Abordagem Bayesiana não-paramétrica para análise de valores extremos

por

Fernando Ferraz do Nascimento

Universidade Federal do Rio de Janeiro

Instituto de Matemática

Departamento de Métodos Estatísticos

2009

Abordagem Bayesiana não-paramétrica para análise de valores extremos

Fernando Ferraz do Nascimento

Tese de Doutorado submetida ao programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários à obtenção do grau de Doutor em Estatística.

Orientadores: **Dani Gamerman e Hedibert Freitas Lopes**

Rio de Janeiro, Dezembro de 2009

Abordagem Bayesiana não-paramétrica para análise de valores extremos

Fernando Ferraz do Nascimento
Orientadores: Dani Gamerman e Hedibert Freitas Lopes

Tese de Doutorado submetida ao programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários à obtenção do grau de Doutor em Estatística.

Aprovada por:

Presidente Prof. Dani Gamerman
IM–UFRJ

Prof. Hedibert Freitas Lopes
University of Chicago

Prof. Helio dos Santos Migon
IM–UFRJ

Prof[a]. Nancy Lopes Garcia
IMECC–UNICAMP

Prof. José Miguel Bernardo
Universitat de València

Rio de Janeiro, Dezembro de 2009

Nascimento, Fernando Ferraz

Abordagem Bayesiana não-paramétrica para análise de valores extremos/ Fernando Ferraz do Nascimento. – Rio de Janeiro: UFRJ/IM, 2009.

xvii, 133 f. : il. ; 31cm.

Tese (Doutorado) – UFRJ/IM. Programa de Pós-Graduação em Estatística, 2009.

Orientadores: Dani Gamerman, Hedibert F. Lopes

Referências bibliográficas: p.167–170.

1. Estatística Matemática - Tese. I. Gamerman, Dani. II. Universidade Federal do Rio de Janeiro. Instituto de Matemática. III. Título.

AGRADECIMENTOS

Gostaria de agradecer meus orientadores Dani e Hedibert, que foram fundamentais para o meu crescimento como pesquisador e educador. Com certeza meu trabalho não seria melhor se eu tivesse escolhido fazer o doutorado em outro lugar. Espero continuar fazendo pesquisa com eles e com outros professores e alunos com quem tive contato nestes quatro anos.

Gostaria de agradecer também as pessoas que conheci nestes quatro anos. Em primeiro e mais importante minha esposa Valmária, que me motivou e sempre esteve presente em todos os momentos desta caminhada e colocou Jesus na minha vida, sou grato a Ele por tudo o que conquistei e irei conquistar. Também aos professores da UFRJ, aos meus amigos, entre os que conheci na UFRJ: Luiz Ledo, Cleide, Fidel, Vinícius Israel, Luzia, Alexandre, Geraldo Marcelo, Adelmo, Esther, Flávio e Vinícius Mayrink. Aos colegas que já conhecia como o Angelo, e que conheci, como a Laninha. Meus colegas de república Waguinho, Fabio e Josi Baldo.

À CAPES, CNPQ e FAPERJ pelo auxílio durante este período.

À minha família que me ajudou na minha formação como pessoa e sempre serei grato: Meus pais Neide e Mauri, meu irmão Rafael, aos meus tios Marlene e Odilon, meus avós paternos Maria e Moacir. Meus avós maternos José e Alcedira. Meus sogros Maria e Valmir, meus cunhados, sobrinhos e amigos que ganhei lá no Piauí.

Resumo

Este trabalho consiste na estimação de densidade de valores extremos. A distribuição de Pareto generalizada (GPD) acima de um limiar é combinada com uma distribuição para valores abaixo do limiar utilizando uma estimação não-paramétrica. Esta configuração semi-paramétrica generaliza algumas abordagens já existentes e fornece uma estimação de densidade sobre todo o espaço amostral. A estimação é feita utilizando o paradigma Bayesiano, que ajuda a identificar as componentes do modelo. A estimação de todas as componentes do modelo, incluindo o limiar, os quantis extremos, limites superiores das observações, quando apropriadas, e predição para observações futuras são fornecidas. Inferência a Posteriori é feita utilizando métodos de Monte Carlo via cadeias de Markov (MCMC). Estudos de simulação sugerem alguns métodos para avaliar a relevância dos procedimentos propostos. Os modelos são então aplicados em duas aplicações de dados ambientais, em vazão de rios em Porto Rico e níveis de chuva em Portugal.

Uma outra parte deste trabalho consiste em estudar dados extremos na presença de informações auxiliares relevantes. A principal novidade é a introdução de uma estrutura de regressão para explicar a variação dos excessos sobre todos os parâmetros da cauda. Os modelos são aplicados no estudo de dois conjuntos de dados de temperatura: máximos nos Estados Unidos e mínimo no estado do Rio de Janeiro, e comparados com outros modelos.

Na última parte deste trabalho são analisados extremos, considerando que os parâmetros da distribuição GPD variam no tempo. Foi introduzido um Modelo Linear Dinâmico (DLM), uma classe geral de modelos de séries temporais, para modelar os parâmetros de forma e escala ao longo do tempo. O modelo proposto foi aplicado em três dados financeiros de séries temporais: Índices BOVESPA, Petrobras e Vale do Rio Doce.

Palavras-Chave: Inferência Bayesiana, teoria de valores extremos, estimação não-paramétrica por mistura, métodos MCMC, modelos dinâmicos, modelos de regressão.

Abstract

This work is concerned with extreme value density estimation. The generalized Pareto distribution (GPD) beyond a given threshold is combined with a nonparametric estimation approach above the threshold. This semiparametric setup is shown to generalize a few existing approaches and enables density estimation over the complete sample space. Estimation is performed via the Bayesian paradigm, which helps identify model components. Estimation of all model parameters, including the threshold, higher quantiles, upper bounds for the data, where appropriate, and prediction for future observations are provided. Posterior inference is performed through Markov chain Monte Carlo (MCMC) methods. Simulation studies suggests a few useful guidelines to evaluate the relevance of the proposed procedures. Models are then applied to environmental data sets, in river flow in Porto Rico and rainfall levels in Portugal.

Another part in this work is concerned with the study of extreme data in the presence of relevant auxiliary information. The main novelty is the introduction of a regression structure to explain the variation of the exceedances through all tail parameters. The models are applied to the study of two temperature datasets: maxima in the U.S.A. and minima in Brazil, and compared to other related models.

In the last part of this work was analyzed the extremal, allowing the parameters of GPD to vary with time. Was introduced the use of dynamic linear model (DLM), a very general class of time series models, to model the shape and scale parameter changes across time. The proposed model was applied to three real financial time series: the Brazilian BOVESPA, Petrobras and Vale do Rio Doce index.

Key-Words: Bayesian Inference, extreme value theory, nonparametric estimation by mixtures, MCMC methods, dynamic models, regression models.

Sumário

Lista de Tabelas

Lista de Figuras

Capítulo 1

Introdução

Análise de dados extremos tem se tornado uma grande ferramenta nas mais diversas áreas do conhecimento, ajudando a prever grandes ganhos e perdas. Duas áreas onde esta análise se destaca são nas áreas ambientais e econômicas.

Problemas com enchentes, devido a altos índices de chuva, sempre preocuparam as autoridades devido a grande perda material e financeira que estes eventos acarretam para a sociedade. Ter o conhecimento da frequência com qual estes eventos ocorrem é de grande importância para as autoridades, e principalmente à população, para que possa ser feito um trabalho preventivo, para evitar ou diminuir as perdas. Este raciocínio também se aplica a outros fenômenos naturais, como em vazão de rios e velocidade do vento.

Em uma aplicação financeira, um acionista tem interesse em saber qual o risco de ter uma grande perda, ou a chance de ter um grande ganho, para poder fazer uma movimentação de retirada ou acréscimo do montante da aplicação.

Devido a importância que valores extremos possui nas situações citadas, entre outras, a Teoria de Valores Extremos (TVE) tem auxiliado na predição destes valores.

1.1 Estudos em valores extremos

Pesquisas sobre a TVE têm crescido nas últimas décadas, com diversas aplicações

em diversas áreas, principalmente ambiental e econômica. Para isto, foram propostas várias distribuições para dados extremos. Mais especificamente, para a cauda deste tipo de dados, Pickands (1975) introduziu o teorema que diz que se X pertence ao domínio de atração de uma distribuição de valores extremos generalizada GEV (*Generalized Extreme Value distribution*), então os excessos acima de um determinado limiar u podem ser modelados por uma distribuição de Pareto generalizada, conhecida também como GPD (*Generalized Pareto Distribution*). A partir deste teorema, podem ser encontradas boas aproximações, baseado na distribuição GPD para dados acima de um determinado limiar.

Trabalhos baseados em estimação Bayesiana com a cauda possuindo distribuição GPD foram desenvolvidos nos últimos anos. Em Tancredi et al. (2006), a modelagem da parte central dos dados é feita baseada numa mistura de distribuições Uniformes, sendo considerado um modelo Poisson-GPD, onde o número de observações acima de um limiar é tratado como sendo um parâmetro do modelo. Além disso, a modelagem da cauda é feita utilizando os parâmetros em função da distribuição GEV.

Uma outra abordagem para este problema é apresentada em Frigessi et al. (2002). A modelagem é feita dando um peso para o centro e um peso para a cauda da distribuição.

Outro trabalho que aborda de maneira diferente a estimação do limiar está em Bermudez, Turkman e Turkman (2001). Neste trabalho é proposta uma abordagem Bayesiana ao método *Peaks of Threshold* (POT), que é um método que ajusta um modelo estocástico aos excessos ou aos picos acima de determinado limiar. Esta abordagem tem como finalidade estimar quantis extremos além do intervalo dos dados. Apesar da escolha do limiar ser feita pela média de vários modelos prováveis através da abordagem preditiva, ainda assim tem que ser determinado indiretamente, já que a posteriori depende de sua escolha. Para os dados abaixo do limiar, é proposta uma abordagem não-paramétrica baseada na distribuição empírica.

1.2 Estimação não-paramétrica por mistura de distribuições

Métodos computacionalmente mais poderosos têm ajudado na estimação de modelos mais complexos. Uma ilustração interessante neste aspecto é modelo de mistura. Embora estes modelos estejam dentro de uma família paramétrica, eles são uma interessante alternativa à modelagem não-paramétrica (Diebolt e Robert, 1994 e Titterington et al., 1985). O modelo de mistura de distribuições de k componentes é dado por

$$f(x) = \sum_{j=1}^{k} p_j f_j(x), \quad \text{onde } \sum_{j=1}^{k} p_j = 1, \tag{1.1}$$

e as densidades f_j são de uma mesma família, com vetores de parâmetros θ_j, $j = 1, \ldots, k$.

O avanço de técnicas Bayesianas de estimação, como *Markov chain Monte Carlo* (MCMC, para detalhes sobre esta técnica ver Gamerman e Lopes, 2006) foram motivações importantes para o desenvolvimento de técnicas que consistem em mistura de distribuições. O amostrador de Gibbs para estimação de mistura de distribuições foi desenvolvido no trabalho de Diebolt e Robert (1994).

1.3 Objetivo

O objetivo deste trabalho é propor uma nova metodologia para analisar dados extremos, dando um enfoque especial para a cauda da distribuição. Baseado na TVE, a partir de um limiar grande, a cauda deste tipo de dado pode ser modelada por uma distribuição GPD. Para a parte central dos dados, pode-se a princípio supor uma ampla classe de distribuições candidatas para realizar a modelagem. Uma metodologia que abrange este leque de classes consiste em uma abordagem não-paramétrica, baseado numa aproximação por mistura finita de distribuições de uma mesma família paramétrica.

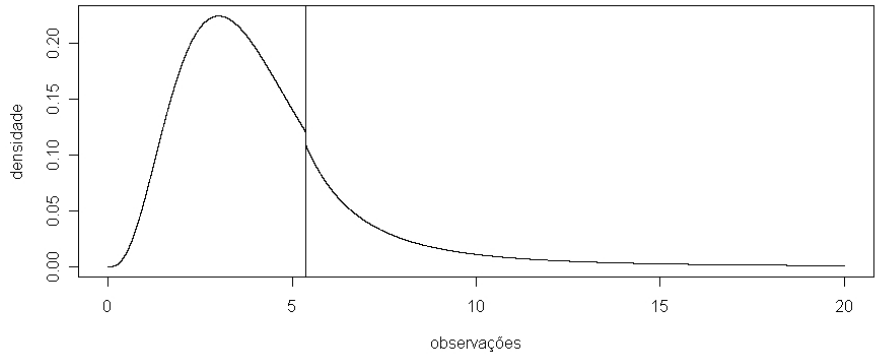
Este trabalho consiste numa extensão do trabalho de Behrens et al. (2004), que realizou uma estimativa através de uma abordagem Bayesiana para dados extremos no

modelo, considerando uma distribuição paramétrica conhecida para a parte central da distribuição, e uma distribuição GPD acima de um limiar. A função de distribuição do modelo é dada por

$$F(x \mid \Theta, \Psi) = \begin{cases} H(x \mid \Theta), & \text{se} \quad x < u \\ H(u \mid \Theta) + [1 - H(u|\Theta)]G(x|\Psi), & \text{se} \quad x \geq u \end{cases}, \tag{1.2}$$

onde Θ são os parâmetros da parte central dos dados, Ψ são os parâmetros da cauda, $H(.|\Theta)$ é a função de distribuição da parte central dos dados, onde Behrens et al. (2004) considera H sendo Gama, e $G(.|\Psi)$ é a função de distribuição GPD, para as observações acima do limiar, que depende de $\Psi = (\xi, \sigma, u)$, onde u é o limiar. Um exemplo de mistura de uma Gama com GPD é dada pela Figura 1.1.

Figura 1.1: Densidade de mistura de uma Gama com GPD.
A linha vertical representa o limiar, a curva à esquerda é a densidade da Gama e a curva a direita é a densidade da GPD.



Neste trabalho, será realizada uma abordagem Bayesiana não-paramétrica para a função de distribuição H em (1.2), assumindo que H é uma mistura finita de distribuições de uma mesma família como em (1.1). Baseado no trabalho de Wiper et al. (2001), foi utilizada a família Gama para mistura de distribuições, que é a mais adequada para dados que pertencem aos reais positivos. Pela Figura 1.1, percebe-se que há um salto no limiar. Este salto pode não ocorrer na prática, mas neste modelo, a estimação abrange as duas

situações, com e sem o salto no limiar, pois aqui o mais importante é determinar o limiar onde ocorre a mudança de distribuição.

Algumas das modelagens propostas serão aplicadas em dados ambientais, em que o espaço amostral são os reais positivos, tais como níveis de chuva e vazão de rios, a família da distribuição da mistura também tem que pertencer aos reais positivos. Assim, a distribuição escolhida para componente da mistura também tem que ter suporte em $\mathcal{R}_+$. Este trabalho utilizou mistura de distribuições Gama, baseado no trabalho de Wiper et al. (2001), que mostrou que a Gama é a família mais adequada para realizar uma aproximação não-paramétrica por misturas. Para dados que também assumem valores negativos, pode-se considerar também mistura de distribuições Normais, como já foram desenvolvidos em trabalhos como em Diebolt e Robert (1994), Roeder e Wasserman (1997), Richardson e Green (1997) entre outros.

Um outro aspecto importante na mistura de distribuições está em determinar qual o melhor número de componentes a ser utilizado. Neste trabalho, o número de componentes foi escolhido de acordo com critérios de comparação de modelos, como o BIC (*Bayesian Information Criterion*) e o DIC (*Deviance Information Criterion*).

1.4 Estrutura do trabalho

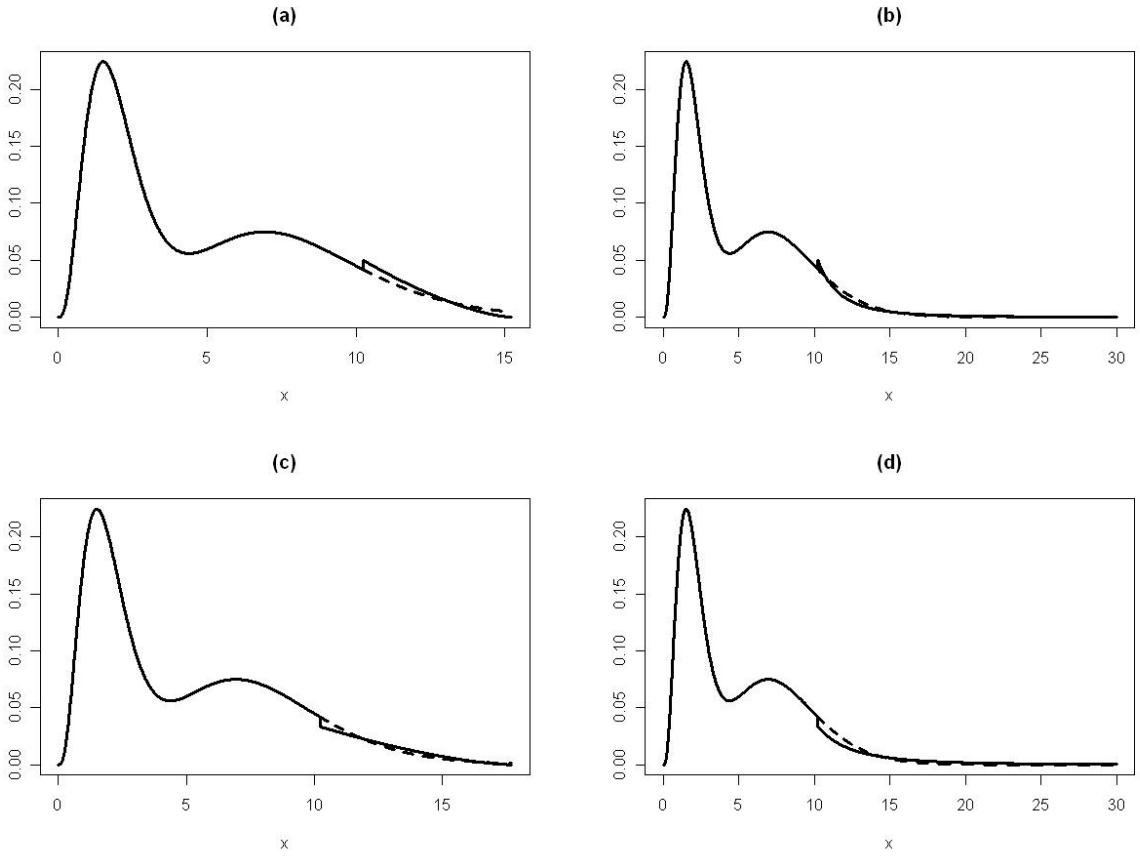
O Capítulo 2 irá mostrar o modelo de mistura de distribuições Gama, mostrando como utilizar técnicas Bayesianas via MCMC para estimação da distribuição a posteriori dos parâmetros das componentes da mistura. Também serão comentados alguns tópicos importantes na mistura, como imposição de uma restrição nos parâmetros das componentes, para evitar o problema de não-identificabilidade do modelo.

O Capítulo 3 irá apresentar uma nova metodologia para avaliar dados extremos, considerando uma mistura de distribuições Gama para o centro, e distribuição GPD para a cauda. Exemplos de mistura de Gamas com GPD é dada pela Figura 1.2. Embora visualmente pareça haver pouca diferença entre o modelo proposto e o que utiliza só mistura,

esta diferença é notória no cálculo de quantis extremos e máximos da distribuição, que são as medidas mais importantes quando se analisa valores extremos. Foram analisados dois conjuntos de dados. O primeiro consiste em dados de vazão de rios em Porto Rico. Os dados estão disponíveis no site `http://waterdata.usgs.gov`. O segundo conjunto de dados consiste em níveis de chuvas em Portugal. Os dados podem ser obtidos no site `http://snirh.pt/`.

Figura 1.2: Funções de densidade do modelo proposto

(a) - $\sigma = 2$ e $\xi = -0,4$; (b) - $\sigma = 2$ e $\xi = 0,4$; (c) - $\sigma = 3$ e $\xi = -0,4$; (d) - $\sigma = 3$ e $\xi = 0,4$. A parte central dos dados é uma mistura de duas Gamas. As linhas tracejadas representam a continuação da mistura de Gamas além do limiar.



O Capítulo 4 consiste em realizar uma extensão do modelo proposto no Capítulo 3,

considerando estrutura de modelos de regressão para os parâmetros da cauda. Nesta estrutura, os parâmetros da distribuição GPD estão em função de variáveis explicativas. Assim, estes parâmetros são escritos como

$$\Psi = g(\mathbf{z}, \gamma),$$

onde $\mathbf{z}$ é o vetor das variáveis explicativas com vetor de parâmetros γ. Esta estrutura é baseada no trabalho de Cabras et al.(2009), onde é feita uma reparametrização dos parâmetros da GPD, encontrado uma distribuição a priori de Jeffreys com distribuição Uniforme para os parâmetros da cauda, modelados por uma estrutura de regressão. Como Cabras et al. (2009) considera o limiar fixo, neste trabalho foi proposta uma modelagem do limiar utilizando uma estrutura de regressão linear. Neste capítulo, foram analisados dois conjuntos de dados, o primeiro de temperaturas máximas nos Estados Unidos. Os dados podem ser obtidos no site `www.engr.udayton.edu/weather`. O segundo conjunto de dados analisado neste capítulo foi de temperaturas mínimas no estado do Rio de Janeiro.

O Capítulo 5 trata a variação dos parâmetros da distribuição GPD ao longo do tempo, considerando os parâmetros de forma e escala variando no tempo. Foi considerada uma estrutura de modelos dinâmicos de primeira ordem, baseado em Lopes et al. (2009), que consiste na seguinte estrutura

$$\begin{aligned}
l\xi_t &= \theta_{\xi,t} + v_{\xi,t}, \quad v_{\xi,t} \sim N(0, 1/V_\xi), \\
\theta_{\xi,t} &= \theta_{\xi,t-1} + w_{\xi,t}, \quad w_{\xi,t} \sim N(0, 1/W_\xi), \\
l\sigma_t &= \theta_{\sigma,t} + v_{\sigma,t}, \quad v_{\sigma,t} \sim N(0, 1/V_\sigma), \\
\theta_{\sigma,t} &= \theta_{\sigma,t-1} + w_{\sigma,t}, \quad w_{\sigma,t} \sim N(0, 1/W_\sigma),
\end{aligned}$$

onde $\xi_t = \exp(l\xi_t) - 1$ e $\sigma_t = \exp(l\sigma_t)$. Este modelo foi aplicado em dados do mercado financeiro, em retornos absolutos de bolsa de valores do Brasil, no índice da BOVESPA, Petrobras e Vale do Rio Doce.

1.5 Distribuições

Esta seção apresenta algumas das distribuições que serão descritas neste trabalho e suas propriedades.

1.5.1 Distribuição Gama

A função densidade da distribuição Gama é denotada por $G(\mu, \eta)$, com $\mu = \eta/\beta$ é dada por

$$f_G(x|\eta/\beta, \eta) = \begin{cases} (1/\Gamma(\eta))\beta^\eta x^{\eta-1}\exp(-\beta x), & \text{se} \quad x \geq 0 \\ 0, & \text{se} \quad x < 0. \end{cases} \tag{1.3}$$

Na distribuição Gama, a esperança e a variância são dadas por $E(X) = \mu$ e $V(X) = \eta/\beta^2$.

