dia como limiares. Os limiares de 0.1, 1.0, 5.0 mílimetros/dia foram considerados em (LANA et al., 2006).

Independente do limiar adotado, a estimação da probabilidade de ocorrer sequências de dias chuvosos de comprimento k, $k \geq 1$, é extremamente útil para planejamento e gestão na agricultura, meio ambiente e em muitos outros setores. É sabido que longos períodos de chuva ou de seca afetam de maneira significativa produções agrícolas e energéticas. Como apontado em (ZEKÂI, 2015), todas as informações sobre o comprimento de períodos secos e chuvosos torna-se mais iminente em muitos campos, como em projetos de gestão de recursos hídricos, planejamento agrícola, estudos de inundação e em diversas atividades industriais.

A análise do comprimento da sequência de uma variável climatológica em um determinado estado não se restringe a precipitação e a descrição de seu comportamento tem uma série de implicações práticas. Por exemplo, longos períodos com baixa temperatura aumentam o consumo de energia e também afetam a produção de leite e o consumo de ração McCalla, Day e Millward (1978).

Basicamente duas suposições excludentes são impostas na análise do comprimento das sequências dos dias chuvosos e secos Mathugama e Peiris (2011). A primeira considera que o estado — chuvoso ou seco — em um determinado dia, depende do que ocorreu no dia ou dias anteriores (veja, por exemplo, (SONNADARA; JAYEWARDENE, 2015)). A segunda, considerada neste artigo, que os comprimentos das sequências num certo estado são indepententes e identicamente distribuidos. Sob estas suposições a análise se inicia pela escolha e ajuste de uma ou mais distribuições discretas de probabilidade com suporte no intervalo $[1, L]$. O valor de L depende do período considerado, por exemplo, se o objetivo é analisar os comprimentos das sequências mensais, nos meses de abril, junho, setembro e novembro o comprimento máximo observado será $L = 30$ dias.

Naturalmente, antes da escolha e ajuste de uma distribuição de probabilidade, a hipótese de que os comprimentos das sequências são independentes e identicamente distribuídos (isto é, formam uma amostra aleatória) deve ser testada. Esta suposição é necessária, por exemplo, para estimar os parâmetros das distribuições candidatas pelo método da máxima verossimilhança. Neste artigo as suposicões supracitadas são avalidas a partir dos valores das estatísticas dos testes de Bartels Bartels (1982), Wald-Wolfwitz Wald e Wolfowitz (1943) e Mann-Kendall Mann (1945) e Kendall e Gibbons (1990).

Várias funções de probabilidade tem sido introduzidas e utilizadas na descrição do comportamento dos comprimentos das sequências dos dias secos e chuvosos. (DENI; JEMAIN; IBRAHIM, 2008) ajusta sete distribuições aos comprimentos das sequências de dias chuvosos e secos observados entre 1971 e 2005 em várias estações meteorológicas da Ma-

laysia Peninsular. (DENI; JEMAIN, 2009a) e (DENI; JEMAIN; IBRAHIM, 2010) analisaram o comportamento dos comprimentos das sequências utilizando, respectivamente, 12 e 13 funções de probabilidade, dentre elas a mistura de distribuições logarítmicas, a mistura da logarítmica e Poisson e a mistura da logarítmica e geométrica. A mistura de duas distribuições também foram consideradas por (DOBI-WANTUCH; MIKA; SZEIDL, 2000), (DENI; JEMAIN, 2008), (DENI; JEMAIN; IBRAHIM, 2009), (DENI; JEMAIN, 2009b), (DENI; JEMAIN, 2012) dentre outros.

Apesar de nos últimos anos uma série de trabalhos terem utilizado misturas finitas de distribuições Titterington, Smith e Makov (1985), Everitt e Hand (1981) para a análise do comportamento dos comprimentos das sequências, essas não foram consideradas neste artigo. Justificamos o não uso de misturas devido, sem ser exaustivo, aos motivos: (i) dificuldades em estabelecer o número de componentes da mistura, (ii) dificuldades em interpertar os parâmetros da mistura, (iii) dificuldades em estabelecer quais distribuições devem ser misturadas (iv) dificuldades em encontrar as estimativas dos parâmetros em comparação a distribuições usuais (v) dificuldades em estabelecer a relação entre a variância empírica dos dados e a explicada pelos modelos de misturas e (vi) possível falta de identificabilidade dos parâmetros.

É importante ressaltar que os trabalhos recentes que utilizam misturas de distribuições não se atentam as dificuldades supracitadas. Entretanto, as mesmas podem ser superadas e são objetos de estudo em um outro artigo em execução.

Neste artigo consideramos algumas distribuições alternativas que no melhor de nosso conhecimento não foram consideradas na análise do comprimento das sequências dos dias secos e chuvosos. Estas distribuições são apresentadas na Seção 4.2. Os dados utilizados são descritos na Seção 4.3. Os resultados obtidos dos ajustes das distribuições são apresentados e discuticos na Seção 4.4. A Seção 4.5 finaliza este artigo.

4.2 Distribuições de Probabilidade

As próximas subseções apresentam alguns detalhes matemáticos sobre as distribuições de probabilidade discretas que estamos considerando para modelar o comprimento de sequências de dias chuvosos.