Neste trabalho, ao invés de trabalhar com a notação da distribuição Gama como em (1.3), é mais conveniente realizar uma reparametrização mantendo η, mas agora considerando $\mu = \eta/\beta$ um parâmetro da distribuição. Assim, a média da distribuição Gama é o parâmetro μ. Realizando a reparametrização, a função de densidade da distribuição $G(\mu, \eta)$ será dada por:

$$f_G(x|\mu, \eta) = \begin{cases} (1/\Gamma(\eta))(\eta/\mu)^\eta x^{\eta-1}\exp(-(\eta/\mu)x), & \text{se} \quad x \geq 0 \\ 0, & \text{se} \quad x < 0. \end{cases} \tag{1.4}$$

Na distribuição Gama reparametrizada, a esperança é dada por $E(X) = \mu$ e a variância é $V(X) = \mu^2/\eta$.

1.5.2 Distribuição Gama inversa

Se X possui distribuição Gama com parâmetros η e β, então $1/X$ possui distribuição Gama inversa, denotada por $IG(\eta, \beta)$, com a seguinte função de densidade

$$f_{IG}(x|\eta,\beta) = \begin{cases} (1/\Gamma(\eta))\beta^{\eta}x^{-\eta-1}\exp(-\beta/x), & \text{se} \quad x \geq 0 \\ 0, & \text{se} \quad x < 0. \end{cases} \tag{1.5}$$

Na distribuição Gama inversa, a esperança é dada por $E(X) = \beta/(\eta-1)$ para $\eta > 1$ e a variância é dada por $V(X) = \frac{\beta^2}{(\eta-1)^2(\eta-2)}$ para $\eta > 2$.

1.5.3 Distribuição Dirichlet

Um vetor de variáveis $\mathbf{X} = (X_1, \ldots, X_k)$, com $\sum_{i=1}^{k} X_i = 1$ possui distribuição Dirichlet com parâmetros $\gamma_1, \ldots, \gamma_k$, e é denotada por $D_k(\gamma_1, \ldots, \gamma_k)$, se possui a seguinte função densidade:

$$f_D(\mathbf{x} \mid \gamma_1, \ldots, \gamma_k) = \frac{(\Gamma(\gamma_0))}{\left(\prod_{j=1}^{k}\Gamma(\gamma_j) - 1\right)} \prod_{j=1}^{k} x_j^{\gamma_j - 1}, \text{ se } \; x_i \geq 0, i = 1, \ldots, k,$$

onde $\gamma_0 = \sum_{j=1}^{k}\gamma_j$, $j = 1, \ldots, k$. Na distribuição Dirichlet, a esperança das marginais são $E(X_i) = \gamma_i/\gamma_0$, $i = 1, \ldots, k$ e variância $V(X_i) = \frac{\gamma_i(\gamma_0-\gamma_i)}{\gamma_0^2(\gamma_0+1)}$, para $i = 1, \ldots, k$.

1.5.4 Distribuição Multinomial

O vetor de variáveis $(X_1, \ldots, X_k)$, com $X_j \in \{0, 1\}$, $j = 1, \ldots, k$ e $\sum_{i=1}^{k} X_i = n$, possui distribuição multinomial com parâmetros $p_1, \ldots, p_k$, com $\sum_{j=1}^{k} p_j = 1$ e denotada por $MN_k(n; p_1, \ldots, p_k)$, se possui a seguinte função de probabilidade:

$$P_{MN}(X_1 = x_1, \ldots, X_k = x_k \mid p_1, \ldots, p_k) = \frac{n!}{\prod_{j=1}^{k} x_j!} \prod_{j=1}^{k} p_j^{x_j}, \tag{1.6}$$

onde $E(X_i) = np_i$ e $V(X_i) = np_i(1-p_i)$, $i = 1, \ldots, k$.

1.5.5 Distribuição Normal truncada

Denotando por $N(a, b)$ a distribuição Normal com $E(X) = a$ e $V(X) = b$, se $X \sim N(a, b)$ com função densidade

$$f_N(y) = \frac{1}{\sqrt{2\pi b}} e^{-\frac{(x-a)^2}{2b}}. \tag{1.7}$$

Se X é a distribuição de Y truncada em um intervalo (c, d), X possui distribuição Normal truncada, denotada por $N(a, b)I(c, d)$, e possui a seguinte função de densidade

$$f_{NT}(x|a, b, c, d) = \frac{\phi\left(\frac{x-a}{\sqrt{b}}\right)}{\Phi\left(\frac{d-a}{\sqrt{b}}\right) - \Phi\left(\frac{c-a}{\sqrt{b}}\right)}, \quad c \leq x \leq d, \tag{1.8}$$

onde ϕ e Φ são respectivamente a função de densidade e de distribuição da $N(0, 1)$.

Capítulo 2

Estimação não-paramétrica por mistura

Neste capítulo, será feita uma revisão dos modelos de mistura finita de distribuições, com objetivo de entender melhor este método, usado principalmente para estimação não-paramétrica de curvas. Os resultados obtidos nos exercícios de simulação utilizados neste capítulo servem de base para a aplicação dos modelos de mistura para os modelos propostos neste trabalho, que serão vistos nos capítulos seguintes para valores extremos.

Neste capítulo, são comparados dois diferentes algoritmos para mistura de curvas de distribuição, um primeiro que utiliza diretamente a distribuição a posteriori do vetor de parâmetros, e um segundo que insere uma variável latente que facilita os cálculos da distribuição a posteriori de cada parâmetro.

2.1 Introdução

Métodos não-paramétricos são utilizados quando há incerteza sobre qual é a distribuição de um conjunto de dados, ou quando esta distribuição não possui forma paramétrica conhecida. Modelos baseados em mistura de distribuições fornecem uma interessante alternativa para modelagem não-paramétrica. O avanço de técnicas computacionais tem

ajudado para um melhor desenvolvimento nesta área, pois possibilitaram trabalhar com estimação de modelos mais complexos.

Abordagem Bayesiana para a estimação de mistura não havia sido bem desenvolvida até o início da década de 1990, devido principalmente a obstáculos computacionais. Diebolt e Robert (1994) propuseram a inserção de variáveis latentes para poder amostrar distribuições de mistura utilizando o amostrador de Gibbs. Celeux et al. (2000) estudaram as dificuldades na estimação da distribuição a posteriori em modelos de mistura pela técnica MCMC. Jasra et al. (2005) revisam alguns métodos de estimar mistura de distribuições, citando o amostrador de Gibbs (Diebold e Robert, 1994), *Tempering* MCMC (Neal, 1996) e *Relabelling Algorithms* (Stephens, 1997).

2.2 Inferência Bayesiana

A abordagem Bayesiana é desenvolvida na presença de observações x cujos valores inicialmente vêm de distribuições incertas e descritos por uma função de densidade $f(x \mid \theta)$. A quantidade θ representa uma característica de interesse que descreve o processo da distribuição da variável. A situação canônica é aquela na qual uma amostra aleatória $X = (X_1, ..., X_n)$ segue uma distribuição com densidade $f(X \mid \theta)$. Neste caso, as observações são iid (independentes e identicamente distribuidas), condicional em θ. A quantidade θ não é somente um índice, mas o principal interesse de estudo, pois determina a característica da população.

2.2.1 Distribuição a priori e a posteriori

É desejável que o pesquisador tenha algum conhecimento sobre o parâmetro de interesse θ. Este conhecimento pode ser formalmente incorporado na análise. A abordagem Bayesiana incorpora esta informação na análise através de uma densidade $p(\theta)$, chamada de distribuição a priori, que geralmente é uma informação subjetiva.

A inferência Bayesiana contém dois ingredientes: a distribuição observacional $f(x \mid \theta)$

e a distribuição $p(\theta)$. Esta última distribuição pode também ser especificada com ajuda de outros parâmetros, chamados de hiperparâmetros, ou seja, são os parâmetros da distribuição dos parâmetros. A distribuição observacional de uma amostra fornece a função de verossimilhança $l(\theta) = f(x \mid \theta)$. Com estes dois ingredientes, pode-se encontrar a distribuição de θ após observar x. Esta distribuição é chamada de distribuição a posteriori, que é obtida através do Teorema de Bayes

$$\pi(\theta \mid x) = \frac{f(x \mid \theta)p(\theta)}{\int f(x \mid \theta)p(\theta)d\theta},$$

onde o denominador é uma constante em relação a θ. Assim, a distribuição a posteriori pode ser dada por $\pi(\theta \mid x)$, escrita de uma forma mais compacta por

$$\pi(\theta \mid x) \propto f(x \mid \theta)p(\theta)$$

Em uma amostra aleatória de tamanho n, a função de verossimilhança é dada por $l(\theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$ e assim a distribuição a posteriori de θ dado o vetor $(x_1, ..., x_n)$ é proporcional a

$$\pi(\theta \mid x_1, ..., x_n) \propto l(\theta)p(\theta)$$

2.3 Monte Carlo via cadeias de Markov

A técnica de Monte Carlo via cadeias de Markov, conhecida como MCMC, que pode ser vista em referências como, por exemplo, Gamerman e Lopes (2006), tem uma vasta área de aplicação em modelos Bayesianos. Esta técnica consiste em simular pontos de uma distribuição multivariada. Antes de mostrar a técnica, considere a identidade a posteriori:

$$P(\theta \mid Y) = \int_Z P(\theta \mid Y, Z)P(Z \mid Y)dZ,$$

e a identidade preditiva

$$P(Z \mid Y) = \int_\theta P(Z \mid Y, \theta)P(\theta \mid Y)d\theta$$

uma iteração entre os dois, tomando a estimativa corrente da identidade a posteriori, simulando uma realização de Z da identidade preditiva, e usando estas realizações para re-estimar a identidade a posteriori.

Uma abordagem modificada é a chamada *chained data algorithm*, onde são feitas iterações sucessivas de $P(Z \mid \theta, Y)$ e $P(\theta \mid Z, Y)$ para obter uma sequência de realizações $(Z^{(1)}, \theta^{(1)})$,...,$(Z^{(n)}, \theta^{(n)})$, onde em cada caso as simulações são feitas condicionadas aos valores correntes das variáveis. Intuitivamente, espera-se que esta cadeia encontre uma distribuição de equilíbrio $P(\theta, Z \mid Y)$. Além disso, a partir de um certo período de aquecimento, chamado de *burn-in*, a sequência $(Z^{(k)}, \theta^{(k)})$,...,$(Z^{(n)}, \theta^{(n)})$ são realizações da distribuição proposta.

2.3.1 Amostrador de Gibbs

O amostrador de Gibbs generaliza o *chained data algorithm* para a situação multivariada. Para obter uma amostra da distribuição multivariada $p(\theta_1, ..., \theta_d)$, o procedimento é dado por

Algoritmo Gibbs

1. Inicializar $\theta = (\theta_1^{(0)}, ..., \theta_d^{(0)})$
2. Simular $\theta_1^{(1)}$ da condicional $\theta_1 \mid \theta_2^{(0)}, ..., \theta_d^{(0)}$
3. Simular $\theta_2^{(1)}$ da condicional $\theta_2 \mid \theta_1^{(1)}, \theta_3^{(0)}, ..., \theta_d^{(0)}$
4. ...
5. Simular $\theta_d^{(1)}$ da condicional $\theta_d \mid \theta_1^{(1)}, ..., \theta_{d-1}^{(0)}$
6. Iterar o procedimento

Assim, após um período de *burn-in*, $(\theta_1^{(k)}, ..., \theta_d^{(k)})$,...,$(\theta_1^{(n)}, ..., \theta_d^{(n)})$ são realizações da distribuição de interesse.

Pela forma do Teorema de Bayes, onde a distribuição a posteriori é dada por

$$p(\theta \mid x) = \frac{p(\theta)p(x \mid \theta)}{\int p(\theta)p(x \mid \theta)d\theta},$$

o amostrador de Gibbs pode obter amostras diretamente da distribuição a posteriori

conjunta utilizando as distribuições do algoritmo, conhecidas também como distribuições condicionais completas, sem ter que se preocupar com a integral do denominador.

2.3.2 Metropolis-Hastings

O algoritmo mais Geral do MCMC é o Metropolis-Hastings. Suponha que o objetivo é simular pontos de uma distribuição (multivaridada) $\pi(\theta)$. Seja $q(\theta, \delta)$ uma probabilidade de transição arbitrária. O algoritmo de Metropolis-Hastings é dado da seguinte maneira

Algoritmo M-H

1. Dada a posição corrente de $\theta_n = \theta$ gerar um novo candidato δ^* de $q(\theta, \delta)$
2. Calcular $\alpha(\theta, \delta) = \min\left\{\frac{\pi(\delta)q(\theta,\delta)}{\pi(\theta)q(\theta,\delta)}, 1\right\}$, com $\delta = \delta^*$.
3. Com probabilidade $\alpha(\theta, \delta^*)$ aceitar $\theta_{n+1} = \delta^*$ caso contrário manter $\theta_{n+1} = \theta$.
4. Voltar ao passo 1.

A vantagem do Algoritmo de Metropolis-Hastings em relação ao amostrador de Gibbs é que não é necessário conhecer todas as distribuições condicionais completas. A variável θ pode ser um vetor de parâmetros $(\theta_1, ..., \theta_d)$.

Além destes algoritmos, outras variações podem ser propostas no MCMC. Por exemplo, fazer o algoritmo de Gibbs por passos de Metropolis, onde a distribuição condicional completa de cada parâmetro é a distribuição $\pi(x)$ do algoritmo de Metropolis-Hastings, e Metropolis por blocos, onde blocos $(\theta_1, ..., \theta_{k_1})$,...,$(\theta_{k_{p-1}}, ..., \theta_{k_p})$ são amostrados um por vez pelo algoritmo de Metropolis-Hastings.

2.4 Modelo de mistura

Uma distribuição de mistura de k componentes tem função de densidade dada por

$$h(x|\theta, \mathbf{p}) = \sum_{j=1}^{k} p_j f_j(x|\theta_j),$$

onde p_j é o peso associado à componente j, com função de densidade $f_j(.|\theta_j)$ com vetor de parâmetros θ_j, $j = 1, \ldots, k$. A restrição nos pesos é dada por $\sum_{j=1}^{k} p_j = 1$. Neste modelo,

os parâmetros são os pesos p_j e os parâmetros de cada componente θ_j, $j = 1, \ldots, k$.

Para a família de componentes f_j, é necessário escolher a família mais adequada para variáveis que pertencem aos reais positivos.

2.4.1 Aproximação não-paramétrica por mistura

Modelos de mistura de Gamas fornecem uma boa aproximação a estimação não-paramétrica de curvas nos dados reais positivos. De Vore e Lorenz (1993, pág 14), mostraram para qualquer função contínua $f(x)$ em $(0, \infty)$, com limite zero quando $x \to \infty$, que $\lim_{u\to\infty} S_u(x) = f(x)$ uniformemente para $0 < x < \infty$, onde $S_u(x)$ é definido por

$$S_u(x) = e^{-ux} \sum_{k=0}^{\infty} \frac{(ux)^k}{k!} f\left(\frac{k}{u}\right), \quad u > 0.$$

Além disso, f pode ser aproximada por uma mistura de densidades Gama $G(k+1/u, k+1)$ com pesos $(1/u)f(k/u)$. Asmussen (1987, pág 76) mostrou que mistura de distribuições Erlang convergem fracamente para qualquer medida de probabilidade em (0,∞).

Em particular, Wiper et al. (2001) propõem utilizar mistura de distribuições Gama para estimação de densidade não- paramétrica, pois a família Gama é mais ampla que as citadas previamente, não tendo um dos parâmetros restrito a um inteiro. Assim, as consequências são aproximações mais parcimoniosas que as citadas previamente.

O modelo de mistura de Gamas com k componentes, com vetor de parâmetros $(\mu, \eta, \mathbf{p})$, é denotado por $MG_k(\mu, \eta, \mathbf{p})$, ou MG_k, quando não for necessário explicitar os parâmetros. Este modelo possui a seguinte função de densidade:

$$h(x|\mu, \eta, \mathbf{p}) = \sum_{j=1}^{k} p_j f_G(x|\mu_j, \eta_j).$$

onde f_G é a função de densidade da distribuição Gama,

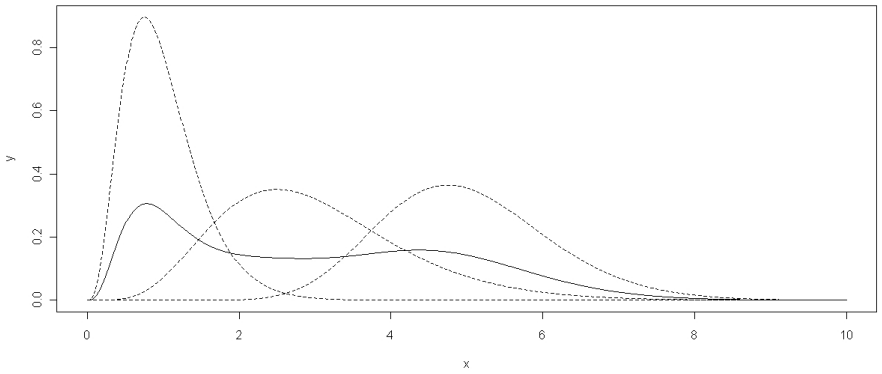
Um exemplo da mistura de Gamas e suas marginais está na Figura 2.1

2.5 Restrição nos parâmetros

Um outro aspecto importante quando é feita a modelagem Bayesiana não-paramétrica

Figura 2.1: Densidade de mistura de 3 Gamas.

Linhas tracejadas: componentes da mistura. Linha cheia: mistura das três componentes.



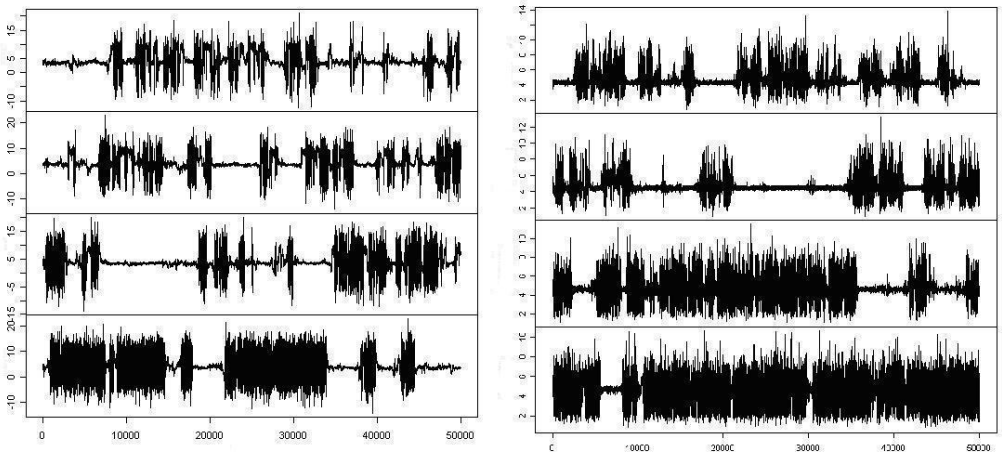
baseada em mistura de distribuições está na identificabilidade das componentes da mistura. Realizando a estimação sem nenhuma restrição nos parâmetros, pode ocorrer do método utilizado na estimação não identificar quais são cada uma das componentes ou misturá-las.

Um exemplo da falta de identificabilidade da distribuição de mistura pode ser visto em exemplos com dados simulados. Resultados de simulações feitas para distribuição Gama e Normal mostram este tipo de comportamento. O gráfico à esquerda da Figura 2.2 mostra a série da distribuição a posteriori das médias de mistura de quatro Normais. Pelo gráfico, percebe-se que a variância ao longo de uma série muda de comportamento, ao mesmo tempo em que a variância de uma componente aumenta, a outra diminui, ou seja, uma componente passa a ter as características da outra. Fazendo a mistura de Gamas, a parte à direita da Figura 2.2 mostra que este comportamento também acontece para o parâmetro de média da distribuição Gama.

Um dos primeiros trabalhos a discutir esta restrição foi Richardson e Green (1997), onde no caso de mistura de distribuição Normais, é sugerido impor a restrição nas médias $(\mu_1, \mu_2, \ldots, \mu_k)$, de tal forma que $\mu_1 < \mu_2 < \ldots < \mu_k$.

Este problema foi amplamente discutido no trabalho de Frühwirth-Schnatter (2001),

Figura 2.2: Problemas da não-identificabilidade para misturas Normal e Gama.
A série à esquerda mostra um exemplo de estimação para 4 componentes para as médias da mistura de Normais, e à direita a média de mistura de Gamas.



onde no caso da mistura de duas distribuições Normais, foi necessário impor a condição de que a média na primeira componente é sempre menor do que a segunda. Este trabalho mostrou também que a não imposição da condição pode levar ao problema da distribuição a posteriori possuir dois máximos locais, e com isso o modelo pode convergir para uma distribuição que não é a verdadeira. Frühwirth-Schnatter (2001) também mostrou que a imposição da restrição na variância da mistura de Normais não é necessária e pode levar o algoritmo MCMC a amostrar de uma distribuição a posteriori truncada.

Ao impor a restrição na média, utilizar a função Gama reparametrizada em (1.4) tem a vantagem da restrição de identificabilidade estar imposta somente em um parâmetro, ao contrário do caso da forma da distribuição Gama em (1.3), que levaria a restrição nos dois parâmetros da distribuição Gama. Assim, é utilizada a forma da distribuição Gama reparametrizada e imposta a restrição de identificabilidade $R = \{(\mu_1, ..., \mu_k) \in (0, \infty); \mu_1 < \ldots < \mu_k\}$.

Embora a restrição seja importante para identificar os parâmetros de cada componente

da mistura, ela não é necessária para estimar corretamente a curva de densidade da distribuição de mistura. Ao fazer a estimação sem a restrição, embora os valores dos parâmetros não sejam necessariamente os mesmos que os valores corretos, a curva de densidade será a mesma curva da estimação que impõe a restrição nas médias.

2.6 Distribuição a priori para mistura

Para o parâmetro η_j, observa-se na densidade da Gama em (1.4) que ele não possui uma distribuição a priori conjugada conhecida. Como este parâmetro só assume valores positivos, a priori escolhida foi a Gama.

Para o parâmetro μ_j, percebe-se, pela função densidade da Gama em (1.4), fixando η_j, que uma priori conjugada para esta distribuição é a Gama inversa. Na priori, também é necessário impor a restrição nos parâmetros R, citada na Seção 2.5.

Para os pesos, por serem variáveis que estão entre 0 e 1 e a soma ser 1, foi atribuída uma priori com distribuição Dirichlet para o vetor de pesos.

A princípio, tem-se que

$$\eta_j \sim G(a_j/b_j, a_j), j = 1, \ldots, k, \quad \mu_j \sim IG(c_j, d_j), j = 1, \ldots, k \text{ e } \mathbf{p} \sim D(\gamma_1, \ldots, \gamma_k).$$

Para o vetor de parâmetros $(\mu_1, \ldots, \mu_k)$, considerando a restrição na média dos parâmetros, a priori pode ser escrita da seguinte maneira:

$$p(\mu_1, \ldots, \mu_k) = \prod_{i=1}^{k} f_{IG}(\mu_i \mid c_i, d_i) = W \prod_{i=1}^{k} f_{IG}(\mu_i \mid c_i, d_i) I_R(\underline{\mu}),$$

onde $W^{-1} = \int_{I_R(\underline{\mu})} \prod_{i=1}^{k} p(\mu_i) d(\mu_1, \ldots, \mu_k)$ e $I_R(\underline{\mu})$ é a função indicadora da restrição nas médias.

Supondo independência entre esses grupos de parâmetros η, μ e $\mathbf{p}$, e denotando o vetor de todos os parâmetros do modelo por Θ, tem-se que

$$\Theta = (\eta_1, \ldots, \eta_k, \mu_1, \ldots, \mu_k, p_1, \ldots, p_k).$$

A densidade a priori de Θ é dada por

$$\pi(\Theta) \propto \prod_{i=1}^{k} \left(\eta_j^{a_j - 1} \exp(-b_j \eta_j) \mu_j^{-c_j - 1} \exp(-d_j/\mu_j) p_j^{\gamma_j - 1} \right) I_R(\underline{\mu}).$$

2.7 Distribuição a posteriori

Considerando a função de verossimilhança de Θ baseada no modelo de mistura, juntamente com a distribuição a priori de Θ, obtém-se, pelo Teorema de Bayes, a seguinte forma da distribuição a posteriori

$$\pi(\Theta|\mathbf{x}) \propto \prod_{i=1}^{n} \left(\sum_{j=1}^{k} p_j f_G(x_i|\mu_j, \eta_j) \right) \prod_{j=1}^{k} \left(f_G(\eta_j|a_j/b_j, a_j) f_{IG}(\mu_j|c_j, d_j) p_j^{\gamma_j - 1} \right) I_R(\underline{\mu})$$

$$\propto \prod_{i=1}^{n} \left(\sum_{j=1}^{k} \left(p_j \frac{1}{\Gamma(\eta_j)} \left(\frac{\eta_j}{\mu_j} \right)^{\eta_j} x_i^{\eta_j} e^{-\left(\frac{\eta_j}{\mu_j}\right) x_i} \right) \right) \prod_{j=1}^{k} \left(\eta_j^{a_j - 1} e^{-b_j \eta_j} \mu_j^{-c_j - 1} e^{-d_j/\mu_j} p_j^{\gamma_j - 1} \right) I_R(\underline{\mu}), \qquad (2.1)$$

onde $\mathbf{x} = (x_1, \ldots, x_n)$ são as observações.

Pode-se notar que a equação (2.1) não tem uma forma conhecida para nenhum parâmetro, devido ao fato da função densidade conter uma série de parcelas que não se combinam analiticamente.

Neste caso, uma primeira alternativa para fazer inferência seria utilizar MCMC pelo algoritmo de Gibbs em blocos com passos de Metropolis, para amostrar da distribuição condicional completa em cada vetor de parâmetros da componente j da mistura, além do vetor de pesos. O algoritmo para amostrar pontos da distribuição a posteriori dos parâmetros é dado no Apêndice 2A deste capítulo.

Para sintonizar a variância das distribuições propostas, foi utilizado o método de Roberts e Rosenthal (2006), que diz que o valor ótimo da taxa de aceitação do algoritmo de Metropolis para propostas unidimensionais é aproximadamente 0,44.

Seja $V^{(l)}$ a variância da proposta no início das iterações com $l = 0$. A cada 50 iterações, o contador l vai para $l + 1$ e verifica-se a taxa de aceitação. Se esta é menor do que 0,44, então $\log\left(V^{(l+1)}\right) = \log\left(V^{(l)}\right) + \delta(l+1)$ caso contrário $\log\left(V^{(l+1)}\right) = \log\left(V^{(l)}\right) - \delta(l+1)$, onde $\delta(l+1) = \min(0,01; (l+1)^{-0,5})$.

2.8 Inserção de uma variável latente

Para facilitar os cálculos da distribuição a posteriori, Diebolt e Robert (1994) utilizaram a ideia proposta em Tanner e Wong (1987). Eles abordaram o problema inserindo uma variável aleatória latente $\mathbf{z}$ de dimensão $n \times k$, em que $z_{i,j}$ é uma variável que indica em qual componente j da mistura está a observação i, assumindo valor 1 para a componente em que ela está e 0 nas outras componentes. Além disso, utilizando o algoritmo MCMC, Diebolt e Robert (1994) mostraram que amostrar θ da distribuição a posteriori $\pi(\theta|\mathbf{z}, x)$ equivale a amostrar θ da distribuição a posteriori $\pi(\theta|x)$. Desta maneira,

$$f(x \mid \mathbf{z}) = \prod_{j=1}^{k} f_j(x)^{z_j}.$$

Sendo $\mathbf{z} \sim MN_k(1; p_1, \ldots, p_k)$, então, $f_{MN}(\mathbf{z}) = \prod_{j=1}^{k} p_j^{z_j}$ e,

$$f(x) = \sum_{j=1}^{k} f(x, \mathbf{z}_j) = \sum_{j=1}^{k} f(x \mid \mathbf{z}_j) f_{MN}(\mathbf{z}_j) = \sum_{l=1}^{k} \left(\prod_{j=1}^{k} \left(f_j(x) p_j \right)^{z_{j,l}} \right) = \sum_{l=1}^{k} p_l f_l(x).$$

Portanto,

$$h(x \mid \theta, \mathbf{p}, \mathbf{z}) = \prod_{j=1}^{k} (p_j f(x|\theta_j))^{z_j}, \tag{2.2}$$

onde $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ e $\mathbf{z}_i = (z_{i,1}, \ldots, z_{i,k})$, $i = 1, \ldots, n$.

Como o valor da variável $\mathbf{z}$ não é conhecido, ela também é considerada mais um parâmetro no modelo. Utilizando as mesmas distribuições a priori da seção anterior, para uma amostra de n observações, a distribuição a posteriori terá a seguinte forma

$$\begin{aligned}
\pi(\Theta, \mathbf{z}|\mathbf{x}) &\propto \prod_{i=1}^{n} \left(\prod_{j=1}^{k} (f_G(x_i|\mu_j, \eta_j))^{z_{i,j}} p_j^{z_{i,j}} \right) \times \\
\times \quad & \prod_{i=1}^{k} \left(\eta_j^{a_j - 1} \exp(-b_j \eta_j) \mu_j^{-c_j - 1} \exp(-d_j/\mu_j) p_j^{\gamma_j - 1} \right) I_R(\underline{\mu}) \\
\propto \quad & \prod_{j=1}^{k} \left(\left(\frac{1}{\Gamma(\eta_j)} p_j \left(\frac{\eta_j}{\mu_j} \right)^{\eta_j} \right)^{\sum_{i=1}^{n} z_{i,j}} \left(\prod_{i=1}^{n} x_i^{z_{i,j}(\eta_j - 1)} \right) \exp\left(- \sum_{i=1}^{n} z_{i,j} \left(\frac{\eta_j}{\mu_j} \right) x_i \right) \right) \times
\end{aligned}$$

$$\times \quad \left(\eta_j^{a_j-1}\exp(-b_j\eta_j)\mu_j^{-c_j-1}\exp(-d_j/\mu_j)p_j^{\gamma_j-1}\right) I_R(\underline{\mu})$$

$$\propto \quad \prod_{j=1}^{k}\left(p_j^{\left(\sum_{i=1}^n z_{i,j}+\gamma_j-1\right)}\mu_j^{-c_j-\eta_j\sum_{i=1}^n z_{i,j}-1}\exp\left(-\left(d_j+\eta_j\sum_{i=1}^n z_{i,j}x_i\right)/\mu_j\right)\times\right.$$

$$\left.\times \quad \frac{\eta_j^{a_j+\eta_j\sum_{i=1}^n z_{i,j}-1}}{\Gamma(\eta_j)^{\sum_{i=1}^n z_{i,j}}}\left(\prod_{i=1}^{n}x_i^{z_{i,j}(\eta_j-1)}\right)\exp(-b_j\eta_j)\right) I_R(\underline{\mu}). \tag{2.3}$$

Reescrevendo a posteriori desta maneira, com exceção de $\eta_j, j = 1,\ldots,k$, todos os outros parâmetros têm distribuições condicionais completas conhecidas. Detalhes das condicionais completas são dados no Apêndice 2B deste capítulo.

2.9 Comparação de modelos

Em uma estimação utilizando abordagem não-paramétrica por mistura de distribuições, é necessário saber qual o número de componentes mais adequado em cada caso. Para isso, há duas alternativas: considerar k um parâmetro do modelo e estimar por RJMCMC (*Reversible Jump Markov Chain Monte Carlo*, Richardson e Green, 1997), ou comparar modelos com diferentes valores de k através de um critério de comparação de modelos.

Como nos dados que possuem caudas com distribuição GPD, problemas de mistura de distribuição de dados reais resultam em modelos com poucas componentes na mistura, o algoritmo RJMCMC resulta num alto custo computacional, para se alcançar a distribuição de equilíbrio. Aplicações deste trabalho mostraram que no máximo três componentes na mistura é o suficiente para alcançar o melhor modelo. Neste trabalho, foram adotados três critérios de comparação de modelos.

2.9.1 Distribuição preditiva

Um critério gráfico para avaliar o bom ajuste do modelo consiste na curva da função de densidade preditiva do modelo. A função distribuição preditiva é dada por

$$f_p(y_{n+1}|\mathbf{y}) = \int f_p(y_{n+1},\theta|\mathbf{y})d\theta = \int f_p(y_{n+1}|\theta)p(\theta|\mathbf{y})d\theta = E_{\theta|\mathbf{y}}\left(f_p(y_{n+1}|\theta)\right).$$

Amostrando I valores de θ da distribuição a posteriori e usando o fato de que, como $f_p(y_{n+1} \mid y) = E_{\theta|y}(f_p(y_{n+1}|\theta))$, pela técnica de Monte Carlo, que é um estimador razoável para ser utilizado, $\hat{f}_p(y_{n+1}|\mathbf{y}) = \frac{1}{I}\sum_{i=1}^{I} f_p(y_{n+1}|\theta^{(i)})$, onde $\theta^{(i)}$ é o i-ésimo valor amostrado da distribuição a posteriori.

Considerando que neste capítulo, f_p é uma mistura de k distribuições Gama com parâmetros $(\mu_j, \eta_j), j = 1, \ldots, k$, a função de densidade preditiva pode ser aproximada por

$$\hat{f}_p(y_{n+1}|\mathbf{y}) = \frac{1}{I}\sum_{i=1}^{I}\sum_{j=1}^{k} p_j^{(i)} f_G(y_{n+1}|\mu_j^{(i)}, \eta_j^{(i)}).$$

2.9.2 Critério de Informação Bayesiano (BIC)

Este critério foi proposto por Schwarz (1978), e é um dos primeiros e mais utilizados critérios de comparação de modelos. Este método penaliza o número de parâmetros de acordo com o tamanho da amostra.

O BIC é dado por

$$BIC = -2\frac{\sum_{i=1}^{I} \log\left((f_p(\mathbf{y} \mid \Theta_i)\right)}{I} + q\log(n), \tag{2.4}$$

onde $f_p(\mathbf{y} \mid \Theta_i)$ é a função de verossimilhança, Θ_i é o vetor de parâmetros Θ na iteração i, q é o número de parâmetros do modelo e n é o tamanho da amostra. O primeiro termo do BIC avalia o ajuste do modelo e o segundo termo é a penalização de acordo com o número de parâmetros. Comparando vários modelos, o melhor, segundo o BIC, é aquele que tiver o menor valor.

Quando é feita a estimação via MCMC no modelo com mistura de Gamas, onde $\Theta_i = (\mu_{i,1}, \ldots, \mu_{i,k}, \eta_{i,1}, \ldots, \eta_{i,k}, p_{i,1}, \ldots, p_{i,k})$, tem-se que

$$\log(f_p(\mathbf{y} \mid \Theta_i)) = \sum_{l=1}^{n} \log\left(\sum_{j=1}^{k} p_j^{(i)} f_G(y_l|\mu_j^{(i)}, \eta_j^{(i)})\right),$$

onde $\mu_{i,j}, \eta_{i,j}$ e $p_{i,j}$, $j = 1, \ldots, k$ e $i = 1$, são os valores dos parâmetros na componente j da mistura na i-ésima iteração do MCMC, num total de I iterações.

2.9.3 Critério de Informação dos Desvios (DIC)

O DIC, introduzido por Spiegelhalter et al. (2002), se baseia na função de distribuição a posteriori da estatística dos desvios.

Para encontrar o DIC, considere uma medida $D(\Theta \mid \mathbf{y})$, dada por:

$$D(\Theta \mid \mathbf{y}) = -2\log(f_p(\mathbf{y}|\Theta)),$$

onde $f_p(\mathbf{y}|\Theta)$ é a função de verossimilhança.

O DIC é calculado por

$$DIC = \bar{D}(\Theta \mid \mathbf{y}) + pD(\Theta \mid \mathbf{y}),$$

onde pD é conhecido como número efetivo de parâmetros, e avalia a complexidade do modelo, e pode ser calculado por

$$pD(\Theta \mid \mathbf{y}) = \bar{D}(\Theta \mid \mathbf{y}) - \hat{D}(\Theta \mid \mathbf{y}),$$

onde $\hat{D}(\Theta \mid \mathbf{y}) = -2\log(f(Y|\hat{\Theta}))$, com $\hat{\Theta}$ sendo uma estimativa da média a posteriori de Θ, ou seja, num caso de um vetor $\Theta = (\theta_1, \ldots, \theta_m)$,

$$\hat{D}(\Theta) = D\left(\frac{1}{I}\sum_{i=1}^{I}\theta_1^{(i)}, \ldots, \frac{1}{I}\sum_{i=1}^{I}\theta_m^{(i)}\right).$$

No caso da mistura de Gamas com k componentes,

$$\hat{D}(\Theta \mid \mathbf{y}) = -2\log\left(\sum_{j=1}^{k}\bar{p}_j f_G(y_i|\bar{\mu}_j, \bar{\eta}_j)\right),$$

onde $\bar{\mu}_j = \sum_{i=1}^{I}\mu_j^{(i)}/I$, $\bar{\eta}_j = \sum_{i=1}^{I}\eta_j^{(i)}/I$ e $\bar{p}_j = \sum_{i=1}^{I}p_j^{(i)}/I$, $j = 1, \ldots, k$ e

$$\bar{D}(\Theta \mid \mathbf{y}) = -2\sum_{i=1}^{I}\log\left(\sum_{j=1}^{k}p_{i,j}f_G(y_i|\mu_{i,j}, \eta_{i,j})\right)/I.$$

Embora este método tenha sido utilizado com muita frequência nos últimos anos, recomenda-se ter cuidado com algumas restrições, como por exemplo o número efetivo de parâmetros, que em alguns casos pode ser negativo. O DIC pode apresentar dificuldades

nos casos onde não há a garantia que a função de verossimilhança seja log-côncava, quando há mais de uma moda na distribuição dos parâmetros e em alguns casos onde há uma forte assimetria na distribuição dos parâmetros. Estes ítens citados estão diretamente relacionados com este estudo de distribuição de mistura.

Na discussão do artigo de Spiegelhalter et al.(2002), Richardson cita exemplo de DIC para mistura de duas distribuições Normais, e em um dos casos, fazendo uma simulação para $0,75N(0;1) + 0,25N(1,5;0,5)$, o menor DIC encontrado foi considerando 4 componentes na mistura e não 2, que seria o modelo correto.

Neste trabalho, foram calculadas as medidas DIC de cada simulação para averiguar mais detalhadamente as limitações do DIC em relação à mistura de distribuições.

2.10 Simulações

Para realizar comparações foram aplicados dois algoritmos: o primeiro considerando apenas o algoritmo de Gibbs com passos de Metropolis (Seção 2.7) e outro que utiliza a variável latente (Seção 2.8). Em ambos os casos, o algoritmo foi feito impondo a restrição na média $R = \{(\mu_1, ..., \mu_k) \in (0, \infty); \mu_1 < \ldots < \mu_k\}$. Foram feitas simulações considerando mistura com uma, duas e três componentes.

Para cada simulação, foi feita a estimação considerando diversos valores de k, com intuito de verificar se o modelo que obtém os melhores resultados é aquele que tem o mesmo número de componentes da mistura usada na simulação. Por facilidade de notação, será denotado por k_{sim} o número de componentes na mistura utilizados na simulação e por k_{est} o número de componentes utilizados na estimação.

Para fazer esta comparação, foram analisados os resultados da distribuição preditiva, verificando em qual caso a densidade preditiva é mais próxima da verdadeira densidade. Além disso, foram utilizados critérios de comparação de modelos, como BIC e DIC.

Foram realizadas 3 simulações de 1000 pontos da distribuição de mistura com k_{sim}=1, 2 e 3. Para cada caso, na estimação por MCMC, foram realizadas 50 mil iterações, com

burn-in de 10 mil e coletadas uma amostra a cada 10 observações. A verificação de convergência foi feita utilizando duas cadeias começando de valores iniciais diferentes.

Distribuição a priori: Para cada simulação, a distribuição a priori será como em (2.1), onde $\mu_j \sim IG(2,1;5,5)$ e $\eta_j \sim G(6;0,5)$, $j = 1,\ldots,k_{est}$, e $\mathbf{p} \sim \mathrm{D}_{k_{est}}(1,\ldots,1)$. Estes parâmetros foram escolhidos de modo a se obter uma média a priori próxima dos valores verdadeiros e com variância alta, numa situação onde a distribuição a priori fornece pouca informação sobre os parâmetros.

Foi usada a linguagem Ox versão 4 (veja Doornik, 1996) para desenvolver os algoritmos de MCMC em um computador Acer Intel Atom N270, 1,60GHz, 1GB RAM.

2.10.1 Simulação com uma Gama

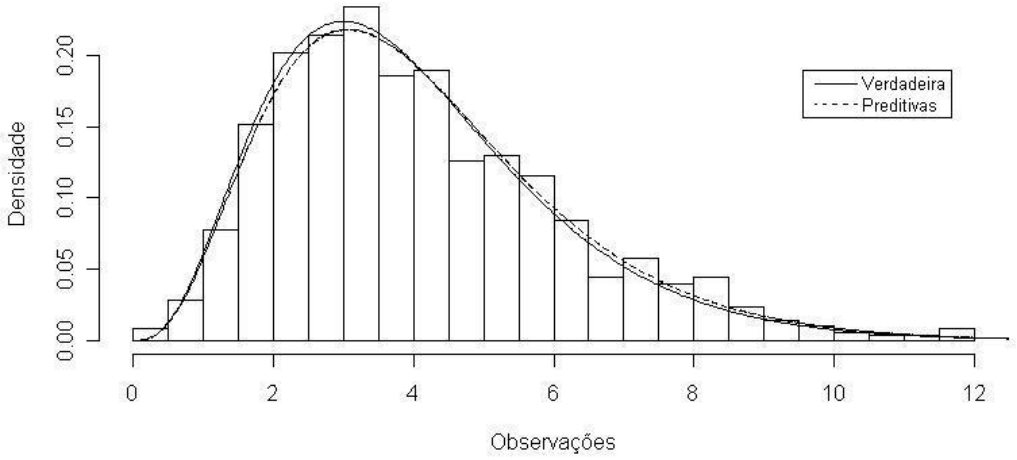
Neste primeiro exercício, foram simulados 1000 pontos de uma distribuição Gama com parâmetros $\mu = 4$ e $\eta = 4$. Supondo que não se conhece o número de componentes de mistura, foi feita a modelagem nestes dados considerando k_{est}=1, 2 e 3 para os algoritmos com e sem a variável latente.

A Figura 2.3 mostra as densidades preditivas para diversos valores de k_{est}, além da verdadeira densidade, juntamente com o histograma das simulações.

Pela Figura 2.3, nota-se que, independente do valor de k_{est}, a estimativa da densidade preditiva é praticamente a mesma, sendo que todas elas estão sobrepostas e muito próximas da verdadeira função de densidade. Isto mostra que, misturar distribuições ($k_{est} > 1$) irá ter como resultado um modelo tão eficiente quanto no modelo com $k_{est} = 1$.

A Figura 2.4 apresenta as taxas de aceitação baseado no método de calibração da variância a posteriori, do trabalho de Roberts e Rosenthal (2006). Pelo gráfico, verifica-se que o algoritmo que utiliza a variável latente chega mais rapidamente a taxa em torno de $0,44$ (gráfico à esquerda da Figura 2.4), enquanto que, ao utilizar o algoritmo de Gibbs com passos de Metropolis para todos os parâmetros, a taxa só fica em torno de $0,44$ após cerca de 10000 iterações, que é o valor utilizado para *burn-in*. Para estas simulações, utilizar a taxa de $0,44$ fornece bons resultados no comportamento das cadeias, não havendo a

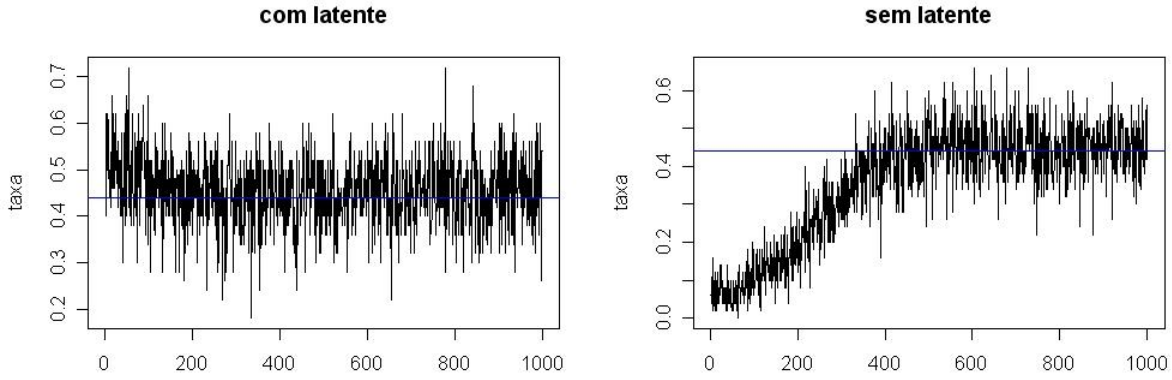
Figura 2.3: Densidade preditiva na simulação com $k_{sim} = 1$ e $k_{est} = 1, 2, 3$.



necessidade de diminuir a taxa.

Figura 2.4: Taxas de aceitação a cada 50 iterações para $k_{sim} = 1$, $k_{est} = 1$ para a estimação da média da distribuição Gama.

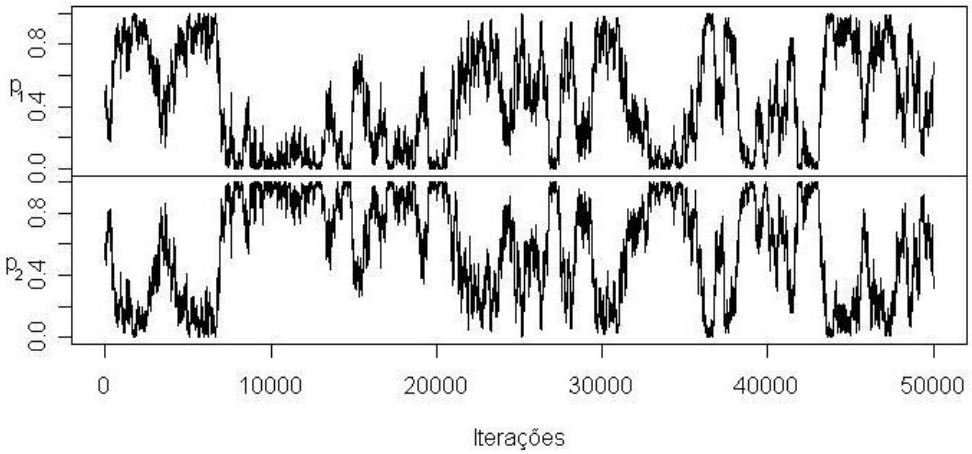
A linha vertical representa a taxa de 0,44.



Quando é utilizado o algoritmo com variáveis latentes, a Figura 2.5 mostra que há um problema de identificabilidade nos pesos. Quando $k_{est} > 1$, o peso de uma componente

troca com o peso da outra. Portanto, para este método, não houve convergência nas cadeias quando $k_{est} > k_{sim}$. Isto acaba afetando também na estimação dos parâmetros das componentes de mistura.