4.2.1 A Distribuição Binomial Negativa

A distribuição Binomial negativa (NBD) é talvez a distribuição estatística mais frequentemente aplicada para modelar os dados da contagem quando não podemos garantir equidispersão e ela pode ser obtida de muitas maneiras diferentes. Como apontado em (JAIN; CONSUL, 1971) a distribuição Binomial negativa é talvez a primeira distribuição de probabilidade cuja variância é maior do que sua média. Sua função massa de probabilidade (PMF), derivado de uma mistura de distribuição Poisson-Gama Hilbe (2011), pode ser escrita como:

$$P\left(X=x \mid \theta, \alpha\right)=\frac{\Gamma\left(x+\frac{1}{\alpha}\right)}{\Gamma\left(x+1\right)\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{\theta\alpha}{1+\theta\alpha}\right)^{x}\left(\frac{1}{1+\theta\alpha}\right)^{\frac{1}{\alpha}} \tag{4.1}$$

para $x=0,1,\ldots,$ $\alpha>0$ e $\theta>0$. A variável aleatória X tem média $E(X)=\theta$ e variância $V(X)=\theta\left(1+\alpha\theta\right)$ que excede a média. Na literatura Cameron e Trivedi (1998) é referido como a distribuição NEGBIN2 e α tende para 0 que têm a distribuição de Poisson e para $\alpha=1$ a distribuição Geométrica. O parâmetro α é o parâmetro de dispersão e sua estimação esta sendo estudada. A estimativa de máxima verossimilhança de α existe e é única se a variância amostral exceder a média amostral Aragón, Eberly e Eberley (1992). (WANG, 1996), (SAHA; PAUL, 2005) e (AL-KHASAWNEH, 2010) outras referências importantes sobre este assunto.

A NBD é naturalmente uma distribuição para modelar dados de contagem com sobredispersão desde que $V(X)$ for sempre maior que $E(X)$.No entanto, a NBD não pode ser aplicada diretamente para descrever a distribuição dos comprimentos dos períodos de chuva, porque estes períodos tem sempre um comprimento de pelo menos um dia. Esta situação é semelhante quando vamos modelar, por exemplo, a distribuição do comprimento de permanência hospitalar em dias, o número de artigos publicados por professores efetivos e assim por diante Böhning e Kuhnert (2006).

Para superar o fato de que o valor zero não pode ocorrer, definimos a distribuição binomial negativa truncada em zero (ZTNBD) Johnson, Kemp e Kotz (2005). Uma discussão detalhada do modelo ZTNB pode ser encontrada em (GROGGER; CARSON, 1991).

A partir de (4.1) temos $P\left(X=0 \mid \theta, \alpha\right)=\left(\frac{1}{1+\theta\alpha}\right)^{\frac{1}{\alpha}}$, então uma distribuição ZTNB tem FP:

$$P\left(X=x \mid X>0, \theta, \alpha\right)=\frac{\Gamma\left(x+\frac{1}{\alpha}\right)}{\Gamma\left(x+1\right)\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{\theta\alpha}{1+\theta\alpha}\right)^{x}\left(\frac{1}{1+\theta\alpha}\right)^{\frac{1}{\alpha}}\left[1-\left(\frac{1}{1+\theta\alpha}\right)^{\frac{1}{\alpha}}\right]^{-1} \tag{4.2}$$

for $x = 1, 2, \ldots$ e 0 caso contrário. A média e a variância são escritas, respectivamente, como: $E(X) = \theta\left[1 - \left(\frac{1}{1+\theta\alpha}\right)^{\frac{1}{\alpha}}\right]^{-1}$ e $V(X) = E(X)\left[(1+\theta+\theta\alpha) - E(X)\right]$ que são um pouco maiores em relação a distribuição binomial negativa. Para $\theta > \frac{\left(\frac{\alpha}{1+\alpha}\right)^{-\alpha}-1}{\alpha}$ a variável aleatória X com distribuição ZTNB tem $V(X) > E(X)$ e $V(X) < E(X)$ caso contrário.

Como uma alternativa para ZTNBD para modelar comprimentos de períodos consideramos o deslocamente da origem por um dia, que nos dá a distribuição binomial negativa deslocada (SNBD). A versão deslocada da origem é a maneira mais simples de transformar uma variável aleatória de contagem emu uma vairiável aleatóriade contagem estritamente positiva. A definição é a seguinte: Dada qualquer variável aleatória de contagem X pode-se definir a versão deslocada de X como $Y = X + 1$ e então $P(Y = y) = P(X = y - 1)$. A partir de (4.1) temos a distribuição SNBD com FP:

$$P(Y = y \mid \theta, \alpha) = \frac{\Gamma\left((y-1) + \frac{1}{\alpha}\right)}{\Gamma(y)\,\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{\theta\alpha}{1+\theta\alpha}\right)^{y-1}\left(\frac{1}{1+\theta\alpha}\right)^{\frac{1}{\alpha}} \text{ for } y = 1, 2, \ldots \tag{4.3}$$

e zero caso contrário. A média e a variância são, respectivamente, $E(Y) = 1 + \theta$ e $V(Y) = \theta(1 + \alpha\theta)$. Como a distribuição ZTNBD a distribuição SNBD pode ser usada para modelar dados de contagem sobredispersos ou subdispersos. Para $\theta > \frac{1}{\sqrt{\alpha}}$ temos $V(X) > E(X)$.

Uma outra modificação da NBD usada na análise do comprimentos da sequência de dias chuvosos e a distribuição Binomial negativa ponderada (SBNB) Mir (2009). Considerando como função peso a expressão $w(X) = X$, Patil e Rao (1978) com X distribuído segundo (4.1) temos a seguinte FP:

$$P(X = x \mid \theta, \alpha) = \frac{\Gamma\left(x + \frac{1}{\alpha}\right)}{\theta\Gamma(x)\,\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{\theta\alpha}{1+\theta\alpha}\right)^{x}\left(\frac{1}{1+\theta\alpha}\right)^{\frac{1}{\alpha}} \text{ for } x = 1, 2, \ldots \tag{4.4}$$

e zero caso contrário. A média de a variância são, respectivamente, $E(X) = \theta(1+\alpha) + 1$ e $V(X) = \theta(1+\alpha)(1+\alpha\theta)$ de modo que $V(X)$ é maior que $E(X)$ para $\theta > \frac{1}{\sqrt{\alpha}}$.