Figura 2.5: Série dos pesos utilizando a variável latente com $k_{sim} = 1$, $k_{est} = 2$.



Utilizando o algoritmo de Gibbs com passos de Metropolis, pela Figura 2.6, observa-se que a série dos pesos converge para uma distribuição em torno de 1 para a primeira componente e 0 para a segunda, ou seja, o algoritmo praticamente só usa a primeira componente para fazer a estimação, não reconhecendo a existência da segunda componente, pois a simulação é feita utilizando $k_{sim} = 1$.

A densidade preditiva de $k_{est} = 3$ apresentou praticamente o mesmo comportamento do que no caso onde $k_{est} = 2$, assim como a série dos pesos que o algoritmo de Gibbs com passos de Metropolis mostra que a terceira componente leva peso 1 e as outras duas 0.

Para comparar todos os resultados das simulações com $k_{sim} = 1$, foram calculados o BIC e o DIC para cada caso estudado. Como quando $k_{est} > k_{sim}$ o algoritmo com variável latente não atinge a distribuição de convergência, as medidas de ajuste que indicam o modelo com melhores componentes foi feito apenas para o algoritmo que não utiliza variável latente.

Pela Tabela 2.1, pode-se concluir que os valores do DIC foram praticamente os mesmos

Figura 2.6: Série dos pesos sem variável latente com $k_{sim} = 1$, $k_{est} = 2$.

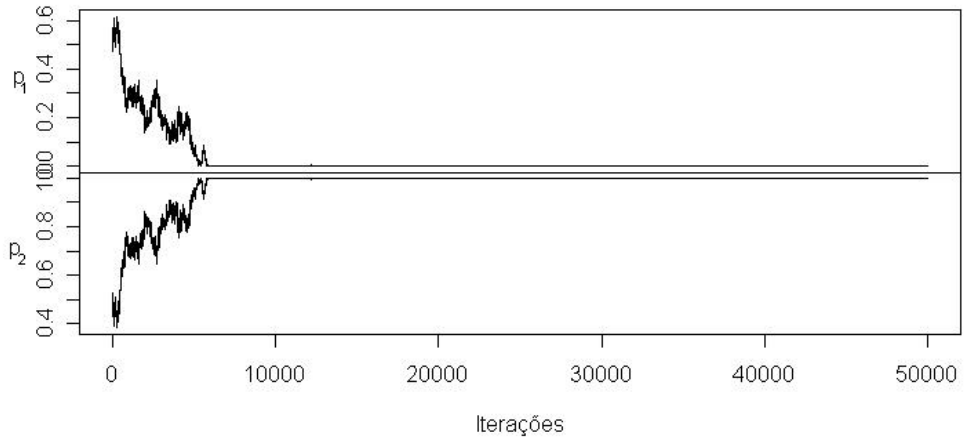


Tabela 2.1: Medidas de ajuste de mistura para $k_{sim} = 1$.

k_{est}	pD	DIC	BIC
1	2,0	**4002,7**	**4027,6**
2	1,9	4002,9	4048,3
3	2,0	4002,8	4069,1

para os diferentes números de componentes na mistura. Para o BIC, os menores valores das medidas foram para a estimação com uma componente na mistura.

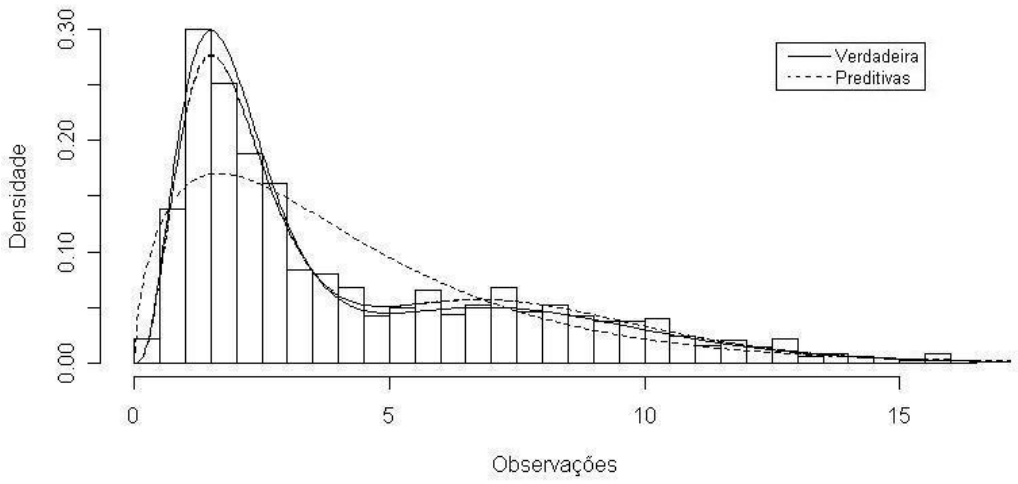
2.10.2 Simulação com mistura de duas Gamas

Na segunda simulação realizada, foram simulados 1000 pontos de mistura de duas Gamas, com parâmetros $\mu = (2; 8)$, $\eta = (4; 8)$ e $p = (0,66; 0,33)$. Foi feita a modelagem nestes dados considerando k_{est} de 1 a 4 para os algoritmos com e sem a variável latente.

A Figura 2.7 mostra a comparação entre as densidades verdadeira e preditivas para diferentes valores de k_{est}, com e sem a presença da variável latente. A densidade que está mais longe da verdadeira é para $k_{est} = 1$. Para $k_{est} > 1$, as funções de densidade preditiva

estão muito próximas da verdadeira densidade.

Figura 2.7: Densidade preditiva da simulação com $k_{sim} = 2$ e $k_{est} = 1, 2, 3, 4$.
A linha tracejada com $k_{est} = 1$ é a única que fica distante da verdadeira densidade. Para os outros valores de k_{est}, as densidades estão sobrepostas e próximas da verdadeira densidade.



No algoritmo que utiliza o amostrador de Gibbs com passos de Metropolis, quando $k_{est} \geq 2$, apenas dois pesos parecem ser significativos, ou seja, o modelo só identifica o número verdadeiro de componentes utilizados na simulação. Pode-se verificar na Figura 2.8 com $k_{est} = 4$, que a primeira e a terceira componentes vão a 0 e a segunda e quarta componentes estacionam numa distribuição em torno dos valores verdadeiros dos pesos de 0,66 e 0,33.

Para comparar todos os resultados das simulações com $k_{sim} = 2$, foram calculados o BIC e DIC. Estes resultados estão na Tabela 2.2.

Pela tabela, percebe-se que o DIC é maior quando $k_{est} = 1$ e há pouca diferença entre os outros valores de $k_{est} > 1$. Isto ocorre porque, para três ou quatro componentes, somente dois pesos são significativos e os outros acabam não tendo nenhuma relevância na estimação. Percebe-se também que, para $k_{est} \geq 2$, o número efetivo de parâmetros se altera pouco, o que indica que eles estão estimando praticamente o mesmo modelo.

Figura 2.8: Série dos pesos sem variável latente com $k_{sim} = 2$, $k_{est} = 4$.

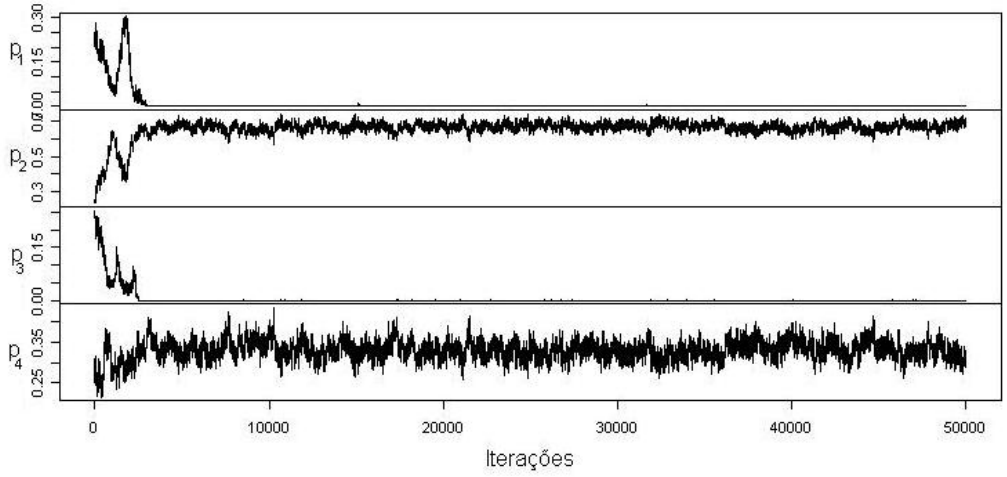


Tabela 2.2: Medidas de ajuste de mistura para $k_{sim} = 2$.

k_{est}	pD	DIC	BIC
1	2,1	4662,9	4688,1
2	5,0	4425,1	**4476,6**
3	5,2	4424,8	4497,3
4	5,3	**4424,5**	4518,0

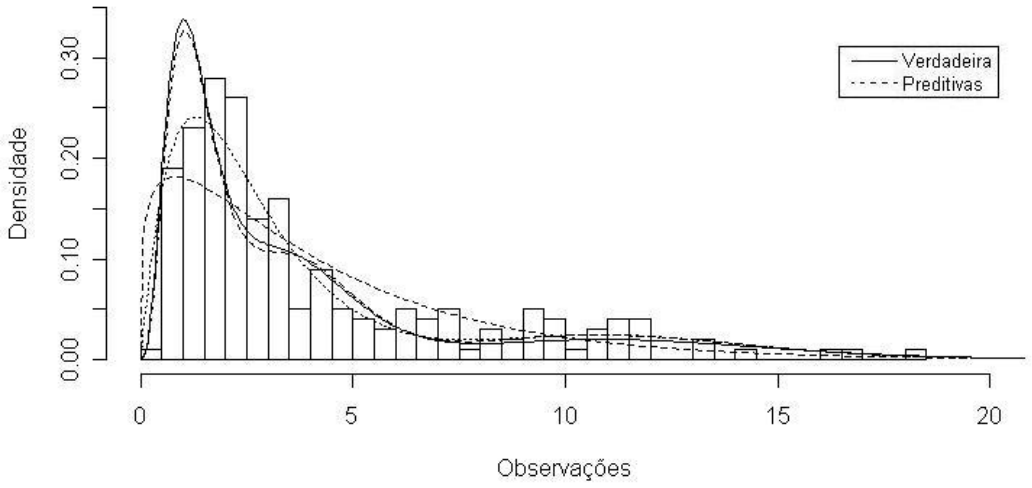
Para o BIC, o menor valor foi exatamente quando $k_{est} = 2$.

2.10.3 Simulação com mistura de três Gamas

Na terceira simulação, foram simulados 1000 pontos de mistura de três Gamas, com parâmetros $\mu = (1,33; 4; 12)$, $\eta = (4; 8; 12)$ e $p = (0,50; 0,333; 0,166)$. Supondo que não se conhece o número de componentes de mistura, foi feita a modelagem nestes dados considerando k_{est}=1, 2, 3, 4 e 5 para os algoritmos com e sem a variável latente. A Figura 2.9 mostra a comparação entre as densidades verdadeira e preditiva para os diferentes valores de k_{est}. Pela Figura, duas componentes na mistura não resultam num bom ajuste,

enquanto que para os outros valores de $k_{est} > 2$, as densidades preditivas são praticamente as mesmas e próximas à verdadeira densidade de mistura.

Figura 2.9: Densidade preditiva com simulação com $k_{sim} = 3$, $k_{est} = 1, 2, 3, 4$.
A linha tracejada com $k_{est} = 1$ é a mais distante da verdadeira densidade. A linha pontilhada representa $k_{est} = 2$ e fica mais próxima. Para os outros valores de k_{est}, as densidades estão sobrepostas e muito próximas da verdadeira densidade.



Foram utilizados o BIC e o DIC como medidas de ajuste para cada caso. Pela Tabela 2.3, o menor DIC foi para $k_{est} = 5$, enquanto que o menor BIC foi para $k_{est} = 3$. Portanto, para simulação de 3 componentes, a melhor medida de ajuste foi o BIC.

2.10.4 Conclusões

Baseado nos resultados encontrados nas simulações de mistura de distribuições, pode-se tirar as seguintes conclusões:

- para o algoritmo que utiliza variável latente, só foi possível realizar a estimação nos casos onde $k_{est} \leq k_{sim}$. Nas outras situações, pode-se notar que a série dos pesos não convergiu para uma distribuição estacionária;

Tabela 2.3: Medidas de ajuste de mistura para $k_{sim} = 3$.

k_{est}	pD	DIC	BIC
1	2,0	4840,3	4865,3
2	5,1	4519,5	4571,4
3	4,8	4475,8	**4544,8**
4	5,5	4473,8	4565,6
5	6,9	**4468,7**	4586,2

- para o algoritmo que não utiliza a variável latente, para os modelos onde $k_{est} \geq k_{sim}$, não houveram diferenças na densidade preditiva, pois o número efetivo de pesos significantes na estimação foi de k_{sim} mesmo quando $k_{est} > k_{sim}$;

- a medida BIC mostrou-se mais eficiente para comparar os modelos, pois em todos os casos o modelo estimado com menor BIC foi o modelo simulado. Em relação ao DIC, é necessário ter um pouco mais de cautela, pois em alguns casos, quando $k_{est} \geq k_{sim}$, há pouca diferença entre os valores do DIC e do número efetivo de parâmetros. Isto pode indicar que nestes casos, diferentes valores de k_{est} podem estar levando à mesma estimação, pois alguns pesos tendem a 0 para modelos com $k_{est} > k_{sim}$.

- O algoritmo que não utiliza a variável latente mostrou-se mais eficiente na identificação do número de componentes na mistura, pois independente do número de componentes estimados na mistura, a estimação converge para os parâmetros do modelo verdadeiro, realizando uma estimação correta não só da curva de densidade, mas também dos parâmetros. Além disso, esta técnica é mais rápida computacionalmente do que a técnica que utiliza a variável latente.

Apêndice 2A - Algoritmo para mistura de Gamas

O algoritmo para amostrar pontos de uma distribuição de mistura de Gamas sem utilizar a variável latente é dado a seguir:

Algoritmo 1

A partir da iteração s do MCMC, os parâmetros são amostrados por:

a) Na componente j da distribuição de mistura, a amostragem é feita da seguinte maneira.

Amostrando μ_j e η_j

Como η_j é um valor que só pode ser positivo, η_j^* foi adotada como distribuição proposta para este parâmetro uma Gama dada por

$$\eta_j^* | \eta_j^{(s)} \sim G(\eta_j^{(s)}, \eta_j^{(s)^2}/V_{\eta_j}),$$

onde $\eta_j^{(s)}$ é o valor de η_j na iteração s da cadeia e V_{η_j} é a variância da distribuição proposta. Note que, de acordo com essa proposta, $E(\eta_j^* \mid \eta_j^{(s)}) = \eta_j^{(s)}$, e $Var(\eta_j^* \mid \eta_j^{(s)}) = V_{\eta_j}$, $j = 1, \ldots, k$.

Como μ é um valor que só pode ser positivo, foi adotada como distribuição proposta para este parâmetro uma Gama truncada na restrição das médias, denotada por

$$\mu_j^* | \mu_j^{(s)} \sim G(\mu_j^{(s)}, \mu_j^{(s)^2}/V_{\mu_j}) I_R(\underline{\mu}),$$

onde $\mu_j^{(s)}$ é o valor de μ_j na iteração s da cadeia e V_{μ_j} é a variância da distribuição proposta.

Os valores $\eta_j^{(s+1)} = \eta_j^*$ e $\mu_j^{(s+1)} = \mu_j^*$ são aceitos com probabilidade α_{μ_j,η_j}, onde

$$\alpha_{\mu_j,\eta_j} = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x}) f_G(\mu_j^{(s)}|\mu_j^*, \mu_j^{*2}/V_{\mu_j}) f_G(\eta_j^{(s)}|\eta_j^*, \eta_j^{*2}/V_{\eta_j}) I_R(\underline{\mu})}{\pi(\tilde{\Theta}|\mathbf{x}) f_G(\mu_j^*|\mu_j^{(s)}, \mu_j^{(s)2}/V_{\mu_j}) f_G(\eta_j^*|\eta_j^{(s)}, \eta_j^{(s)2}/V_{\eta_j}) I_R(\underline{\mu})}\right\},$$

$\Theta^* = (\eta_{<j}^{(s+1)}, \eta_j^*, \eta_{>j}^{(s)}, \mu_{<j}^{(s+1)}, \mu_j^*, \mu_{>j}^{(s)}, p^{(s)})$ e $\tilde{\Theta} = (\eta_{<j}^{(s+1)}, \eta_{\geq j}^{(s)}, \mu_{<j}^{(s+1)}, \mu_{\geq j}^{(s)}, p^{(s)})$.

b) Amostrando $\mathbf{p}$.

Como p é um vetor de dimensão k com $\sum_{j=1}^k p_j = 1$, $\mathbf{p}^*$ foi amostrado de uma distribuição proposta Dirichlet, com parâmetros $(V_p p_1^{(s)}, \ldots, V_p p_k^{(s)})$, onde V_p é um valor constante que determina a variância da distribuição proposta. Resultados de simulações

mostraram que um valor adequado é de $V_p = 50$. Assim, $\mathbf{p}^{(s+1)} = \mathbf{p}^*$ com probabilidade

$$\alpha_p = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x}) f_D(\mathbf{p}^{(s)}|\mathbf{p}^*)}{\pi(\tilde{\Theta}|\mathbf{x}) f_D(\mathbf{p}^*|\mathbf{p}^{(s)})}\right\},$$

onde $\Theta^* = (\eta^{(s+1)}, \mu^{(s+1)}, \mathbf{p}^*)$ e $\tilde{\Theta} = (\eta^{(s+1)}, \mu^{(s+1)}, \mathbf{p}^{(s)})$.

Para a escolha da variância das distribuições propostas V_{η_j} e V_{μ_j}, $j = 1, \ldots, k$ poderia-se a princípio escolher um valor fixo, que poderia resultar numa taxa de aceitação muito alta ou baixa no algoritmo de Metropolis. Porém, taxas altas resultam em problemas de convergência da cadeia e escolha de valores impróprios para a distribuição, enquanto que taxas muito baixas resultam em poucas amostras para a estimação dos parâmetros da distribuição a posteriori.

No Algoritmo 1, o método de sintonização de Roberts e Rosenthal (2006) é feito para V_{μ_j} e V_{η_j}, $j = 1, \ldots, l$. Em simulações, este algoritmo não pareceu ser eficiente para sintonizar V_p, que foi dado por um valor fixo.

Neste trabalho, como as propostas não são unidimensionais, em algumas situações foi necessário diminuir a taxa de aceitação calibrada.

Apêndice 2B - Condicionais completas no caso com variável latente

Considerando a modelagem de mistura de Gamas, com a variável latente, encontra-se as seguintes condicionais completas, baseado na distribuição a posteriori em (2.3):

- Observando a equação (2.3), percebe-se que, se $(\mu_1 < \ldots < \mu_k)$, a distribuição condicional completa de μ_j, $j = 1, \ldots, k$ é dada por

$$\pi(\mu_j|\mu_{-j}, \eta, p, \mathbf{z}) \propto \mu_j^{-c_j - \eta_j \sum_{i=1}^n z_{i,j} - 1} \exp\left(-\left(d_j + \eta_j \sum_{i=1}^n z_{i,j} x_i\right)/\mu_j\right) I_R(\underline{\mu}),$$

onde $\mu_{-j} = (\mu_1, \ldots, \mu_{j-1}, \mu_{j+1}, \ldots, \mu_k)$.

Assim, se $(\mu_1 < \ldots < \mu_k)$,

$$\mu_j|\mu_{-j}, \eta, \mathbf{p}, \mathbf{z} \sim IG\left(c_j + \eta_j \sum_{i=1}^{n} z_{i,j}, d_j + \eta_j \sum_{i=1}^{n} z_{i,j} x_i\right) I_R(\underline{\mu}), \quad j = 1, \ldots, k.$$

- Pela equação (2.3), tem-se que a distribuição condicional completa em $\mathbf{p}$ é dada por

$$\pi(\mathbf{p}|\mu, \nu, z) \propto \prod_{j=1}^{k} p_j^{\left(\sum_{i=1}^{n} z_{i,j} + \gamma_j - 1\right)},$$

e portanto,

$$\mathbf{p} \mid \mu, \eta, \mathbf{z} \sim D_k\left(\left(\sum_{i=1}^{n} z_{i,1} + \gamma_1\right), \ldots, \left(\sum_{i=1}^{n} z_{i,k} + \gamma_k\right)\right).$$

- De acordo com a equação (2.3), a distribuição condicional completa em $\mathbf{z}_i$ é dada por

$$\pi(\mathbf{z}_i|\mu, \nu, \mathbf{p}, \mathbf{z}_{-i}) \propto \prod_{j=1}^{k} (p_j f_G(x_i|\mu_j, \eta_j))^{z_{i,j}},$$

onde $\mathbf{z}_{-i} = (\mathbf{z}_1, \ldots, \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, \ldots, \mathbf{z}_n)$, $i = 1, \ldots, n$.

Assim,

$$\mathbf{z}_i|\mu, \eta, \mathbf{p}, \mathbf{z}_{-i} \sim MN_k\left(1; \frac{p_1 f(x_i|\mu_1, \eta_1)}{C}, \ldots, \frac{p_k f(x_i|\mu_k, \eta_k)}{C}\right), \tag{2.5}$$

onde $C = \sum_{j=1}^{k} p_j f\left(x_i|\mu_j, \eta_j\right)$.

- Condicional em η_j

Para amostrar η_j, $j = 1, \ldots, k$, por (2.3), como não é possível encontrar uma distribuição condicional completa conhecida, é utilizado o mesmo procedimento do Algoritmo 1 (Gibbs com passos de Metropolis), também utilizando como proposta uma distribuição Gama.

Capítulo 3

Mistura de distribuições e TVE

Neste capítulo, será apresentado um novo modelo para dados extremos. Neste modelo, a cauda da observação, ou seja, valores maiores que um determinado limiar, segue uma distribuição GPD. A distribuição abaixo do limiar é estimada utilizando uma aproximação não-paramétrica, por uma mistura finita de distribuições Gama.

3.1 Teoria de valores extremos

Análise de dados extremos tem sido uma grande ferramenta nas mais diversas áreas do conhecimento nas últimas décadas, ajudando a prever grandes ganhos e perdas. Duas áreas onde esta análise se destaca são nas áreas ambientais e econômicas. Em estudos de temperatura, o aquecimento global tem sido um assunto em destaque nos últimos anos devido a rápidas alterações climáticas que o planeta vem sofrendo nas últimas décadas. Estas mudanças implicam em alterações no número de eventos extremos de temperatura, seja esta de máximos ou mínimos. Eventos de temperaturas extremas são mais responsáveis por mudanças na natureza do que mudanças na média (Parmesan et al., 2000).

Verões extremamente quentes ou invernos extremamente frios podem influenciar na agricultura, no consumo de energia, e em problemas de saúde para a população. Entender com qual frequência eventos extremos ocorrem em um período de tempo é importante

para amenizar o impacto na sociedade das perdas e danos que estes eventos possam vir a acarretar.

A TVE pode ajudar a resolver estes problemas, através de modelos que ajudam a entender o comportamento de dados extremos, calculando a probabilidade de quantis altos e prevenindo a população contra possíveis perdas.