4.2.2 A distribuição Poisson-Lindley

A nova distribuição, denominada como distribuição Poisson-Lindley (PL), foi introduzida e estudada por Sankaran (1970). A distribuição PL pode ser utilizada como alternativa da distribuição Poisson ou da distribuição Binomial negativa em muitos problemas que envolvem a análise de dados de contagem. Como a distribuição NB, a distribuição PL é obtida

por um processo de composição como segue. Assumindo que $X \mid \lambda \sim Poisson\,(\lambda)$ tal que $\lambda \sim Lindley\,(\theta)$, então a distribuição marginal de X define a distribuição PL com FP escrita como:

$$P(X = x \mid \theta) = \frac{\theta^2\,(2+\theta+x)}{(1+\theta)^{3+x}}$$

para $x = 0, 1, \ldots$ e 0 caso contrário. Para todo θ, $\theta > 0$, a variável aleatória X é subdispersa, isso é $E\,(X) < V\,(X)$ onde $E\,(X) = \frac{2+\theta}{\theta\,(1+\theta)}$ e $V(X) = \frac{2+6\,\theta+4\,\theta^2+\theta^3}{\theta^2(1+\theta)^2}$.

Uma recente extensão da distribuição Poisson-Lindley chamada de Poisson-Lindley generalizada foi apresentada em (MAHMOUDI; ZAKERZADEH, 2010) o qual considera a composição entre as distribuições Poisson e Lindley generalizada Zakerzadeh e Dolati (2009). Esta extensão não é considerada neste artigo.

Considerada por (GHITANY; AL-MUTAIRI; NADARAJAH, 2008), a distribuição Poisson-Lindley truncada em zero (ZTPL) tem função de probabilidade:

$$P(X = x \mid \theta) = \frac{\theta^2\,(2+\theta+x)}{(1+\theta)^x\,(1+3\,\theta+\theta^2)} \text{ para } x = 1, 2, \ldots \text{ e } \theta > 0. \tag{4.5}$$

A média e variância se X são dadas, respectivamente, por $E\,(X) = \frac{(2+\theta)(1+\theta)^2}{\theta(1+3\,\theta+\theta^2)}$ e $V(X) = \frac{(1+\theta)^2\left(2+10\,\theta+6\,\theta^2+\theta^3\right)}{\theta^2(1+3\,\theta+\theta^2)^2}$. Para $\theta \cong 1.259$ temos equidispersão, sobredispersão para $\theta < 1.259$ e subdispersão para $\theta > 1.259$.

Duas outras modificações da distribuição Poisson-Lindley, úties quando não se observa zeros, são a distribuição Poisson-Lindley deslocada (SPL) e a distribuição Poisson-Lindley ponderada.

Uma variável aleatória com distribuição SPL tem função de probabilidade:

$$P(X = x \mid \theta) = \frac{\theta^2\,(1+\theta+x)}{(1+\theta)^{2+x}} \text{ para } x = 1, 2, \ldots \text{ e } \theta > 0,$$

que é equidispersa para $\theta \cong 1.297$, sobredispersa para $\theta < 1.297$ e subdispersa para $\theta > 1.297$. A média e variância de X são dadas, respectivamente, por: $E\,(X) = \frac{2+2\,\theta+\theta^2}{\theta(1+\theta)}$ e $V(X) = \frac{2+6\,\theta+4\,\theta^2+\theta^3}{\theta^2(1+\theta)^2}$.

A distribuição de Poisson-Lindley ponderada com $w(X) = X$ denominada "size-biased", Rao (1965), e tem FP dada por:

$$P(X = x \mid \theta) = \frac{\theta^3 x\,(2+\theta+x)}{(2+\theta)\,(1+\theta)^{2+x}} \text{ para } x = 1, 2, \ldots \text{ e } \theta > 0,$$

e foi introduzida por (GHITANY; AL-MUTAIRI, 2008). A média e variância se X são dadas, respectivamente, por $E(X) = \frac{6+4\theta+\theta^2}{\theta(2+\theta)}$ e $V(X) = \frac{2\left(6+12\theta+6\theta^2+\theta^3\right)}{\theta^2(2+\theta)^2}$. Para $\theta \cong 1.671$ temos equidispersão, sobredispersão para $\theta < 1.671$ e subdispersão para $\theta > 1.671$.

4.2.3 Outras Distribuições

Nesta seção nós apresentamos, brevemente, três outras distribuições frequentemente utilizadas para descrever o comprimento de períodos secos ou chuvosos. São elas: a distribuição geométrica (G), a distribuição Poisson truncada em zero (ZTPD) e a distribuição de séries logarítmica (LD). Todas elas possuem suporte definido em $\mathbb{N}^+$.

Uma variável aleatória discreta X tem distribuição Geométrica, também chamada de distribuição Geométrica truncada em zero, se a FP for escrita como $P(X = x \mid \theta) = \theta(1-\theta)^{x-1}$, onde $0 < \theta < 1$. A média e a variância são, respectivamente, $E(X) = \frac{1}{\theta}$ e $V(X) = \frac{1-\theta}{\theta^2}$. Para $0.5 < \theta < 1$ temos $E(X) > V(X)$ e $E(X) < V(X)$ para $0 < \theta < 1$.

A distribuição Poisson truncada em zero tem FP: $P(X = x \mid \theta) = \frac{\theta^x \exp(-\theta)}{x![1-\exp(-\theta)]}$ tal que a $E(X) > V(X)$ para todo θ, $\theta > 0$, e as mesmas são escritas, respectivamente, como $E(X) = \frac{\theta}{[1-\exp(-\theta)]}$ e $V(X) = \frac{\theta}{[1-\exp(-\theta)]} - \frac{\theta^2 \exp(-\theta)}{[1-\exp(-\theta)]^2}$.

A distribuição de séries logarítmica tem $E(X) > V(X)$ para $0.6321 < \theta < 1$ e $E(X) < V(X)$ for $0 < \theta < 0.6321$ em que $E(X) = -\frac{\theta}{\log(1-\theta)(1-\theta)}$ e $V(X) = -\frac{\theta}{\log(1-\theta)(1-\theta)^2}\left[1 + \frac{1}{\log(1-\theta)}\right]$. Sua FP é dada por $P(X = x \mid \theta) = -\frac{\theta^x}{\log(1-\theta)x}$.

4.3 Os Dados

Foram utilizadas as séries formadas pelas precipitações diárias observadas em 33 estações meteorológicas de superfície localizadas nos estados do Rio grande do Sul (18 estações), Paraná (8 estações) e Santa Catarina (7 estações). As séries de preciptação diárias foram obtidas do "Banco de dados meteriológicos para educação e pesquisa" do Instituto Nacional de Meteriologia (<http://www.inmet.gov.br>) de 1 de janeiro até 31 de dezembro de 2014.

Os comprimentos das sequências foram contados dentro de cada mês e adotamos como chuvosos os dias com precipitação superior a 5.0 mm. Os meses que mostram qualquer valor diário ausente foram excluídos da série. Uma vez que os comprimentos das sequências foram contados dentro de cada mês, meses em que o número de dias de medição não

coincidiram com o número de dias do calendário foram descartados.

Apresentamos na Tabela 12, para cada uma das estações, algumas características numéricas dos comprimentos das sequências dos dias chuvosos. A saber, a cidade onde se localiza a estação (Cidade), o código da estação segundo a Organização Mundial de Meteorologia (OMM), o período usado na consolidação da série (Data Inicial e Data Final), o número de dias no período (n_1), o número efetivo de dias que a estação operou no período (n_2), a porcentagem de dias que a estação não operou no período (Perdidos), o número de dias com preciptação ≥ 5.0 mm (Precipitação) e a média e o desvio padrão dos períodos chuvosos (Média e DP).

Tabela 12 – Características numéricas das estações

Cidade	OMM	Data Inicial	Data Final	n_1	n_2	Perdidos	Precipitação	Média	DP
Londrina	83766	1995-01-01	2014-12-31	7305	6446	11.76%	1217	1.59	1.01
Maringa	83767	1995-01-01	2014-11-30	7274	6233	14.31%	1207	1.55	0.97
Campo Mourao	83783	1995-01-01	2014-12-31	7305	6666	8.75%	1321	1.52	0.90
Ivai	83811	1996-01-01	2014-12-31	6940	6175	11.02%	1263	1.54	0.86
Castro	83813	1996-01-01	2014-12-31	6940	5928	14.58%	1172	1.52	0.86
Irati	83836	1995-01-01	2014-12-31	7305	7057	3.39%	1480	1.56	0.92
Curitiba	83842	1995-01-01	2014-12-31	7305	6904	5.49%	1406	1.53	0.90
Paranagua	83844	1995-01-01	2014-12-31	7305	6416	12.17%	1724	1.76	1.17
Indaial	83872	1995-01-01	2014-12-31	7305	6298	13.79%	1409	1.56	0.93
Irai	83881	1995-01-01	2014-12-31	7305	6418	12.14%	1362	1.54	0.92
Chapeco	83883	1995-01-01	2014-12-31	7305	6815	6.71%	1468	1.52	0.87
Campos Novos	83887	1995-01-01	2014-12-31	7305	6420	12.11%	1520	1.58	0.95
Lages	83891	1995-01-01	2014-12-31	7305	7057	3.39%	1546	1.51	0.88
Florianopolis	83897	1995-01-01	2014-12-31	7305	5872	19.62%	1311	1.52	0.89
Sao Luiz Gonzaga	83907	1995-01-01	2014-12-31	7305	6663	8.79%	1280	1.52	0.85
Cruz Alta	83912	1995-01-01	2014-12-31	7305	6757	7.50%	1382	1.51	0.89
Passo Fundo	83914	1995-01-01	2014-12-31	7305	6329	13.36%	1364	1.51	0.86
Lagoa Vermelha	83916	2000-01-01	2014-12-31	5479	4319	21.17%	930	1.56	0.92
Bom Jesus	83919	1995-01-01	2014-12-31	7305	6634	9.19%	1500	1.54	0.87
Sao Joaquim	83920	1995-01-01	2014-12-31	7305	6057	17.08%	1489	1.59	0.94
Urussanga	83923	2001-09-01	2014-08-31	4748	4687	1.28%	980	1.61	0.98
Uruguaiana	83927	1995-01-01	2014-12-31	7305	6414	12.20%	983	1.44	0.76
Santa Maria	83936	1995-01-01	2014-12-31	7305	7026	3.82%	1333	1.43	0.78
Bento Goncalves	83941	1995-01-01	2013-12-31	6940	5536	20.23%	1144	1.52	0.83
Caxias Do Sul	83942	1995-01-01	2014-12-31	7305	6698	8.31%	1460	1.51	0.88
Torres	83948	1995-01-01	2014-12-31	7305	6082	16.74%	1223	1.53	0.90
Santana Do Livramento	83953	1998-06-01	2013-11-30	5662	4350	23.17%	685	1.42	0.70
Encruzilhada Do Sul	83964	1995-01-01	2014-12-31	7305	7026	3.82%	1344	1.46	0.76
Porto Alegre	83967	1995-01-01	2014-12-31	7305	6964	4.67%	1288	1.47	0.77
Bage	83980	1995-01-01	2014-12-31	7305	6660	8.83%	1152	1.45	0.77
Pelotas	83985	1996-01-01	2014-08-31	6818	6113	10.34%	995	1.40	0.72
Rio Grande	83995	1995-01-01	2014-12-31	7305	6965	4.65%	1099	1.37	0.69
Santa Vitoria Do Palmar	83997	1995-01-01	2014-12-31	7305	5996	17.92%	861	1.35	0.68