Seja $x_1, \ldots, x_n$ variáveis aleatórias iid, com função distribuição F. $M_n = \max\{x_1, \ldots, x_n\}$ é o valor máximo destas observações. O resultado clássico da TVE mostra três possíveis distribuições limite para M_n, que é enunciado pelo Teorema de Fisher-Tippet (1928).

Teorema 1 *Se existem sequências de constantes $a_n > 0$ e $b_n \in \Re$, e alguma função distribuição H não-degenerada tal que*

$$P\{(M_n - b_n)/a_n \leq x\} \to H(x),$$

então H é do mesmo tipo de uma entre 3 distribuições:

$$\begin{aligned}
Gumbel \quad &: \quad H_I(x) = \exp\{\exp(-x)\}, \quad x \in \Re \\
Fréchet \quad &: \quad H_{II}(x) = \begin{cases} 0, & x \leq 0, \xi > 0 \\ \exp\left(-x^{-\xi}\right), & x > 0, \xi > 0 \end{cases} \\
Weibull \quad &: \quad H_{III}(x) = \begin{cases} \exp\left\{-\left(-x^{-\xi}\right)\right\}, & x \leq 0, \xi < 0, \\ 1, & x > 0, \xi < 0 \end{cases}
\end{aligned}$$

onde duas distribuições F e F^* são do mesmo tipo se existem constantes a e b tal que $F^*(a + bx) = F(x)$, para todo x.

O teorema mostra distribuições limite para máximos coletados em blocos de tamanho n. Estes blocos, dependendo do tipo de dado em estudo podem ser, por exemplo, meses do ano, coletando os máximos mensais de dados diários.

Ao invés de trabalhar com três distribuições distintas, von Mises (1954) e Jenkinson (1955) propuseram a distribuição GEV, que engloba as três distribuições do Teorema

de Fisher-Tippet. Esta função é denotada por $H_{u,\sigma,\xi}$ e possui a seguinte função de distribuição:

$$H(y|\xi,\sigma,\mu) = \begin{cases} \exp\left\{-\left(1+\xi\left(\frac{(y-\mu)}{\sigma}\right)\right)^{-1/\xi}\right\}, & \text{se} \quad \xi \neq 0 \\ \exp\left\{-\exp\left\{-\left(\frac{y-\mu}{\sigma}\right)\right\}\right\}, & \text{se} \quad \xi = 0 \end{cases}, \tag{3.1}$$

definida em $\{y : 1-\xi(y-\mu)/\sigma > 0\}$.

O conjunto de todas as distribuições F de uma mesma classe para as quais os máximos normalizados tenham a mesma distribuição limite é chamado de *domínio de atração* dessa distribuição limite. A função distribuição F pertence ao domínio de atração de uma função distribuição G se existem constantes $a_n > 0$ e $b_n \in \Re$ tal que $a_n^{-1}(M_n - b_n) \xrightarrow{d} G$.

3.1.1 Distribuição de Pareto generalizada

A distribuição de Pareto generalizada foi desenvolvida por Pickands (1975), e se baseia no seguinte teorema:

Teorema 2 *Se X é uma variável aleatória com função distribuição $F(x)$, que pertence ao domínio de atração de uma distribuição GEV, então, quando $u \to \infty$, $F(x|u) = P(X \leq u + x|X > u)$, possui distribuição GPD, possui a seguinte função de distribuição*

$$G(x|\xi,\sigma,u) = \begin{cases} 1-\left(1+\xi\frac{(x-u)}{\sigma}\right)^{-1/\xi}, & se \quad \xi \neq 0 \\ 1-\exp\{-(x-u)/\sigma\}, & se \quad \xi = 0 \end{cases}, \tag{3.2}$$

onde $u > 0$, $\sigma > 0$, $x - u \geq 0$, se $\xi \geq 0$, e $0 \leq x - u \leq -\sigma/\xi$, se $\xi < 0$. O caso $\xi = 0$ é interpretado como sendo o limite quando $\xi \to 0$, é a distribuição exponencial de parâmetro 1. Os parâmetros são ξ, σ e u e representam a forma, escala e limiar da distribuição.

A função de densidade da distribuição GPD é dada por

$$g(x|\xi,\sigma,u) = \begin{cases} \frac{1}{\sigma}\left(1+\xi\frac{(x-u)}{\sigma}\right)^{-(1+\xi)/\xi}, & \text{se} \quad \xi \neq 0 \\ \frac{1}{\sigma}\exp\{-(x-u)/\sigma\}, & \text{se} \quad \xi = 0 \end{cases}, \tag{3.3}$$

onde $x - u > 0$ para $\xi \geq 0$ e $0 \leq x - u < -\sigma/\xi$ para $\xi < 0$.

Pickands (1975) e Davidson e Smith (1990) mostram propriedades que justificam o uso da GPD e provam também que não há nenhuma outra família de distribuições que satisfaça estas propriedades. Por exemplo, estabilidade do limiar, ou seja, se Y possui distribuição GPD, e se $u > 0$, então a distribuição de $(Y - u|Y > u)$ também possui distribuição GPD.

Em valores extremos, além de encontrar a estimativa dos parâmetros do modelo, também é muito importante encontrar uma forma para determinar os quantis altos, acima do limiar, de tal forma que se X possui distribuição GPD, é importante saber com qual probabilidade ocorre um evento maior ou igual a q, ou seja, $P(X > q) = 1 - p$.

Na distribuição GPD, um quantil q com probabilidade $P(X < q) = p$ é dado em função dos parâmetros, invertendo a função de distribuição acumulada em (3.2).

Assim, para encontrar o quantil q, basta inverter a função $p = G(q \mid \xi, \sigma, u)$ e obtém-se a seguinte forma para os quantis

$$q = \frac{((1-p)^{-\xi} - 1)\sigma}{\xi}.$$

Tão importante quanto realizar a estimação nos parâmetros, é obter as estimações de quantis com p próximos de 1, sendo de fundamental importância na análise de eventos extremos, para demonstrar as vantagens de se utilizar a função GPD para a cauda de dados extremos.

3.2 Mistura de Gamas com GPD

Baseado na TVE e na distribuição de mistura, este capítulo se baseia em uma extensão do trabalho de Behrens et al. (2004). Neste trabalho, utiliza-se uma aproximação não-paramétrica para a distribuição dos dados abaixo do limiar através de mistura de distribuições Gama e GPD acima do limiar.

3.2.1 Mistura no domínio de atração da GEV

Para utilizar a distribuição GPD para excessos de um modelo de mistura finita de distribuições Gama, é necessário mostrar que esta distribuição de mistura pertence ao domínio de atração da distribuição GEV para satisfazer o Teorema de Pickands (1975). Esta seção mostra como isto é satisfeito.

Segundo Embrechts et al. (1997), se F é uma variável aleatória com limite superior $x_F \leq \infty$, então F é uma função von Mises se, e somente se,

$$\lim_{x \uparrow x_F} \frac{\overline{F}(x) f'(x)}{f(x)^2} = -1, \tag{3.4}$$

onde $x \uparrow x_F$ é o limite quando x vai a x_F pela esquerda, $f'(x)$ é a derivada de f em relação a x, e $\overline{F}(x) = 1 - F(x)$.

Proposição 1 *Se F é uma função von Mises, então F pertence ao domínio de atração de uma distribuição Gumbel.*

Esta proposição é utilizada para mostrar que a distribuição Gama é uma função von Mises, pois, na função Gama, a primeira derivada é dada por $f'(x) = f(x)\left[\frac{\alpha-1}{x} - \beta\right]$

Assim, a Equação (3.4) para a Gama é dada por

$$\lim_{x\to\infty} \frac{(1-F(x)) f(x)\left[\frac{\alpha-1}{x} - \beta\right]}{f(x)^2} = \lim_{x\to\infty} \frac{(1-F(x))\left[\frac{\alpha-1}{x} - \beta\right]}{f(x)}.$$

Na equação acima, o limite do numerador e do denominador vão a 0. Assim, para encontrar o limite, aplica-se a regra de L'Hôpital, obtendo o seguinte limite

$$\begin{aligned} & \lim_{x\to\infty} \frac{-f(x)\left[\frac{\alpha-1}{x} - \beta\right] + (1-F(x))\left[-\frac{\alpha-1}{x^2}\right]}{\left[\frac{\alpha-1}{x} - \beta\right] f(x)} \\ = \; & -1 + \lim_{x\to\infty} \frac{(1-F(x))\left[-\frac{\alpha-1}{x^2}\right]}{f(x)(((\alpha-1) - \beta x)/x)} \\ = \; & -1 + \lim_{x\to\infty} \frac{(1-F(x))\left[\alpha-1\right]}{f(x)((\alpha-1)x - \beta x^2)}. \end{aligned}$$

Aplicando a regra de L'Hôpital novamente, tem-se

$$= \ -1 + \lim_{x\to\infty} \frac{-f(x)[\alpha-1]}{f(x)\left[(\alpha-1-2\beta x) + \left[\frac{\alpha-1}{x} - \beta\right]((\alpha-1)x - \beta x^2)\right]} = -1.$$

Portanto, a distribuição Gama é uma função von Mises e pertence ao domínio de atração de uma distribuição Gumbel. Consequentemente ela pertence também ao domínio de atração da distribuição GEV.

Para mostrar que a mistura finita de Gamas pertence ao domínio de atração da GEV, considere agora a seguinte proposição:

Proposição 2 *F e G são equivalentes de cauda se* $x_F = x_G$*, e*

$$\lim_{x\uparrow x_F} \frac{\overline{F}(x)}{\overline{G}(x)} = c. \tag{3.5}$$

Se F e G são equivalentes, então pertencem ao mesmo domínio de atração.

Seja $F = \sum_{j=1}^{k} p_j f_G(x \mid \alpha_j, \beta_j)$ e $G = f_G(x \mid \alpha, \beta)$. Assim $x_F = x_G = \infty$. Considere $\beta = \min_j \beta_j$ e $\alpha = \alpha_l$ para um l tal que $\beta_l = \beta$, então

$$\lim_{x\to\infty} \frac{\overline{F}(x)}{\overline{G}(x)} = \lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{\sum_{j=1}^{k} p_j f_G(x \mid \alpha_j, \beta_j)}{f_G(x \mid \alpha, \beta)} = \sum_{j=1}^{k} \lim_{x\to\infty} p_j \frac{f_G(x \mid \alpha_j, \beta_j)}{f_G(x \mid \alpha, \beta)}.$$

A razão entre duas densidades Gama é dada por

$$\frac{f_G(x \mid \alpha_j, \beta_j)}{f_G(x \mid \alpha, \beta)} = \frac{\Gamma(\alpha)\beta_j^{\alpha_j} x^{\alpha_j - 1} e^{-\beta_j x}}{\Gamma(\alpha_j)\beta^{\alpha} x^{\alpha-1} e^{-\beta x}} = c x^{\alpha_j - \alpha} e^{-(\beta_j - \beta)x}. \tag{3.6}$$

O limite da razão em (3.6) vai para 0 se $\beta_j > \beta$. Assim,

$$\sum_{j=1}^{k} \lim_{x\to\infty} p_j \frac{f_G(x \mid \alpha_j, \beta_j)}{f_G(x \mid \alpha, \beta)} = \sum_{j=1}^{k} p_j \lim_{x\to\infty} \frac{f(x \mid \alpha_j, \beta_j)}{f(x \mid \alpha, \beta)} = p_k,$$

onde k é a componente na mistura onde $\beta_k = \min_j \beta_j$

Assim, a densidade da mistura de Gamas é equivalente de cauda a uma única Gama, que pertence ao domínio de atração da distribuição Gumbel. Deste modo, a mistura finita de distribuições Gama pertence ao domínio de atração da distribuição GEV, sendo possível analisar os valores maiores que um determinado limiar a partir da distribuição GPD, pois satisfaz o Teorema proposto por Pickands (1975).

3.2.2 Modelo

Seja g a função densidade da distribuição GPD como em (3.2). Portanto, a função densidade do modelo é de mistura de Gamas com cauda GPD é dado por

$$f(x|\theta, \mathbf{p}, \Psi) = \begin{cases} \sum_{j=1}^{k} p_j f_G(x|\theta_j), & \text{se} \quad x \leq u \\ \left[1 - \sum_{j=1}^{k} p_j F_G(u|\theta_j)\right] g(x|\Psi), & \text{se} \quad x > u \end{cases} \tag{3.7}$$

onde $F_G(. \mid \theta)$ é a função de distribuição da Gama, $\Psi = (\xi, \sigma, u)$, $\mathbf{p} = (p_1, \ldots, p_k)$, $\theta = (\theta_1, \ldots, \theta_k)$, $\theta_j = (\mu_j, \eta_j)$, $j = 1, \ldots, k$, $\sigma > 0, (x - u) \leq -\sigma/\xi$ se $\xi < 0$ e $x > u$. Os dados exibem um comportamento de cauda pesada se $\xi > 0$.

Neste modelo, os parâmetros são $\Theta = (\theta, \mathbf{p}, \Psi)$ e o modelo de mistura de Gamas com GPD será denotado por $MGPD_k(\theta, \mathbf{p}, \xi, \sigma, u)$, ou simplesmente $MGPD_k$, quando não houver a necessidade de explicitar os parâmetros.

Uma primeira vantagem do modelo $MGPD_k$ em relação ao método totalmente não-paramétrico baseado no modelo MG_k, é tornar possível encontrar diretamente um quantil alto da distribuição. No modelo de distribuição de mistura, um quantil q tal que $P(X < q) = p$ é dado a partir da inversão da função de distribuição acumulada em q, dada por

$$p = H(q \mid \mu, \eta, \mathbf{p}) = \sum_{j=1}^{k} p_j F_G(q \mid \mu_j, \eta_j). \tag{3.8}$$

Porém, não é possível encontrar (3.8) em função de q diretamente, sendo necessário fazer uma grade de valores possíveis de q e observando qual deles apresenta uma probabilidade mais próxima ao p de (3.8).

No modelo $MGPD_k$, para estimar um quantil maior que o limiar, encontra-se q em função dos parâmetros do modelo, pois, para $x > u$, baseado na função distribuição do modelo $MGPD_k$,

$$F(x \mid \mu, \eta, \mathbf{p}, \Psi) = H(u \mid \mu, \eta, \mathbf{p}) + [1 - H(u \mid \mu, \eta, \mathbf{p})]\, G(x|\Psi),$$

é possível isolar o quantil q em função dos parâmetros e de p, obtendo a seguinte forma

$$q = \frac{((1 - p*)^{-\xi} - 1)\sigma}{\xi}, \tag{3.9}$$

onde

$$p* = \frac{p - H(u \mid \mu, \eta, \mathbf{p})}{1 - H(u \mid \mu, \eta, \mathbf{p})}.$$

Assim, além da facilidade dos cálculos, pelo fato da distribuição GPD ser na teoria a distribuição verdadeira para cauda de dados extremos, espera-se que os quantis estimados do modelo $MGPD_k$ sejam mais próximos dos quantis verdadeiros que os do modelo MG_k.

3.2.3 Distribuição a priori

Para realizar a abordagem Bayesiana e encontrar a distribuição a posteriori dos parâmetros, é necessário elicitar quais são as distribuições a priori para cada parâmetro. Para os parâmetros dos valores abaixo da cauda $(\Theta, \mathbf{p})$, foram escolhidas distribuições a priori da mesma família que a do Capítulo 2, ou seja, Gama para η_j, Gama Inversa para μ_j e Dirichlet para o vetor de pesos $\mathbf{p}$, com restrição de identificabilidade para o vetor $\mu_j's$.

Para o limiar, foi tomada uma priori Normal, mas sem utilizar a condição de truncamento como em Behrens et al.(2004). Porém, não é possível utilizar uma priori completamente vaga para este parâmetro. Os resultados de simulações e nas aplicações realizadas neste trabalho irão mostrar as dificuldades na estimação do limiar ao considerar uma distribuição a priori com uma variância relativamente alta.

Para os parâmetros da cauda, Coles e Tawn (1996) sugerem expressar a priori em função de quantis conhecidos da distribuição dos dados. Seguindo outro caminho, Castellanos e Cabras (2007) utilizaram uma priori de Jeffreys para estes parâmetros, prescindindo de suposições iniciais. A priori de Jeffreys para $(\xi, \sigma \mid u)$ é dada por

$$\pi(\sigma, \xi \mid u) \propto \sigma^{-1}(1+\xi)^{-1}(1+2\xi)^{-1/2}, \quad \xi > -0,5, \sigma > 0. \tag{3.10}$$

Castellanos e Cabras (2007) mostraram que, para o caso da distribuição GPD, a priori de Jeffreys irá resultar numa distribuição a posteriori que é própria.

3.2.4 Distribuição a posteriori

A função de verossimilhança do modelo $MGPD_k$ é dada por

$$L(\theta, \mathbf{p}, u, \sigma, \xi|\mathbf{x}) = \prod_{i:x_i<u} \left(\sum_{j=1}^{k} p_j f_G(x_i|\theta_j) \right) \prod_{i:x_i>u} \left(\left[1 - \sum_{j=1}^{k} p_j F_G(u|\theta_j) \right] g(x_i|\xi, \sigma, u) \right). \quad (3.11)$$

Dadas a distribuição a priori e a função de verossimilhança, encontra-se a seguinte forma para o logaritmo da função de densidade da distribuição a posteriori

$$\begin{aligned} \log \pi(\theta, \mathbf{p}, \Psi|\mathbf{x}) &= k + \sum_{i:x_i \geq u} \left[\log \left(\sum_{j=1}^{k} p_j f_G(x_i|\mu_j, \eta_j) \right) \right] + \sum_{i:x_i \geq u} \log \left[1 - \sum_{j=1}^{k} p_j F_G(u|\mu_j, \eta_j) \right] \\ &- \sum_{i:x_i \geq u} \left[\log(\sigma) - \frac{1+\xi}{\xi} \log \left(1 + \frac{\xi(x_i - u)}{\sigma} \right) \right] \\ &+ \sum_{j=1}^{k} \left[(a_j - 1) \log(\eta) - b_j \eta_j - (c_j + 1) \log(\mu) - d_j/\mu_j \right] \\ &- \frac{1}{2} \left(\frac{u - \mu_u}{\sigma_u} \right)^2 - \log(\sigma) - \log(1+\xi) - (1/2) \log(1+2\xi). \end{aligned} \quad (3.12)$$

Não existe forma fechada da distribuição a posteriori, e nenhum parâmetro possui distribuição condicional completa conhecida. Assim, todos os parâmetros do modelo terão que ser estimados pelo algoritmo de Gibbs por passos de Metropolis. Ao contrário de quando há uma simples mistura de distribuições Gama, no caso da mistura do GPD com limiar desconhecido, não é possível aplicar variáveis latentes que sejam vantajosas para a implementação do algoritmo, como na Seção 2.2, pois mesmo inserindo estas variáveis, não é possível encontrar uma forma conhecida para a distribuição a posteriori de cada parâmetro.

Além disso, também foi imposta a restrição na média das distribuições das componentes de mistura, como no Capítulo 2. O algoritmo para o modelo deste capítulo é dado no Apêndice 3.

3.3 Caracterização de excessos via processos pontuais

Existem diferentes caminhos para caracterizar o comportamento de valores extremos de um processo, e uma maneira particular de fazer esta caracterização é derivada da teoria de processos pontuais. Todas as inferências feitas usando a metodologia de processos pontuais podem ser igualmente obtidas utilizando o modelo proposto. A principal razão para considerar esta abordagem é que ela fornece uma interpretação de valores extremos que unifica todos os modelos introduzidos anteriormente.

3.3.1 Teoria básica de processos pontuais

Um processo pontual em um conjunto $\mathcal{A}$ é uma regra estocástica para a ocorrência e posição de pontos, onde $\mathcal{A}$ representa um período de tempo, que pode ser utilizado para descrever a ocorrência de furacões ou terremotos, por exemplo. A probabilidade de um certo número de eventos em um período de tempo pode ser calculada, além do tempo de espera entre dois eventos. Um conjunto $\mathcal{A}$ também pode ser multidimensional. Por exemplo, um processo pontual bi-dimensional pode ser usado para descrever a posição de trincos em um prato de vidro.

Um caminho para caracterizar as propriedades estatísticas de um processo pontual é definir um conjunto de variáveis aleatórias inteiras não-negativas, $N(A)$, tal que $N(A)$ é o número de pontos em um conjunto A, onde $A \subset \mathcal{A}$. A distribuição de probabilidade de cada $N(A)$ determina as características do processo pontual, que pode ser chamado de N. Algumas características importantes podem ser definidas. Em particular, $\Lambda(A) = E(N(A))$ é o número esperado de pontos no conjunto A, e é definido pela medida de intensidade do processo. Se $A = [a_1, x_1] \times \ldots \times [a_k, x_k] \subset R^k$, a função derivada

$$\lambda(x) = \frac{d\Lambda(A)}{dx_1 \ldots dx_k}$$

é a função intensidade do processo. O processo pontual canônico é o processo de Poisson

homogêneo unidimensional. Com $\lambda > 0$, o processo em $\mathcal{A} \subset R$ satisfaz

1. para todo $A = [t_1, t_2] \subset \mathcal{A}$, $N(A) \sim Poisson(\lambda(t_2 - t_1))$
2. para todos os conjuntos disjuntos A e B de $\mathcal{A}$, $N(A)$ e $N(B)$ são independentes.

O processo de Poisson é o modelo estocástico apropriado para pontos que ocorrem sem nenhuma regra específica no tempo. A medida de intensidade correspondente é $\Lambda([t_1, t_2]) = \lambda(t_2 - t_1)$ e a função de intensidade é dada por λ.

Para explicar processos pontuais como uma representação para valores extremos, é necessária uma ideia de convergência que é análoga a convergência de variáveis aleatórias.
Definição: *Seja $N_1, N_2, ...$ uma sequência de processos pontuais em $\mathcal{A}$. A sequência é dita ter convergência em distribuição para N, denotada por*

$$N_n \xrightarrow{d} N,$$

se, para todas as escolhas de m e para todos os conjuntos fechados $A_1, ..., A_m$ tal que $P(N(dA_j) = 0) = 1$, $j = 1, \ldots, m$, onde dA_j é a fronteira de A, a distribuição conjunta de $(N_n(A_1), \ldots, N_n(A_m))$ converge em distribuição para $(N(A_1), \ldots, N(A_m))$.

3.3.2 Processo de Poisson para extremos

Seja $X_1, X_2, ...$ uma série de variáveis aleatórias iid com função de distribuição F. Seja $M_n = \max\{X_1, \ldots, X_n\}$ sequências com constantes $\{a_n > 0\}$ e $\{b_n > 0\}$ tais que

$$P((M_n - a_n)/b_n \leq z) \to H(z \mid \xi, \psi, \mu),$$

onde $H(z \mid \xi, \psi, \mu)$, é a função de distribuição GEV dada em (3.1). Então, define-se uma sequência de processos pontuais N_n em R^2 por

$$N_n = \{(i/(n+1), (X_i - b_n)/a_n) : i = 1, \ldots, n\}.$$

A escala da primeira coordenada assegura que esta estará sempre no intervalo $(0, 1)$. A segunda coordenada estabelece o comportamento dos excessos quanto $n \to \infty$. Um resultado fundamental é dado pelo teorema a seguir, que pode ser visto em Coles (2001).