n_1: número de dias no período.

n_2: número de dias com medições no período. ,

Precipitação: número de dias com precipitação ≥ 5.0 mm.

4.4 Resultados e Discussões

Segundo o teste de Bartels, em 3 das 33 séries a hipótese de comprimentos *i.i.d* foi rejeitada em nível de significância de 1%. Estas séries são provenientes das estações PR-83783, RS-83912 e SC-83887. As séries provenientes das estações PR-83783 e RS-83912 também não passaram no teste de Wald-Wolfwitz adotando-se o mesmo nível de significância. Todas as séries passaram no teste de Mann-Kendall. Os p-valores associados aos valores das estatísticas, de ambos os testes, foram calculados a partir de $B = 10.0000$ permutações Good (2005) das séries originais.

As bibliotecas *lawstat* Gastwirth et al. (2015), *randtests* Caeiro e Mateus (2014) e McLeod (2011), do sistema R R Core Team (2015), foram utilizadas, respectivamente, na aplicacão dos testes. Wald-Wolfwitz, respectivamente.

Apresentamos na Tabela 13 os p-valores associados ao teste qui-quadrado usado para avaliar o ajustes das nove distribuições. Um resumo do número de séries ajustadas por cada distribuição, considerando níveis de significância de 1% e 5%, são apresentado na Tabela 14.

Tabela 13 – P-valores do teste de qualidade de ajuste qui-quadrado.

WMO	G	ZTP	LG	ZTNB	SBN	SBNB	ZTPL	SPL	SBPL
83766	0.18	0.00	0.34	0.55	0.58	0.01	0.08	0.05	0.00
83767	0.16	0.00	0.06	0.26	0.21	0.01	0.07	0.05	0.00
83783	0.06	0.00	0.00	0.03	0.03	0.01	0.04	0.03	0.00
83811	0.10	0.00	0.00	0.37	0.49	0.49	0.17	0.20	0.36
83813	0.09	0.05	0.00	0.40	0.45	0.45	0.17	0.20	0.64
83836	0.31	0.00	0.00	0.55	0.66	0.66	0.46	0.52	0.19
83842	0.33	0.00	0.00	0.20	0.20	0.12	0.31	0.30	0.00
83844	0.40	0.00	0.00	0.31	0.33	0.10	0.25	0.19	0.00
83881	0.18	0.00	0.00	0.10	0.10	0.04	0.14	0.13	0.00
83907	0.03	0.00	0.00	0.13	0.20	0.20	0.06	0.07	0.16
83912	0.06	0.00	0.00	0.03	0.03	0.01	0.04	0.04	0.00
83914	0.60	0.00	0.00	0.70	0.76	0.76	0.72	0.76	0.14
83916	0.15	0.00	0.00	0.15	0.19	0.19	0.20	0.22	0.16
83919	0.20	0.00	0.00	0.32	0.38	0.38	0.30	0.33	0.08
83927	0.57	0.01	0.01	0.50	0.46	0.45	0.65	0.67	0.22
83936	0.30	0.00	0.07	0.23	0.26	0.03	0.25	0.23	0.00
83941	0.07	0.00	0.00	0.16	0.19	0.19	0.11	0.13	0.19
83942	0.86	0.00	0.00	0.74	0.74	0.54	0.85	0.84	0.01
83948	0.75	0.00	0.02	0.59	0.59	0.33	0.70	0.67	0.00
83953	0.09	0.32	0.00	0.36	0.43	0.43	0.12	0.14	0.64
83964	0.04	0.09	0.00	0.49	0.56	0.56	0.08	0.09	0.73
83967	0.11	0.11	0.00	0.84	0.87	0.87	0.20	0.23	0.90
83980	0.37	0.00	0.00	0.49	0.55	0.55	0.47	0.51	0.32
83985	0.24	0.00	0.00	0.16	0.18	0.18	0.27	0.27	0.08
83995	0.65	0.01	0.02	0.49	0.48	0.41	0.70	0.71	0.19
83997	0.89	0.02	0.05	0.86	0.88	0.79	0.93	0.93	0.42
83872	0.47	0.00	0.00	0.62	0.69	0.70	0.64	0.68	0.17
83883	0.53	0.00	0.00	0.58	0.60	0.60	0.65	0.68	0.05
83887	0.63	0.00	0.00	0.46	0.47	0.36	0.61	0.58	0.00
83891	0.69	0.00	0.00	0.56	0.58	0.48	0.71	0.70	0.00
83897	0.62	0.00	0.00	0.46	0.47	0.37	0.61	0.60	0.01
83920	0.63	0.00	0.00	0.80	0.81	0.81	0.80	0.84	0.07
83923	0.22	0.00	0.00	0.13	0.13	0.12	0.21	0.21	0.00

Tabela 14 – Número de séries ajustadas pelas distribuições de acordo com o teste qui-quadradado.