Teorema 3 *Seja $X_1, X_2, \ldots$ uma sequencia de variáveis aleatórias iid nas quais existem sequências $\{a_n > 0\}$ e $\{b_n\}$ tais que*

$$P((M_n - a_n)/b_n \leq z) \to H(z \mid \xi, \psi, \mu),$$

e seja z_- e z_+ os limites superior e inferior de H respectivamente. Então, a sequência de processos pontuais

$$N_n = \{(i/(n+1), (X_i - b_n)/a_n) : i = 1, \ldots, n\}$$

em regiões da forma $(0,1) \times [u, \infty)$, para qualquer $u > z_-$, converge para um processo de Poisson bidimensional, com medida de intensidade em $A = [0,1] \times [0, \infty]$.

Utilizando o teorema para valores extremos, considere o conjunto $A_x = [0,1] \times [x, \infty)$, para algum $x > u$. Pela propriedade da Poisson,

$$\begin{aligned} P\{\text{não ter pontos em } A_x\} &= \exp\{-\Lambda(A_x)\} \\ &= P\{M_n \leq x\} \\ &\to \exp\left\{-\left(1 + \xi\frac{x-\mu}{\psi}\right)^{-1/\xi}\right\} \end{aligned}$$

e então, pela homogeneidade do processo de Poisson,

$$\Lambda(A) = \left[1 + \xi\left(\frac{x-\mu}{\psi}\right)\right]^{-1/\xi}. \tag{3.13}$$

3.3.3 Verossimilhança dos excessos do limiar

Seja $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas. A distribuição dos excessos acima de um limiar possui uma distribuição GPD. Sem perda de generalidade, suponha também que a série $X_1, \ldots, X_n$ corresponde a um ano de observações. Originalmente, a verossimilhança do modelo de excessos ignora X_i menores do que u. Primeiro, pode-se incluir uma informação parcial destas observações. Seja $\varsigma = P(X_i > u)$, tal que, por Coles (2001), pode ser aproximado por

$$\varsigma \approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\psi}\right)\right]^{-\frac{1}{\xi}},$$

onde (μ, ψ, ξ) são os parâmetros da distribuição dos máximos anuais que possui distribuição GEV.

A contribuição para um X_i que não excede u é dada por $P(X_i < u) = 1 - \varsigma$.

Para um X_i que excede u, a contribuição da verossimilhança é dada por

$$P(X_i = x, X_i > u) = P(X_i > u) f(x \mid X_i > u) = \varsigma g(x \mid \Psi),$$

onde $\Psi = (\xi, \sigma, u)$ são os parâmetros da distribuição GPD, com densidade g.

Assim, a função de verossimilhança é dada por

$$(1 - \varsigma)^{n - n_u} \prod_{i=1}^{n_u} \varsigma g(x_i \mid \Psi),$$

onde n é o número total de pontos e n_u é o número de pontos acima do limiar u

Tancredi et al. (2006) formula um modelo que modifica a contribuição em um X_i que não excede u, dada por

$$P(X_i = x, X_i < u) = P(X_i < u) f(x \mid X_i < u) = (1 - \varsigma) h_T(x \mid \theta),$$

onde $h_T(x \mid \theta)$ é uma mistura de densidades Uniformes.

Tancredi et al. (2006) escreve a densidade do modelo da seguinte maneira

$$f_T(x|\theta, \Psi) = \begin{cases} (1 - \varsigma) h_T(x \mid \theta), & \text{se} \quad x \leq u \\ \varsigma g(x|\Psi), & \text{se} \quad x > u. \end{cases}$$

Assim, a função de verossimilhança é dada por

$$\prod_{x_i < u} \left[(1 - \varsigma) h_T(x_i \mid \theta) \right] \prod_{x_i > u} \left[\varsigma g(x_i \mid \Psi) \right].$$

No modelo $MGPD_k$ proposto neste trabalho, a densidade em (3.7) pode ser escrita também como

$$f(x|\theta, \mathbf{p}, \Psi) = \begin{cases} (1 - P(x > u)) \frac{h(x|\mu, \eta, \mathbf{p})}{H(u|\mu, \eta, \mathbf{p})}, & \text{se} \quad x \leq u \\ P(x < u) g(x|\Psi), & \text{se} \quad x > u \end{cases}, \tag{3.14}$$

onde h e H são respectivamente as funções densidade e acumulada de mistura de Gamas e g é a densidade da distribuição GPD. Note que, pela maneira como é formulado o modelo $MGPD_k$, $P(X_i > u)$ não é aproximado por ς e sim é igual a $(1 - H(u \mid \theta))$.

Desta maneira, a contribuição para um x que não excede u é escrita da seguinte maneira

$$P(X_i = x, X_i < u) = P(X_i < u) f(x \mid X_i < u) = H(u \mid \theta, \mathbf{p}) \frac{h(x \mid \theta)}{H(u \mid \theta, \mathbf{p})} = h(x \mid \theta, \mathbf{p}),$$

e, para x que excede u

$$P(X_i = x, X_i > u) = P(X_i > u) f(x \mid X_i > u) = (1 - H(u \mid \theta, \mathbf{p})) g(x \mid \Psi).$$

Assim, tem-se a mesma forma da função de verossimilhança dada por

$$\prod_{x_i < u} \left[h(x \mid \theta, \mathbf{p}) \right] \prod_{x_i > u} \left[(1 - H(u \mid \theta, \mathbf{p})) g(x \mid \Psi) \right],$$

que é a mesma verossimilhança em (3.11).

3.4 Simulações

Para demonstrar a eficiência do método, foram simulados pontos de uma distribuição de mistura. Foram simuladas amostras de tamanho 1000 e 10000 de mistura de duas Gamas com os mesmos parâmetros das simulações da Seção 2.10, ou seja, $\mu = (2; 8)$, $\eta = (4; 8)$ e $p = (0,66; 0,33)$. Após simular as observações, toma-se um percentil dos dados como sendo o valor u do limiar destas distribuições. Foram feitas simulações tomando o limiar no percentil 90 das simulações. Na simulação, todas as observações que ultrapassaram o limiar foram removidas, gerando no lugar dessas observações pontos da distribuição GPD com o limiar no percentil 90, e com diversos valores nos parâmetros da cauda, onde $\sigma = (2; 3; 5)$ e $\xi = (-0,4; 0,4)$. Em cada simulação, foi feita a modelagem considerando $k = 1, 2$ ou 3 componentes na mistura para as observações antes da cauda. A variação destes parâmetros tem os seguintes objetivos:

- Variando o tamanho da amostra: verificar como o número de observações acima da cauda irá interferir na estimação do limiar e dos parâmetros acima dele. Poucas observações podem dificultar a estimação da cauda, pois como o limiar está no quantil 90, cerca de apenas 10% das observações fazem parte da cauda.

- Variando ξ: verificar se o grau de dificuldade da estimação muda de acordo com o peso da cauda(ξ positivo $\rightarrow$ cauda pesada, ξ negativo $\rightarrow$ cauda leve). O valor $-0,4$ foi escolhido para verificar como se comporta a estimação do parâmetro ξ para valores próximos do limite onde este parâmetro possui uma forma regular no estimador de máxima verossimilhança ($\xi > -0,5$).

- Variando σ: verificar o quanto o tamanho e direção do salto no limiar entre a distribuição de mistura e a GPD interfere na estimação do limiar e dos parâmetros da cauda.

- Variando k: verificar se o número de componentes utilizado na mistura irá interferir na estimação do limiar e dos parâmetros da cauda.

Além de comparar o modelo $MGPD_k$ com diversos valores de k, foi comparado também o modelo MG_k, para verificar com qual eficiência o modelo MG_k consegue estimar a densidade de distribuições na cauda. Para comparar graficamente com os modelos $MGPD_k$, será apresentado o modelo MG_k que apresentar o melhor BIC e DIC.

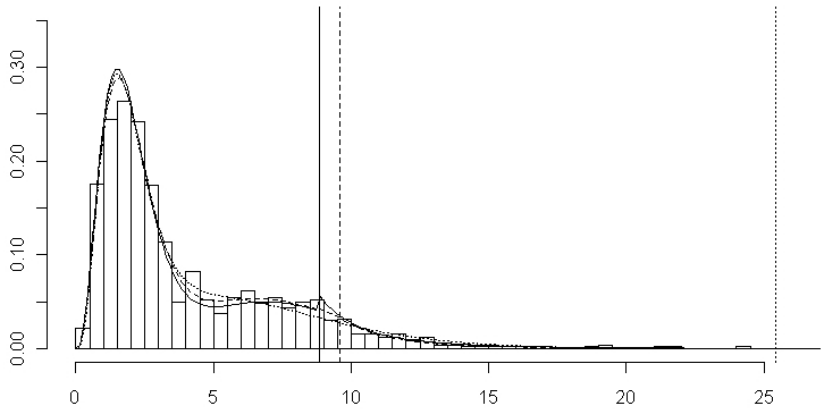
Para aplicar a técnica MCMC, foram feitas cadeias com 30 mil iterações, onde foi considerado um *burn-in* de 10 mil e coletadas uma amostra a cada 10.

A distribuição a priori para as componentes de mistura são as mesmas do Capítulo 2. Para os parâmetros da cauda, é utilizada a priori de Jeffreys vista em (3.10). Em relação ao limiar, para $n = 10000$, foi dada como distribuição a priori uma $N(u_0, 100)$, onde $u_0 = u$. Para amostras de tamanho $n = 1000$, houve dificuldade na estimação do limiar quando se utilizou a mesma distribuição a priori.

A Figura 3.1 mostra a dificuldade de estimação do limiar quando a variância é 100, onde a média a posteriori do limiar foi para o final dos dados, muito longe do verdadeiro valor do limiar utilizado na simulação. A variância teve que ser diminuída para 1 a fim de obter uma boa estimação do limiar. Com esta priori, a probabilidade do limiar ser menor que 8 ou maior que 12 é de 95%.

Figura 3.1: Densidade preditiva para dados simulados com $\xi = 0,4$, $k = 2$ $\sigma = 2$, $n = 1000$ e limiar no valor 9,24.

Linha cheia - distribuição verdadeira; tracejada: estimação com priori do limiar com variância 1; pontilhada: estimação com priori do limiar com variância 100. As linhas verticais representam a média a posteriori do limiar.



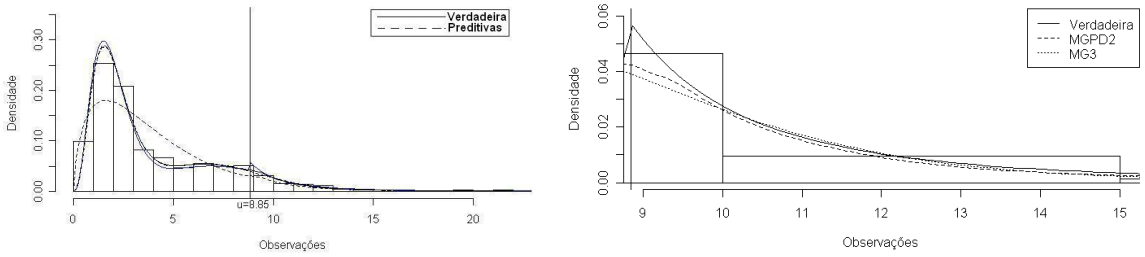
Simulação com $\sigma = 2$

ξ=0,4

A Figuras 3.2 e 3.3 mostram a densidade preditiva para amostras de tamanho 1000 e 10000 para o modelo $MGPD_k$ com k=1, 2 e 3, além do modelo MG_k com melhor BIC e DIC na classe dos modelos só com mistura, de acordo com a Tabela 3.1. Com exceção do modelo $MGPD_1$, as densidades preditivas estão quase sobrepostas à verdadeira densidade, não tendo praticamente diferença para a estimação entre $MGPD_2$ ou $MGPD_3$. Ampliando as figuras das distribuições preditivas na cauda, percebe-se que tanto para $n = 1000$ quanto para $n = 10000$, os modelos MG_k têm maiores dificuldades em predizer a densidade nesta região do que para os modelos $MGPD_k$, $k > 1$, principalmente nos valores próximos do limiar. Nos modelos $MGPD_2$ e $MGPD_3$, para $n = 10000$, a predição

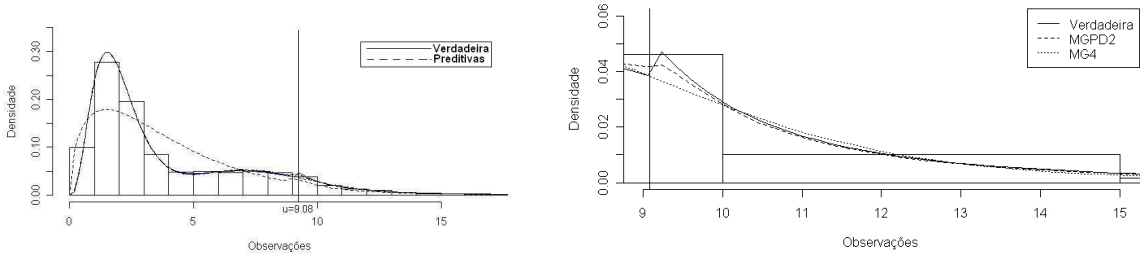
foi mais precisa do que para $n = 1000$, na região em torno do limiar.

Figura 3.2: Densidade preditiva para $\sigma = 2$, $\xi = 0,4$ e $n = 1000$.
No painel à esquerda, o único modelo distante da curva verdadeira é com $k_{est} = 1$.



a) Densidade Preditiva. b) Densidade da cauda.

Figura 3.3: Densidade preditiva para $\sigma = 2$, $\xi = 0,4$ e $n = 10000$.
No painel à esquerda, o único modelo distante da curva verdadeira é com $k_{est} = 1$.

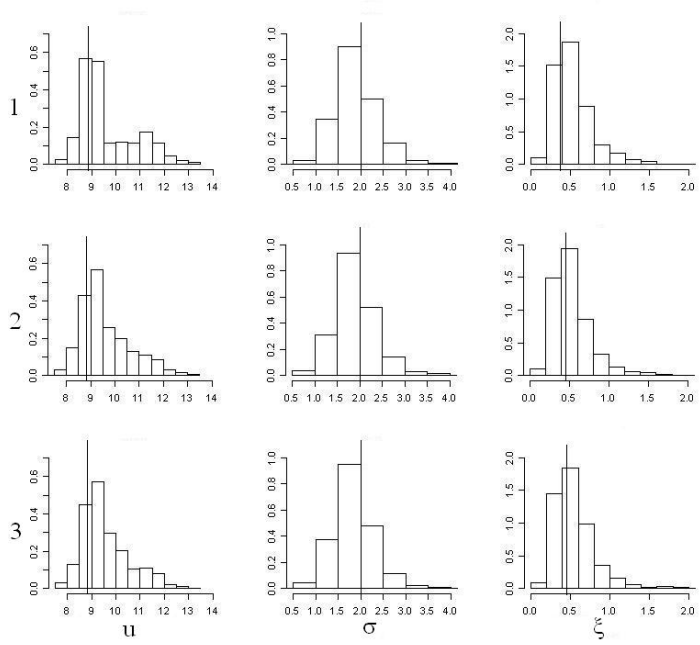


a) Densidade Preditiva. b) Densidade da cauda.

A semelhança entre os modelos também pode ser visto nos parâmetros da cauda do modelo $MGPD_k$, onde o histograma da distribuição a posteriori (Figuras 3.4 e 3.5) são praticamente os mesmos para $k = 1$, $k = 2$ ou $k = 3$. Isto ocorre pois, para diferentes números de componentes, os valores no limiar foram próximos. Com relação ao tamanho da amostra, percebe-se que a modelagem com $n = 10000$ fornece resultados mais precisos do que $n = 1000$, com a densidade preditiva próxima da verdadeira densidade, e com menor variabilidade nos parâmetros da cauda, como pode ser observado na Figura 3.5.

Figura 3.4: Histograma da posteriori dos parâmetros da cauda para $\sigma = 2$, $\xi = 0,4$ e $n = 1000$.

As linhas representam o número de componentes k=1, 2 e 3, e as colunas são os parâmetros da cauda u, σ e ξ, nesta ordem. As linhas verticais são os valores verdadeiros simulados.



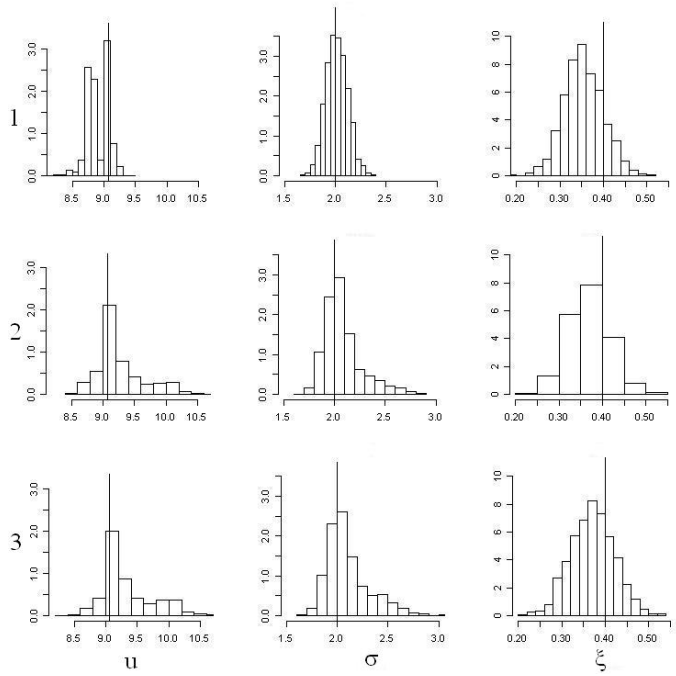
Assim como no Capítulo 2, no modelo $MGPD_k$, quando $k_{est} > k_{sim}$, os pesos em excesso vão para 0 e o número de pesos que são significativamente diferentes de 0 é igual a k_{sim}, como pode-se ver na Figura 3.6, onde um dos pesos do modelo com $k_{est} = 3$ estaciona em torno de 0 e os outros dois pesos convergem para uma distribuição igual aos pesos do modelo com $k_{est} = 2$.

ξ=-0,4

Observando as densidades preditivas (Figuras 3.7 e 3.8), assim como o histograma dos parâmetros da cauda (Figuras 3.9 e 3.10), nos modelos $MGPD_k$ os valores para $k = 2$ e $k = 3$ são próximos, sendo que a densidade que difere significativamente da verdadeira distribuição apenas para $k = 1$. Aqui, mesmo para uma componente, a estimação do

Figura 3.5: Histograma da posteriori dos parâmetros da cauda para $\sigma = 2$, $\xi = 0,4$ e $n = 10000$.

As linhas representam o número de componentes k=1, 2 e 3, e as colunas são os parâmetros da cauda u, σ e ξ, nesta ordem. As linhas verticais são os valores verdadeiros simulados.



limiar é muito próxima em relação ao resultado obtido das outras componentes quando $n = 10000$. Observando a parte b das Figuras 3.7 e 3.8, verifica-se que o modelo $MGPD_2$, no caso de $n = 10000$, a estimação parece ser mais eficiente para a cauda do que os modelos da classe MG_k. Além disso, para $n = 10000$, percebe-se que a estimação pelo modelo $MGPD_2$ detecta o salto da distribuição no limiar.

Foram calculadas as medidas de ajuste BIC e DIC, para os diversos modelos utilizados. Pela Tabela 3.1, para $n = 10000$, os menores BIC e DIC foram aqueles do modelo utilizado na simulação, ou seja, o modelo $MGPD_2$. Para $n = 1000$, em duas situações o melhor modelo apontado foram os da classe MG_k, enquanto que nas outras duas configurações o melhor modelo foi o $MGPD_2$.

Figura 3.6: Série dos pesos para $\sigma = 2$, $\xi = 0,4$ e $n = 10000$.
A esquerda o modelo $MGPD_2$ e à direita o modelo $MGPD_3$.

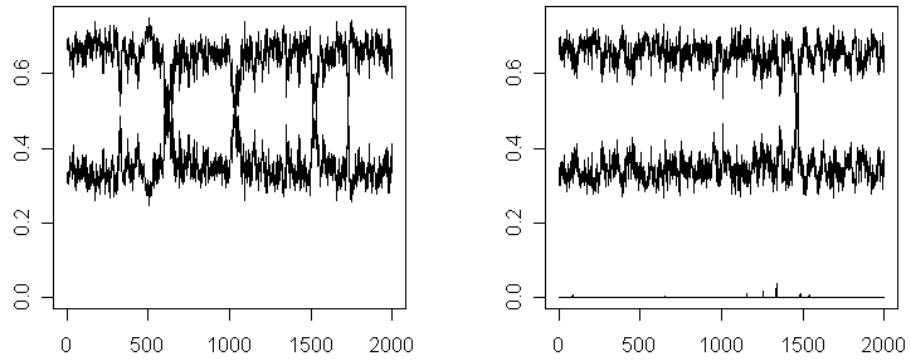
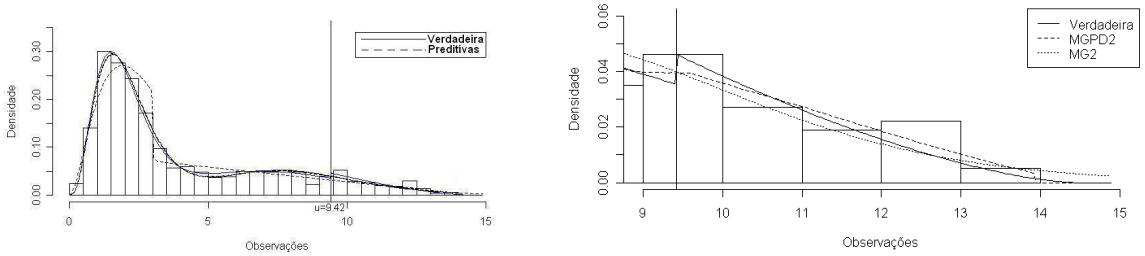


Figura 3.7: Densidade preditiva para $\sigma = 2$, $\xi = -0,4$ e $n = 1000$.
No painel à esquerda, o único modelo distante da curva verdadeira é com $k_{est} = 1$.



a) Densidade Preditiva.

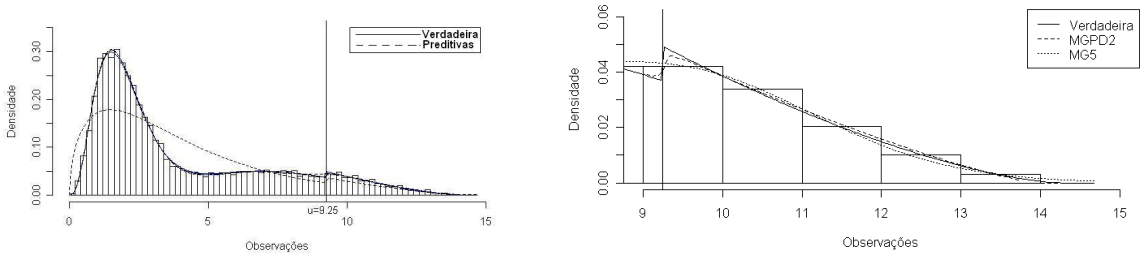
b) Densidade da cauda.

Simulação com $\sigma = 3$

ξ=0,4

A Figuras 3.11 e 3.12 mostram a densidade preditiva com $n = 1000$ e $n = 10000$ para o modelo $MGPD_k$, com k=1, 2 e 3, além do melhor modelo MG_k segundo o BIC e DIC (Tabela 3.2). Com exceção do modelo $MGPD_1$, as densidades apresentam um bom ajuste

Figura 3.8: Densidade preditiva para $\sigma = 2$, $\xi = -0,4$ e $n = 10000$.
No painel à esquerda, o único modelo distante da curva verdadeira é com $k_{est} = 1$.



a) Densidade Preditiva. b) Densidade da cauda.

Figura 3.9: Histograma da posteriori dos parâmetros da cauda para $\sigma = 2$, $\xi = -0,4$ e $n = 1000$.
As linhas representam o número de componentes k=1, 2 e 3, e as colunas são os parâmetros da cauda u, σ e ξ, nesta ordem. As linhas verticais são os valores verdadeiros simulados.

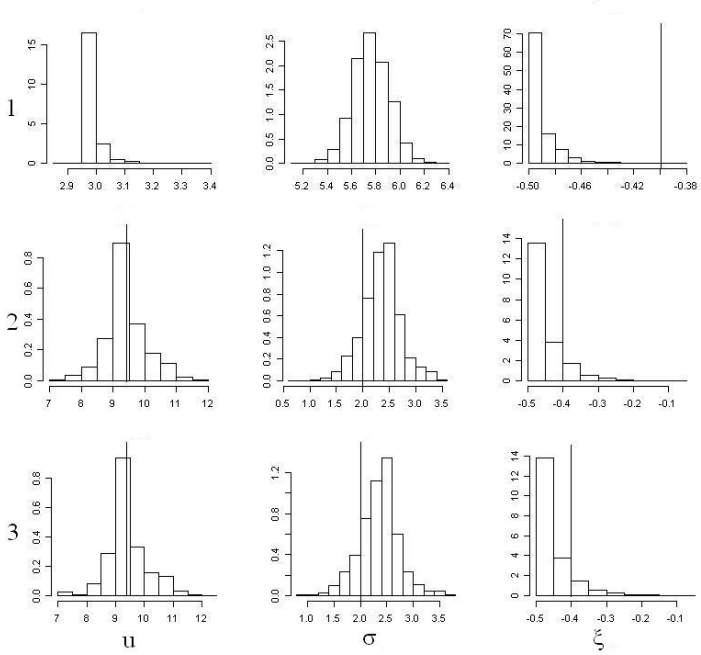
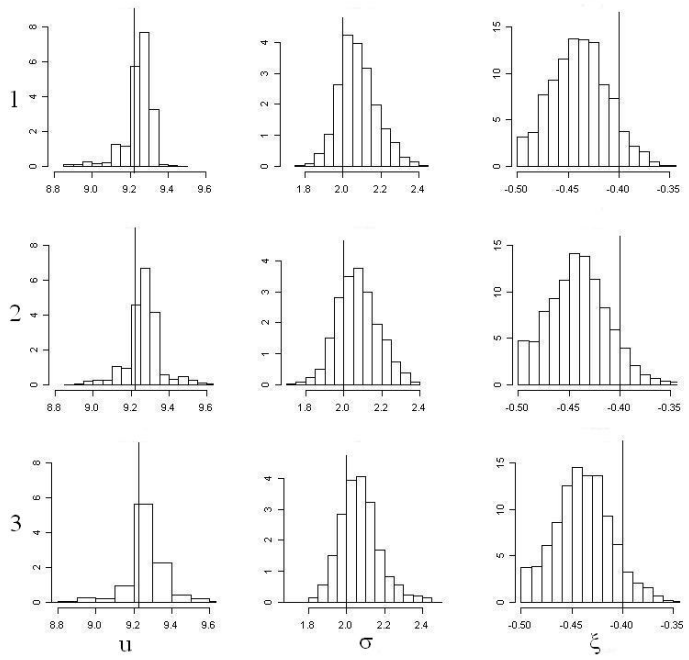


Figura 3.10: Histograma da posteriori dos parâmetros da cauda para $\sigma = 2$, $\xi = -0,4$ e $n = 10000$.

As linhas representam o número de componentes k=1, 2 e 3, e as colunas são os parâmetros da cauda u, σ e ξ, nesta ordem. As linhas verticais são os valores verdadeiros simulados.



em relação à verdadeira densidade na parte central dos dados, como mostra a parte a das figuras.

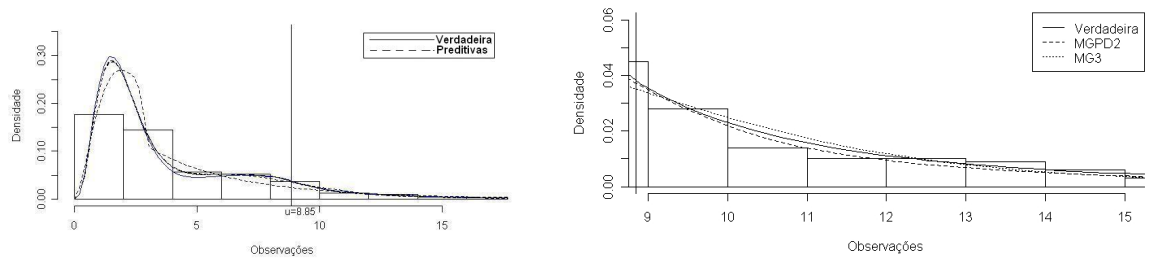
Pelas Figuras 3.13 e 3.14 do histograma dos parâmetros da cauda do modelo $MGPD_k$, o número de componentes da mistura parece ter pouca influência na estimação da cauda para $n = 1000$. A simulação com σ=3 é a que apresenta o menor salto no limiar, praticamente não tendo descontinuidade. Isto dificulta na estimação do limiar, que pelo histograma da Figura 3.14 tem uma distribuição bimodal, onde a segunda moda está em torno do valor verdadeiro do limiar e a primeira um pouco antes.

Tabela 3.1: Medidas de ajuste para σ=2.

Modelo	Pd	DIC	BIC	Pd	DIC	BIC
	n=1000			n=10000		
	$\xi = 0,4$					
$MGPD_1$	0,77	4653,7	4691,5	2,09	46817	46857
$MGPD_2$	6,79	4468,2	**4542,5**	7,77	**44718**	**44814**
$MGPD_3$	6,50	4469,1	4563,1	4,61	44727	44842
MG_3*	7,70	**4466,6**	4543,9	7,68	44734	44859
	$\xi = -0,4$					
$MGPD_1$	2,85	4388,9	4435,3	5,71	45372	45440
$MGPD_2$	6,77	**4345,5**	4421,7	8,65	**43171**	**43272**
$MGPD_3$	6,61	4345,8	4442,3	8,57	43172	43300
MG_2*	5,18	4366,1	**4418,1**	8,94	43211	43368

* Modelo MG_k com menor BIC e DIC.

Figura 3.11: Densidade preditiva para $\sigma = 3$, $\xi = 0,4$ e $n = 1000$.
No painel à esquerda, o único modelo distante da curva verdadeira é com $k_{est} = 1$.



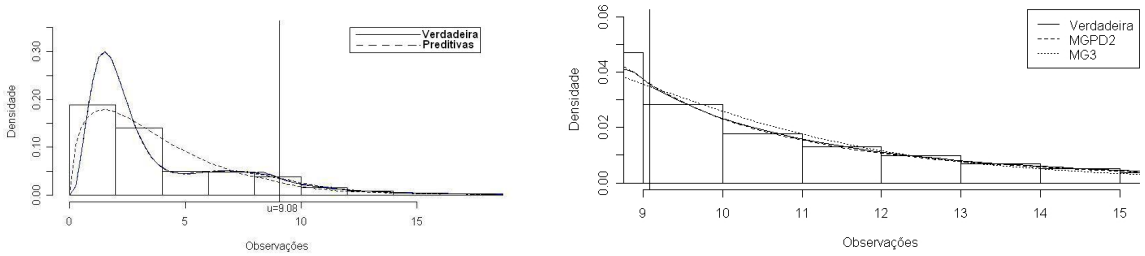
a) Densidade Preditiva.

b) Densidade da cauda.

ξ=-0,4

As Figuras 3.15 e 3.16 da distribuição preditiva apresentam um bom ajuste para todos

Figura 3.12: Densidade preditiva para $\sigma = 3$, $\xi = 0,4$ e $n = 10000$.
No painel à esquerda, o único modelo distante da curva verdadeira é com $k_{est} = 1$.



a) Densidade Preditiva. b) Densidade da cauda.

Figura 3.13: Histograma da posteriori dos parâmetros da cauda para $\sigma = 3$, $\xi = 0,4$ e $n = 1000$.
As linhas representam o número de componentes k=1, 2 e 3, e as colunas são os parâmetros da cauda u, σ e ξ, nesta ordem. As linhas verticais são os valores verdadeiros simulados.

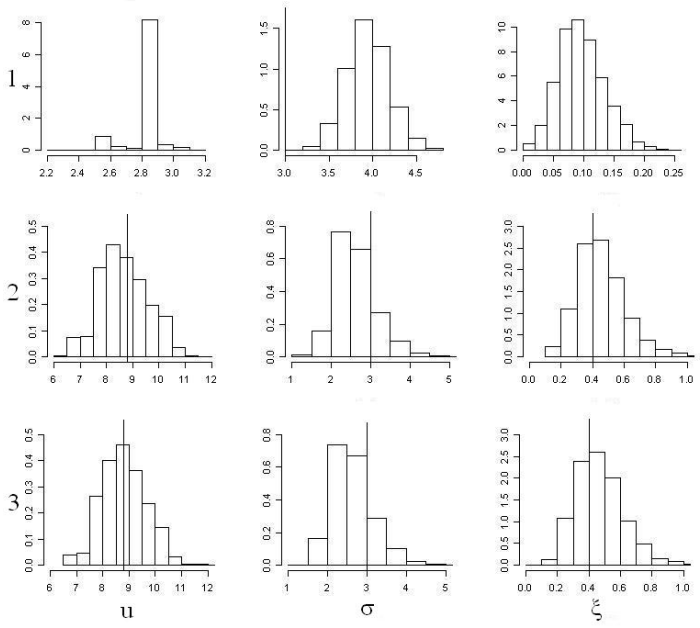
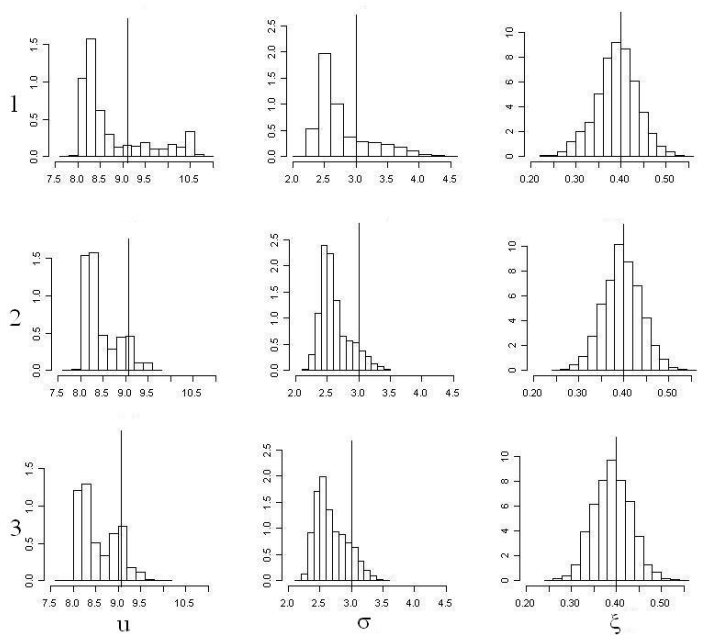


Figura 3.14: Histograma da posteriori dos parâmetros da cauda em $\sigma = 3$, $\xi = 0,4$ e $n = 10000$.

As linhas representam o número de componentes k=1, 2 e 3, e as colunas são os parâmetros da cauda u, σ e ξ, nesta ordem. As linhas verticais são os valores verdadeiros simulados.

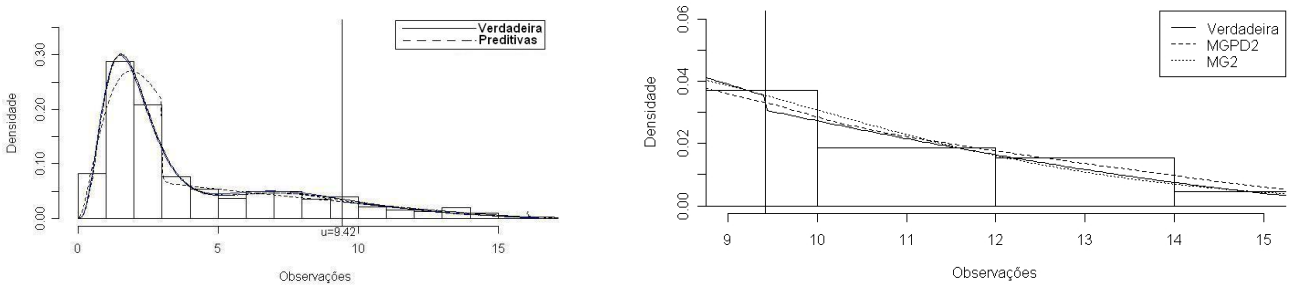


os modelos, com exceção do modelo $MGPD_1$, como mostra a parte a de ambas as figuras. Percebe-se que a estimação pelo modelo $MGPD_k$ é melhor que os modelos MG_k na cauda da distribuição, como mostra a parte b das figuras.

Para cada modelagem de $\sigma = 3$, foram calculadas as medidas de ajuste BIC e DIC. Pela Tabela 3.2, assim como na tabela de $\sigma = 2$, os menores valores do BIC e DIC ocorrem exatamente no modelo verdadeiro simulado, o $MGPD_2$. A única exceção é quando $\xi = -0,4$ e $n = 1000$, onde o melhor modelo foi o MG_2.

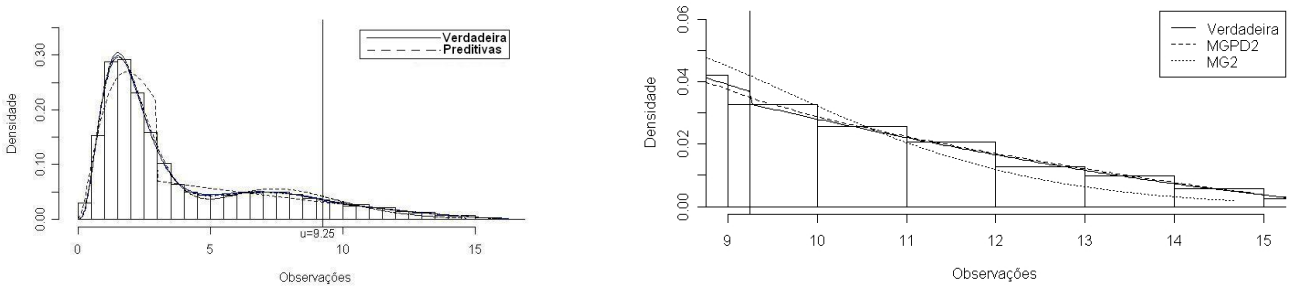
Simulação com $\sigma = 5$

Figura 3.15: Densidade preditiva para $\sigma = 3$, $\xi = -0,4$ e $n = 1000$.
No painel à esquerda, o único modelo distante da curva verdadeira é com $k_{est} = 1$.



a) Densidade Preditiva. b) Densidade da cauda.

Figura 3.16: Densidade preditiva para $\sigma = 3$, $\xi = -0,4$ e $n = 10000$.
No painel à esquerda, o único modelo distante da curva verdadeira é com $k_{est} = 1$.



a) Densidade Preditiva. b) Densidade da cauda.

ξ=0,4

As Figuras 3.17 e 3.18 mostram a distribuição preditiva do modelo com diferentes componentes na mistura para $n = 1000$ e $n = 10000$. Percebe-se que, para $n = 10000$, as densidades preditivas ficam mais próximas da verdadeira densidade do que para $n = 1000$, com exceção do modelo $MGPD_1$, que ficou longe da verdadeira densidade em ambos os casos. A região próxima do limiar é a única em que o melhor modelo da classe MG_k

Tabela 3.2: Medidas de ajuste para σ=3.

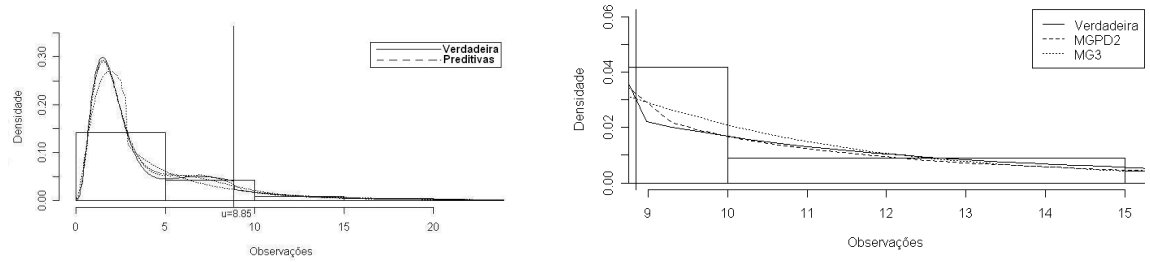
Modelo	Pd	DIC	BIC	Pd	DIC	BIC
	n=1000			n=10000		
	$\xi = 0,4$					
$MGPD_1$	3,61	4603,9	4648,1	2,04	47610	47665
$MGPD_2$	7,81	**4547,1**	**4624,6**	5,68	**45530**	**45621**
$MGPD_3$	7,83	4547,4	4645,4	4,27	45534	45649
MG_3*	7,41	4552,2	4628,4	7,17	45560	45656
	$\xi = -0,4$					
$MGPD_1$	3,56	4462,7	4510,1	1,79	44416	44467
$MGPD_2$	6,53	4426,3	4501,9	7,51	**43986**	**44083**
$MGPD_3$	6,76	**4425,7**	4522,6	4,32	43997	44112
MG_2*	5,27	4440,9	**4493,1**	4,79	44064	44129

* Modelo MG_k com menor BIC e DIC.

fica distante da verdadeira densidade, como percebe-se na parte b das Figuras 3.17 e 3.18, enquanto que para $k > 1$, considerando o modelo $MGPD_k$, a estimação parece ser precisa em toda a cauda. Para $n = 10000$, o modelo $MGPD_2$ consegue detectar o salto da distribuição no limiar.

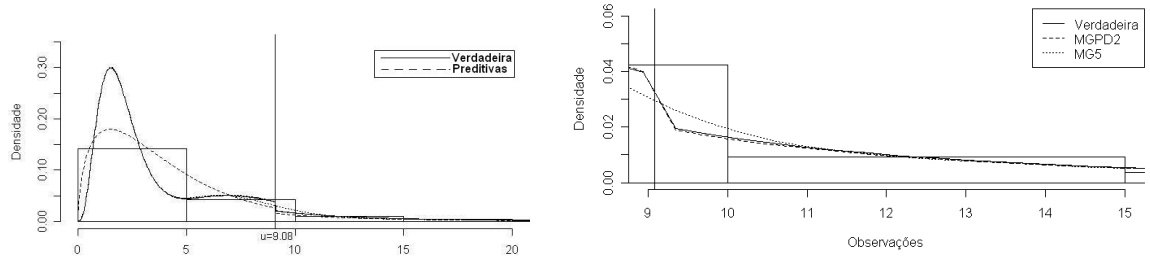
As Figuras 3.19 e 3.20 mostram o histograma da posteriori dos parâmetros da cauda. Observando os histogramas, para $n = 10000$, as distribuições parecem ser similares, independente do número de componentes na mistura, enquanto que, para $n = 1000$ isto só acontece entre $k = 2$ e $k = 3$, pois para $k = 1$ a distribuição antes da cauda é diferente e isto leva a uma estimação de valores diferentes para o limiar, implicando em uma outra distribuição para a cauda.

Figura 3.17: Densidade preditiva para $\sigma = 5$, $\xi = 0,4$ e $n = 1000$.
No painel à esquerda, o único modelo distante da curva verdadeira é com $k_{est} = 1$.



a) Densidade Preditiva. b) Densidade da cauda.

Figura 3.18: Densidade preditiva para $\sigma = 5$, $\xi = 0,4$ e $n = 10000$.
No painel à esquerda, o único modelo distante da curva verdadeira é com $k_{est} = 1$.



a) Densidade Preditiva. b) Densidade da cauda.

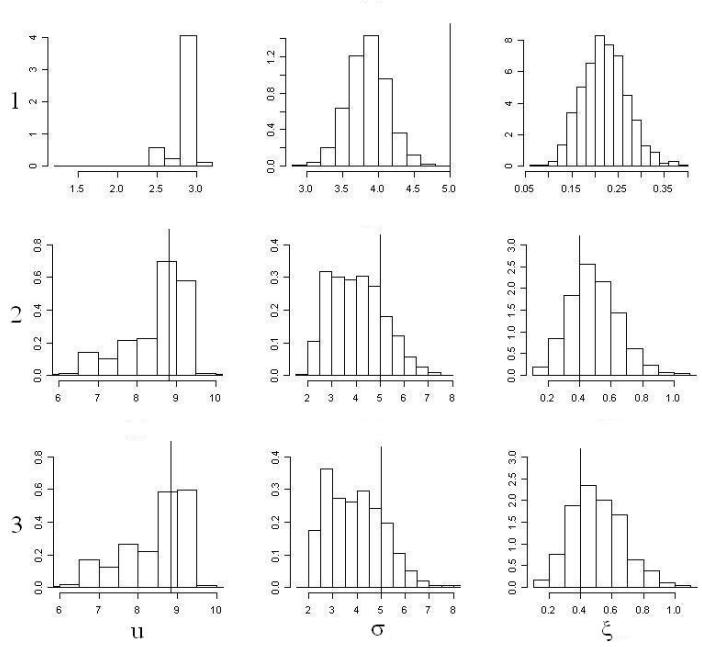
ξ=-0,4

As Figuras 3.21 e 3.22 mostram a dificuldade do modelo MG_k em estimar o início da cauda da distribuição. Com relação ao modelo $MGPD_k$, percebe-se que, para $n = 10000$ a estimação se mostra eficiente na cauda, enquanto que, para $n = 1000$, a densidade preditiva dos modelos $MGPD_k$ tem uma maior dificuldade de acompanhar a distribuição a partir do salto no limiar, como mostra a parte b da Figura 3.21.

Outro fator interessante relacionado ao salto no limiar é que a predição nos saltos foi mais precisa quando o salto é para baixo do que quando é para cima, como se observa nos

Figura 3.19: Histograma da posteriori dos parâmetros da cauda para $\sigma = 5$, $\xi = 0,4$ e $n = 1000$.

As linhas representam o número de componentes k=1, 2 e 3, e as colunas são os parâmetros da cauda u, σ e ξ, nesta ordem. As linhas verticais são os valores verdadeiros simulados.



gráficos das distribuições preditivas quando $\sigma = 2$ (salto para cima) e $\sigma = 5$ (salto para baixo). O tamanho da amostra também interfere da detecção deste salto. Para $n = 1000$, em todos os casos houve dificuldade de encontrar o salto no limiar.

A Tabela 3.3 apresenta BIC e DIC para $\sigma = 5$. Pela tabela, percebe-se as mesmas características das duas tabelas anteriores. Para amostras de tamanho 1000, o BIC mostrou ser uma boa medida de comparação de modelos, enquanto que para $n = 10000$, tanto BIC quanto o DIC mostraram-se boas medidas, pois detectaram como melhor modelo aquele utilizado na simulação, ou seja, o $MGPD_2$.

3.4.1 Cálculo dos quantis

Além de encontrar as medidas de ajuste, como o interesse principal deste trabalho é

Figura 3.20: Histograma da posteriori dos parâmetros da cauda em $\sigma = 5$, $\xi = 0,4$ e $n = 10000$.

As linhas representam o número de componentes k=1, 2 e 3, e as colunas são os parâmetros da cauda u, σ e ξ, nesta ordem. As linhas verticais são os valores verdadeiros simulados.