α	G	ZTP	LG	ZTNB	SBN	SBNB	ZTPL	SPL	SBPL
1%	33	7	6	33	33	30	33	33	19
5%	31	3	3	31	31	27	31	31	19

Pelo critério de informação de Akaike, encontramos que as distribuições G, SBN, ZTPL, SPL e SBPL foram selecionadas, respectivamente, 9, 2, 3, 12 e 7 vezes. Pelo critério de informação de Schwarz, estas distribuições foram selecionadas, respectivamente, 10, 0, 3, 13 e 7 vezes.

4.5 Conclusões

Na literatura, vários modelos discretos de probabilidade tem sido utilizados para caracterizar o comprimento de períodos chuvosos. Este artigo identificou alguns conjuntos novos distribuições discretas univariadas como possíveis alternativas para as distribuições G, ZTP, LG, ZTNB, SBN e SBNB. Nossos resultados mostraram que as distribuições ZTPL e SPL são promissoras para a modelagem dos comprimentos de perídos chuvosos.

Como apontado em (DENI; JEMAIN; IBRAHIM, 2008), a análise da distribuição dos períodos molhados e secos baeados na precipitação diária é cada vez mais importante e tais análises podem ser utilizadas para a finalidade de tomada de decisão ou previsões. Neste sentido é importante avaliar o desempenho de várias distribuições de probabilidade para descrever o comportamento do comprimento de dias chuvosos. É importante salientar que nenhum ajuste de qualquer distribuição em particular para as séries de períodos molhados encontrado é sempre adequado.

Referências

ADEYEMI, S.; OJO, M. O. On a generalization of gumbel distribution. 2003.

AKAIKE, H. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, v. 19, n. 6, p. 716–723, 1974.

AL-KHASAWNEH, M. F. Estimating the negative binomial dispersion parameter. *Asian Journal of Mathematics & Statistics*, v. 3, n. 1, p. 1–15, 2010.

AQUINO, L. et al. Precipitação quinzenal provável na bacia hidrográfica do arroio pelotas (pelotas rs). 2012.

ARAGÓN, J.; EBERLY, D.; EBERLEY, S. Existence and uniqueness of the maximum likelihood estimator for the two-parameter negative binomial distribution. *Statistics & Probability Letters*, v. 15, n. 5, p. 375–379, 1992.

ARYAL, G. R.; TSOKOS, C. P. On the transmuted extreme value distribution with application. *Nonlinear Analysis: Theory, Methods & Applications*, v. 71, n. 12, p. e1401 – e1407, 2009.

BARBOSA, E. et al. Distribuiç ão generalizada de valores extremos (gve): Um estudo aplicado a valores de temperatura mínima da cidade de viçosa-mg. *Revista da Estatística da Universidade Federal de Ouro Preto*, v. 3, n. 3, p. 387–391, 2014.

BARTELS, R. The rank version of von Neumann's ratio test for randomness. *Journal of the American Statistical Association*, v. 77, n. 377, p. 40–46, 1982.

BISPO, E. M. Caracterização agroclimática e probabilidade de precipitação para região de ilha solteira - sp. *Faculdade de Engenharia*, 2008.

BöHNING, D.; KUHNERT, R. Equivalence of truncated count mixture distributions and mixtures of truncated count distributions. *Biometrics*, v. 62, n. 4, p. 1207–1215, 2006.

BOZDOGAN, H. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, Springer, v. 52, n. 3, p. 345–370, 1987.

BRITO, E. *Algumas novas distribuições: desenvolvimento e aplicações*. Tese (Doutorado) — Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2014.

CAEIRO, F.; MATEUS, A. *randtests: Testing randomness in R*. [S.l.], 2014. R package version 1.0. Disponível em: <http://CRAN.R-project.org/package=randtests>.

CAMERON, A. C.; TRIVEDI, P. K. *Regression Analysis of Count Data*. [S.l.]: Cambridge University Press, 1998. (Econometric Society Monographs, v. 30).

CINDRIC, K.; PASARIC, Z.; GAJIC-CAPKA, M. Spatial and temporal analysis of dry spells in Croatia. *Theoretical and Applied Climatology*, v. 102, n. 1-2, p. 171–184, 2010.

COLES, S. et al. *An introduction to statistical modeling of extreme values*. [S.l.]: Springer, 2001.

CORDEIRO, G. M.; NADARAJAH, S.; ORTEGA, E. M. M. The kumaraswamy gumbel distribution. *Statistical Methods & Applications*, Springer-Verlag, v. 21, n. 2, p. 139–168, 2012.

CORDEIRO, G. M.; ORTEGA, E. M. M.; CUNHA, D. C. C. da. The exponentiated generalized class of distributions. *Journal of Data Science*, Academic Journal, v. 11, p. 1–27, 2013.

DASH, S. K. et al. Characteristic changes in the long and short spells of different rain intensities in India. *Theoretical and Applied Climatology*, v. 105, n. 3-4, p. 563–570, 2011.

DAVISON, A. et al. A sketch of statistics of spatial extremes. 2009.

DENI, S. M.; JEMAIN, A. A. Mixed geometric truncated Poisson model for sequences of wet days. *Journal of Applied Sciences*, v. 8, n. 21, p. 3975–3980, 2008.

DENI, S. M.; JEMAIN, A. A. Fitting the distribution of dry and wet spells with alternative probability models. *Meteorology and Atmospheric Physics*, v. 104, n. 1-2, p. 13–27, 2009. ISSN 0177-7971.

DENI, S. M.; JEMAIN, A. A. Mixed log series geometric distribution for sequences of dry days. *Atmospheric Research*, v. 92, n. 2, p. 236 – 243, 2009.

DENI, S. M.; JEMAIN, A. A. Comparison between mixed probability models and Markov chain models for weekly dry and wet spells in Peninsular Malaysia. *Proceedings of the World Congress on Engineering*, v. 1, 2012.

DENI, S. M.; JEMAIN, A. A.; IBRAHIM, K. The spatial distribution of wet and dry spells over Peninsular Malaysia. *Theoretical and Applied Climatology*, v. 94, n. 3-4, p. 163–173, 2008.

DENI, S. M.; JEMAIN, A. A.; IBRAHIM, K. Mixed probability models for dry and wet spells. *Statistical Methodology*, v. 6, n. 3, p. 290 – 303, 2009.

DENI, S. M.; JEMAIN, A. A.; IBRAHIM, K. The best probability models for dry and wet spells in Peninsular Malaysia during monsoon seasons. *International Journal of Climatology*, v. 30, n. 8, p. 1194–1205, 2010. ISSN 1097-0088.

DOBI-WANTUCH, I.; MIKA, J.; SZEIDL, L. Modelling wet and dry spells with mixture distributions. *Meteorology and Atmospheric Physics*, v. 73, n. 3-4, p. 245–256, 2000. ISSN 0177-7971.

DUBEY, S. D. A new derivation of the logistic distribution. *Naval Research Logistics Quarterly*, Wiley Subscription Services, Inc., A Wiley Company, v. 16, n. 1, p. 37–40, 1969.

EVERITT, B. S.; HAND, D. J. *Finite mixture distributions*. [S.l.]: Chapman & Hall, London-New York, 1981. Monographs on Applied Probability and Statistics.

FISHER, R. A.; TIPPETT, L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, v. 24, p. 180–190, 4 1928.

GASTWIRTH, J. L. et al. *lawstat: Tools for Biostatistics, Public Policy, and Law*. [S.l.], 2015. R package version 2.5. Disponível em: <http://CRAN.R-project.org/package=lawstat>.

GHITANY, M. E.; AL-MUTAIRI, D. K. Size-biased Poisson-Lindley distribution and its application. *Metron - International Journal of Statistics*, LXVI, n. 3, p. 299–311, 2008.

GHITANY, M. E.; AL-MUTAIRI, D. K.; NADARAJAH, S. Zero-truncated Poisson-Lindley distribution and its application. *Mathematics and Computers in Simulation*, v. 79, n. 3, p. 279–287, 2008.

GOOD, P. *Permutation, parametric and bootstrap tests of hypotheses*. Third. [S.l.]: Springer-Verlag, New York, 2005. xx+315 p. (Springer Series in Statistics).

GROGGER, J. T.; CARSON, R. T. Models for truncated counts. *Journal of Applied Econometrics*, v. 6, n. 3, p. 225–238, 1991.

GUMBEL, E. Les valeurs extrêmes des distributions statistiques. *Annales de l'institut Henri Poincaré*, Presses universitaires de France, v. 5, n. 2, p. 115–158, 1935.

HARTMANN, M.; MOALA, F. A.; MENDONCA, M. A. Estudo das precipitações máximas anuais em presidente prudente. *Revista Brasileira de Meteorologia*, scielo, v. 26, p. 561 – 568, 12 2011.

HEFFERNAN, J.; STEPHENSON, A. ismev: An introduction to statistical modeling of extreme values. *UR L http://CRAN. R-project. org/package= ismev. R package version*, v. 1, 2012.

HILBE, J. M. *Negative binomial regression*. 2th. ed. [S.l.]: Cambridge University Press, Cambridge, 2011.

JAIN, G. C.; CONSUL, P. C. A generalized negative binomial distribution. *SIAM Journal on Applied Mathematics*, v. 21, n. 4, p. 501–513, 1971.

JENKINSON, A. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, v. 81, p. 158–171, apr 1955.

JOHNSON, N. L.; KEMP, A. W.; KOTZ, S. *Univariate discrete distributions*. Third. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons], 2005. xx+646 p. (Wiley Series in Probability and Statistics).

KENDALL, M.; GIBBONS, J. D. *Rank correlation methods*. Fifth. [S.l.]: Edward Arnold, London, 1990. (A Charles Griffin Title).

LANA, X. et al. Statistical distributions and sampling strategies for the analysis of extreme dry spells in Catalonia (NE Spain). *Journal of Hydrology*, v. 324, n. 1-4, p. 94–114, 2006.

LISKA, G. R. et al. Estimativas de velocidade máxima de vento em piracicaba-sp via séries temporais e teoria de valores extremos. *Revista Brasileira de Biometria, São Paulo*, v. 31, n. 2, p. 295–309, 2013.

MAHMOUDI, E.; ZAKERZADEH, H. Generalized Poisson-Lindley distribution. *Communications in Statistics - Theory and Methods*, v. 39, p. 1785–1798, 2010.

MANN, H. B. Nonparametric tests against trend. *Econometrica*, v. 13, n. 3, p. 245–259, 1945.

MARSHALL, A. W.; OLKIN, I. A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, Biometrika Trust, v. 84, n. 3, p. 641–652, 1997.

MARTINS, F. J. P. *Dimensionamento Hidrológico e Hidrálico de Passagens Inferiores Rodoviárias para águas Pluviais*. Dissertação (Mestrado) — Universidade de Coimbra - Faculdade de Ciências e de Tecnologia, 2000.

MATHUGAMA, S. C.; PEIRIS, T. S. G. Critical evaluation of dry spell research. *International Journal of Basic & Applied Sciences*, v. 11, n. 6, 2011.

MCCALLA, R. J.; DAY, E. E. D.; MILLWARD, H. A. The relative concept of warm and cold spells of temperature: Methodology and application. *Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie B*, v. 25, n. 4, p. 323–336, 1978.

MCLEOD, A. I. *Kendall: Kendall rank correlation and Mann-Kendall trend test*. [S.l.], 2011. R package version 2.2.

MIR, K. A. On size-biased negative binomial distribution and its use in zero-truncated cases. *Measurement Science Review*, v. 9, n. 2, 2009.

NADARAJAH, S. The exponentiated gumbel distribution with climate application. *Environmetrics*, John Wiley & Sons, Ltd., v. 17, n. 1, p. 13–23, 2006. ISSN 1099-095X.

NADARAJAH, S.; KOTZ, S. The beta gumbel distribution. *Mathematical Problems in Engineering*, Hindawi Publishing Corporation, New York, n. 4, p. 323–332, 2004.

NAGHETTINI, M.; PINTO, E. J. A. *Hidrologia Estatística*. [S.l.]: CPRM, 2007.

OJO, M. O. Some relationships between the generalized gumbel and other distributions. *Kragujevac Journal of Mathematics*, Prirodno-matematiki fakultet Kragujevac, v. 23, n. 23, p. 101–106, 2001.

PASCOA, M. A. R. *Extensões da distribuição Gama Generalizada: propriedades e aplicações*. Tese (Doutorado) — Universidade de São Paulo - Escola Superior de Agricultura "Luiz de Queiroz", 2012.

PATIL, G. P.; RAO, C. R. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, v. 34, n. 2, 1978.

PINHEIRO, E. C. *Contribuições em Inferênciae Modelagem de Valores Extremos*. Tese (Doutorado) — Universidade de São Paulo, 2014.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2015.

RAO, C. R. On discrete distributions arising out of methods of ascertainment. *Sankhyā (Statistics). The Indian Journal of Statistics. Series A*, v. 27, p. 311–324, 1965.

SAHA, K.; PAUL, S. Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. *Biometrics*, v. 61, n. 1, p. 179–185, 2005.

SANKARAN, M. The discrete Poisson-Lindley distribution. *Biometrics*, v. 26, p. 145–149, 1970.

SANTOS, W. d. O. et al. Precipitações máximas para o município de mossoró de 1964 a 2011 pela distribuição de gumbel. *IRRIGA - Brazilian Journal of Irrigation and Drainage*, v. 19, p. 207 – 213, 2014.

SCHWARZ, G. Estimating the dimension of a model. *The annals of statistics*, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978.

SHAW, E. M. et al. *Hydrology in practice*. 4th. ed. [S.l.]: Spon Press (Taylor and Francis), 2010.

SHAW, W. T.; BUCKLEY, I. R. The alchemy of probability distributions: Beyond gram-charlier & cornish-fisher expansions, and skew-normal or kurtotic-normal distributions. *Department of Mathematics - King's College London*, 2007.

SILVA, J. C. d. et al. Análise de distribuição de chuva para santa maria, rs. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 11, p. 67 – 72, 02 2007.

SONNADARA, D. U. J.; JAYEWARDENE, D. R. A Markov chain probability model to describe wet and dry patterns of weather at Colombo. *Theoretical and Applied Climatology*, v. 119, n. 1-2, p. 333–340, 2015.

TITTERINGTON, D. M.; SMITH, A. F. M.; MAKOV, U. E. *Statistical analysis of finite mixture distributions*. [S.l.]: John Wiley & Sons, Ltd., Chichester, 1985. (Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics).

TOLIKA, K.; MAHERAS, P. Spatial and temporal characteristics of wet spells in Greece. *Theoretical and Applied Climatology*, v. 81, n. 1-2, p. 71–85, 2005.

VIEIRA, J. a. P. G. et al. Estudo da precipitações mensal durante a estação chuvosa em diamantina, minas gerais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 14, p. 762 – 767, 07 2010.

WALD, A.; WOLFOWITZ, J. An exact test for randomness in the non-parametric case based on serial correlation. *The Annals of Mathematical Statistics*, JSTOR, v. 14, n. 4, p. 378–388, 1943.

WANG, Y. Estimation problems for the two-parameter negative binomial distribution. *Statistics & Probability Letters*, v. 26, n. 2, p. 113–114, 1996.

WATANABE, F. M. *Análise do Método de Gumbel para cálculo de vazões de dimensionamento de vertedouros*. Dissertação (Mestrado) — Universidade de São Paulo - USP / Escola de Engenharia São Carlos - Engenharia Elétrica com Ênfase em Sistemas de Energia e Automação, 2013.

WMO. *Guide to Hydrological Practices*. [S.l.]: Management of Water Resources and Application of Hydrological Practices, 2009.

ZAKERZADEH, H.; DOLATI, A. Generalized Lindley distribution. *Journal of Mathematical Extension*, v. 3, n. 4, p. 13–25, 2009.

ZEKÂI, S. *Chapter Three - Temporal Drought Analysis and Modeling*. Boston: Elsevier, 2015.