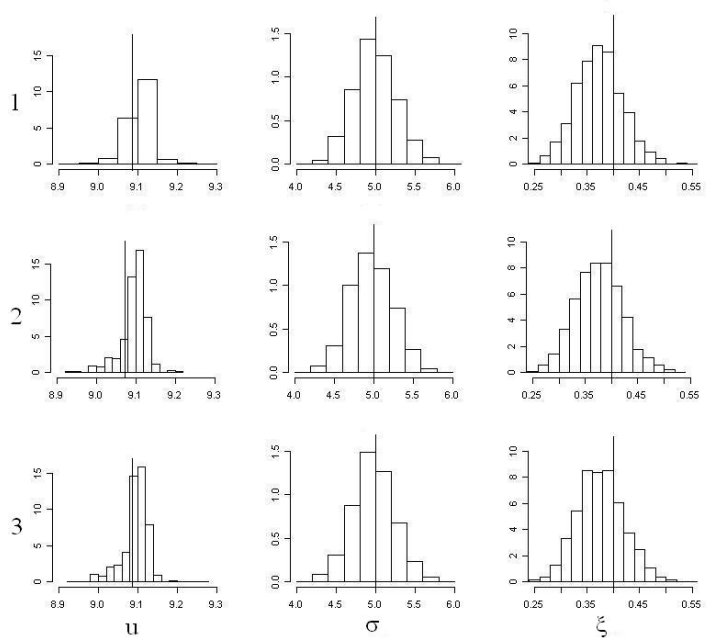
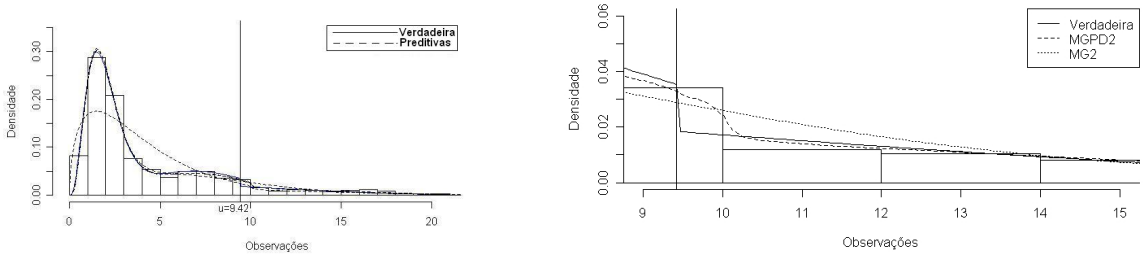


Figura 3.21: Densidade preditiva para $\sigma = 5$, $\xi = -0,4$ e $n = 1000$.

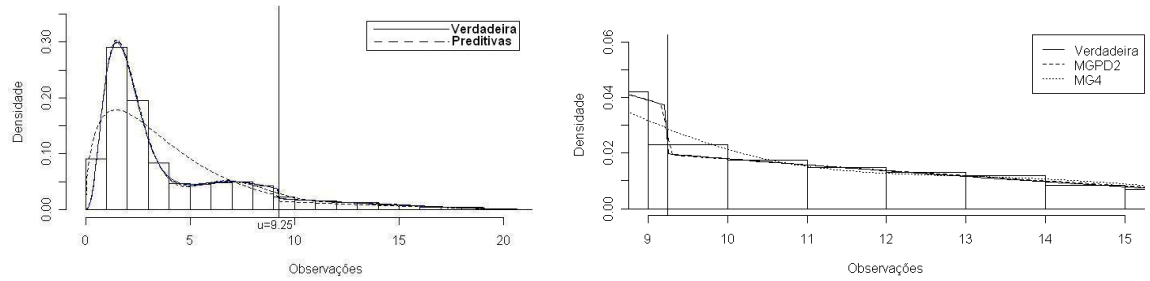
No painel à esquerda, o único modelo distante da curva verdadeira é com $k_{est} = 1$.



a) Densidade Preditiva.

b) Densidade da cauda.

Figura 3.22: Densidade preditiva para $\sigma = 5$, $\xi = -0,4$ e $n = 10000$.
No painel à esquerda, o único modelo distante da curva verdadeira é com $k_{est} = 1$.



a) Densidade Preditiva.

b) Densidade da cauda.

Tabela 3.3: Medidas de ajuste para σ=5.

Modelo	Pd	DIC	BIC	Pd	DIC	BIC
	n=1000			n=10000		
	$\xi = 0,4$					
$MGPD_1$	-5,89	4741,8	4746,3	6,55	48621	48691
$MGPD_2$	4,70	**4660,6**	**4728,1**	7,20	**46548**	**46644**
$MGPD_3$	4,17	4662,5	4749,3	7,07	46548	46671
MG_3*	2,68	4684,8	4742,3	9,98	46594	46753
	$\xi = -0,4$					
$MGPD_1$	-140,22	5119,9	4694,4	3,06	47215	47275
$MGPD_2$	8,81	4522,6	**4603,8**	7,80	**45008**	**45106**
$MGPD_3$	9,03	**4522,1**	4624,6	7,57	45009	45133
MG_2*	5,12	4556,0	4607,9	6,81	45057	45181

* Modelo MG_k com menor BIC e DIC.

analisar dados extremos, também foram calculados os quantis altos dos modelos MG_k, $MGPD_k$, e do quantil empírico. Estes quantis foram comparados com os verdadeiros.

A Tabela 3.4 mostra os resultados dos quantis para as simulações com $\xi = 0,4$ dos modelos $MGPD_k$ e do MG_k com melhores medidas de ajuste BIC e DIC. Percebe-se que quando $n = 1000$, nenhum método parece levar vantagem, pois das 12 situações com este tamanho de amostra, em 5 a estimação mais próximas do quantil verdadeiro foram obtidos a partir da abordagem empírica, enquanto que em 4 foram obtidos dos modelos $MGPD_k$ e 3 dos modelos MG_k. Para $n = 10000$, a estimação dos quantis pelo modelo $MGPD_k$ mostrou-se mais eficiente que os outros métodos, ficando mais próximo do quantil verdadeiro em 8 das 12 situações apresentadas na tabela.

Tabela 3.4: Quantis das simulações para $\xi = 0,4$.

Prob	V	E	1	2	3	MG_k	V	E	1	2	3	MG_k
	n=1000						n=10000					
	$\sigma = 2$											
0,95	10,44	10,49	9,99	**10,41**	10,41	10,48	10,68	10,70	9,99	**10,68**	10,67	10,75
0,99	16,41	15,13	**15,01**	15,98	15,98	15,16	16,64	16,43	15,17	16,39	16,40	**16,51**
0,999	35,41	**33,82**	42,80	44,44	46,19	37,25	35,62	32,01	30,37	33,59	**33,77**	32,48
0,9999	83,09	53,98	5949,99	710,34	516,75	**65,23**	83,32	59,99	65,49	74,57	**75,49**	47,73
	$\sigma = 3$											
0,95	11,24	**11,30**	12,49	11,04	11,04	11,15	11,48	**11,51**	10,38	11,41	11,41	11,36
0,99	20,19	18,27	**20,01**	19,81	19,93	19,04	20,41	**20,09**	18,01	19,96	19,92	21,37
0,999	48,67	46,30	35,90	54,31	56,77	**48,84**	48,90	43,47	42,27	**47,18**	46,71	41,21
0,9999	120,21	76,75	54,85	183,3	209,8	**85,00**	120,45	85,44	103,65	**116,24**	113,67	60,35
	$\sigma = 5$											
0,95	12,84	**12,94**	13,77	12,45	12,42	12,44	13,07	**13,14**	11,37	13,02	13,04	12,76
0,99	27,75	24,56	25,99	27,36	**27,55**	29,03	27,98	**27,44**	24,22	27,27	27,29	29,67
0,999	75,22	**71,27**	53,52	92,13	96,11	64,42	75,45	66,41	63,22	**70,97**	70,80	53,72
0,9999	194,46	**121,68**	100,76	363,03	402,87	109,57	194,69	136,34	156,98	**177,13**	176,11	77,47

Prob=P(X$\leq q$), $V - Verdadeiro, E - Empírico$, 1−MGPD$_1$, 2-$MGPD_2$, 3-$MGPD_3$,$MG_k$-Melhor MG, segundo DIC

Comparando apenas os métodos MG_k contra os métodos $MGPD_k$, para $n = 1000$ em metade das situações o modelo MG_k esteve mais perto do quantil verdadeiro que

o modelo $MGPD_k$. Para $n = 10000$, em 11 das 12 comparações da tabela o modelo $MGPD_k$ estimou um quantil mais próximo do verdadeiro do que o modelo MG_k. Estes resultados mostram que é importante estimar a cauda por uma distribuição GPD, pois ela é mais eficiente na detecção de altos quantis dos dados do que uma aproximação não-paramétrica para a cauda por mistura.

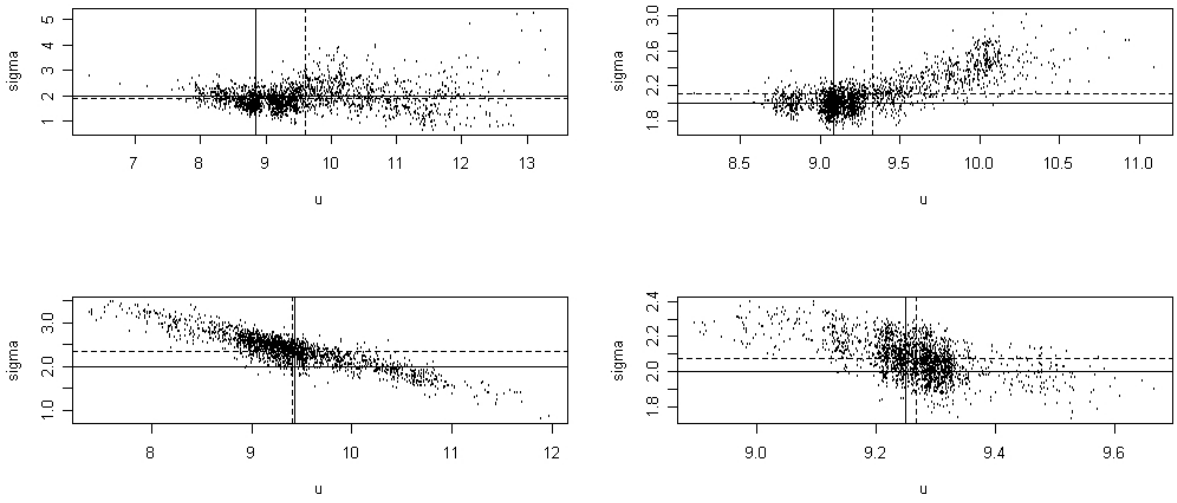
3.5 Identificação dos parâmetros da cauda

Um problema relacionado a distribuição GPD é a correlação entre o limiar e o parâmetro da cauda σ. Se o limiar é alterado para $u' > u$, então os novos excessos também são descritos por uma distribuição GPD com mesmo parâmetro de forma ξ, mas o parâmetro de escala será dado então por $\sigma' = \sigma + \xi(u' - u)$. Isto pode causar um problema de falta de identificabilidade na estimação dos parâmetros. Para verificar qual o grau de identificabilidade dos parâmetros, foi estudada a correlação entre os valores estimados do limiar e do parâmetro de forma σ.

A Figura 3.23 mostra os valores estimados para a distribuição a posteriori dos parâmetros nas 4 simulações onde $\sigma = 2$. Pela Figura, nota-se que, embora exista uma correlação entre os parâmetros, sendo estas positivas para $\xi = 0,4$ e negativas para $\xi = -0,4$ a grande massa dos valores da distribuição a posteriori está próxima da linha dos valores verdadeiros dos parâmetros, ou seja, a estimação conseguiu identificar bem os parâmetros da cauda mesmo havendo correlação entre eles.

A Figura 3.24 mostra os valores estimados para a distribuição a posteriori dos parâmetros nas 4 simulações onde $\sigma = 3$. Pela Figura, observa-se que a estimação do limiar e de σ ficam um pouco distantes da verdadeira quando $n = 10000$ e $\xi = 0,4$, lembrando que nas simulações com $n = 10000$ a distribuição a priori do limiar tem variância 100, enquanto que nas simulações com $n = 1000$ esta variância foi de 1. Portanto, adotando-se uma priori vaga para o limiar, pode ocorrer o problema da identificação dos parâmetros perder eficiência, embora que os valores encontrados da estimação não estejam tão distantes dos

Figura 3.23: Relação dos parâmetros da cauda para $\sigma = 2$.
Primeira linha: $\xi = 0,4$. Segunda linha: $\xi = -0,4$. Primeira coluna: $n = 1000$. Segunda coluna: $n = 10000$. As linhas cheias são os valores verdadeiros e as tracejadas são as médias a posteriori.



verdadeiros.

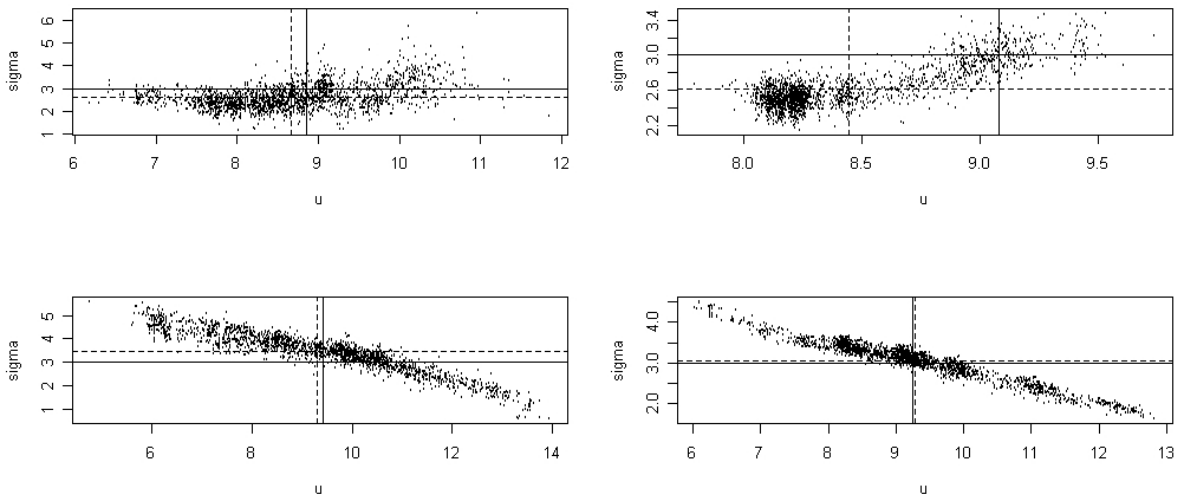
A Figura 3.25 mostra os valores estimados para a distribuição a posteriori dos parâmetros nas 4 simulações onde $\sigma = 5$. Para $\xi = -0,4$, parece haver duas massas de pontos, onde σ não parece possuir correlação com o limiar e este limiar possui uma distribuição bimodal.

3.5.1 Conclusões das Simulações

Baseado nas simulações do modelo $MGPD_2$, foi mostrado que a estimação pelo modelo $MGPD_k$, com $k > 1$ é mais eficiente em relação ao modelo MG_k e $MGPD_1$ pelos seguintes aspectos:

- a estimação da cauda apresentou valores mais próximos à verdadeira densidade, tendo vantagem em relação aos outros métodos, principalmente nos valores próximos do limiar, onde a curva da densidade preditiva foi mais próxima à verdadeira;

Figura 3.24: Relação dos parâmetros da cauda para $\sigma = 3$.
Primeira linha: $\xi = 0,4$. Segunda linha: $\xi = -0,4$. Primeira coluna: $n = 1000$. Segunda coluna: $n = 10000$. As linhas cheias são os valores verdadeiros e as tracejadas são as médias a posteriori.



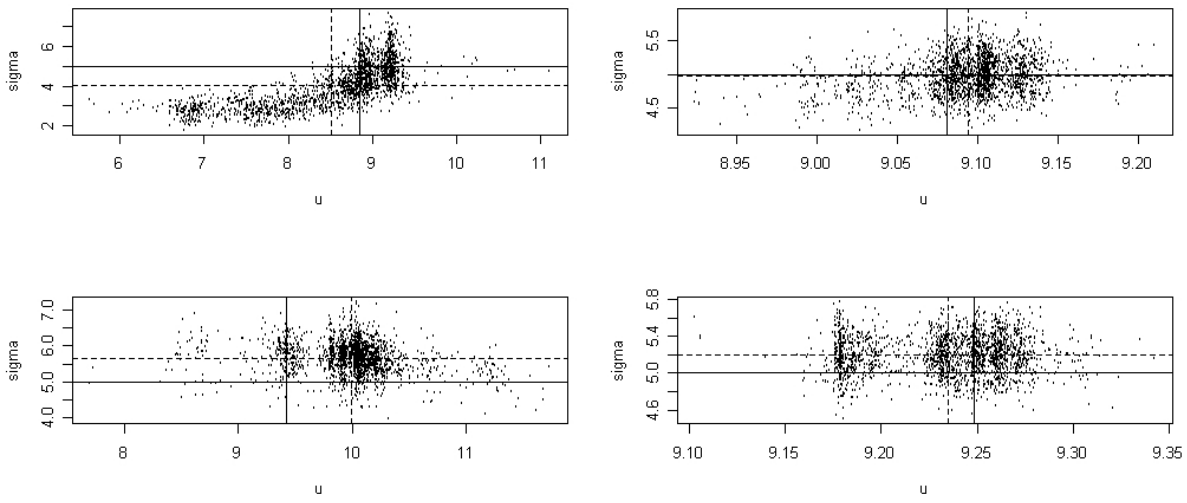
- analisando as medidas de ajuste, foi o modelo em que, na maioria das simulações, apresentou os menores valores de BIC e DIC;

- o BIC e DIC mostraram ser bons métodos para comparar modelos e detectar o número correto de componentes do modelo $MGPD_k$, pois na maioria das simulações estes critérios escolheram como melhor modelo exatamente o $MGPD_2$, que foi o utilizado para fazer a simulação.

- O modelo $MGPD_k$ apresentou melhores estimativas dos quantis altos na grande maioria das situações, sendo superior as estimativas obtidas pelo método empírico e às obtidas pelos modelos MG_k;

- Em relação aos modelos $MGPD_k$, embora exista correlação entre o limiar e o parâmetro da cauda σ, as restrições impostas nos parâmetros fazem com que na maioria das simulações os parâmetros sejam identificáveis, tornando precisa a es-

Figura 3.25: Relação dos parâmetros da cauda para $\sigma = 5$.
Primeira linha: $\xi = 0,4$. Segunda linha: $\xi = -0,4$. Primeira coluna: $n = 1000$. Segunda coluna: $n = 10000$. As linhas cheias são os valores verdadeiros e as tracejadas são as médias a posteriori.



timação dos dois parâmetros e próximas aos valores verdadeiros. Impor uma priori muito informativa resolveria o problema de não-identificabilidade, porém como um dos principais desafios deste trabalho é a estimação do limiar, é necessário encontrar uma distribuição do limiar com menor informação possível, sem prejudicar a identificação dos parâmetros do modelo.

3.6 Aplicações

Dados ambientais não são iid. Por exemplo, em dados de temperatura, um dia quente tem maior probabilidade de ser seguido por outro muito quente do que um dia frio ser seguido de um dia muito quente. Características particulares também ocorrem com nível de chuva e vazão de rios. Para isso, Coles (2001) (capítulo 5, pág 99) sugere algumas alternativas como, por exemplo, agrupar dados e observar o máximo nestes grupos. Nas

análises de dados ambientais deste trabalho, os dados analisados são máximos, quinzenais ou mensais, e não os dados diários.

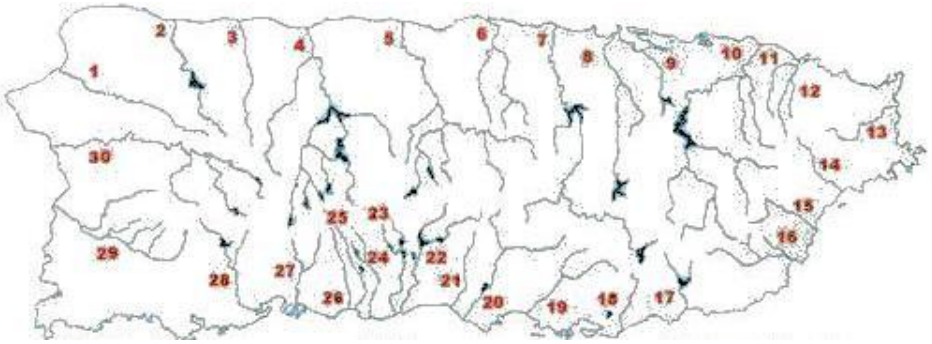
3.6.1 Aplicação 1 - Vazão de rios em Porto Rico

Este conjunto de dados consiste em máximos quinzenais de vazão de dois rios de Porto Rico: Fajardo e Espírito Santo. Os dados foram coletados no período de abril de 1967 a setembro de 2002.

A Figura 3.26 mostra a bacia hidrográfica de Porto Rico. Os dados foram estimados considerando uma abordagem não-paramétrica para todas as observações, pela aproximação pelo modelo MG_k, e pelo modelo $MGPD_k$.

Figura 3.26: Bacia Hidrográfica de Porto Rico.

12-Espírito Santo. 13- Fajardo



Foram realizadas 150000 iterações, nas quais as primeiras 100000 foram utilizadas para *burn-in*. Devido a maior dificuldade de dar um chute inicial adequado para as aplicações, foi necessário utilizar um *burn-in* maior do que nas simulações. Foram tomadas uma a cada 25 iterações e, como houve maior dificuldade de convergência nas aplicações do que nas simulações, a taxa de aceitação foi calibrada em 0,15 ao invés de 0,44, como sugerido no algoritmo de Roberts e Rosenthal (2006).

Para os parâmetros da cauda, a priori dada foi uma Normal, onde a média é o quantil 90 e a variância não pode ser muito vaga. Nas simulações, a variância da priori foi fixada

em 2. Nesta aplicação, como a amplitude dos das observações é cerca de 100 vezes maior do que na simulação, a variância da priori do limiar foi tomada como sendo 200, dando probabilidade de 95% entre os quantis 0,88 e 0,92.

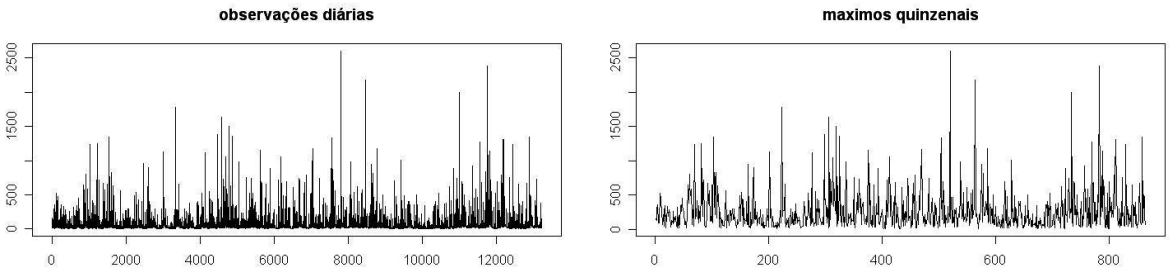
Assim como nas simulações, nestas aplicações, quando se aumentou a variância da priori para 20000, o parâmetro convergiu para uma distribuição com média e mediana muito baixas, com um percentil menor do que 10% das observações. Por isso, é necessário ter cautela na escolha da variância do limiar, não podendo esta priori ser completamente vaga. Por isso, a variância do limiar não pode ser muito maior do que 200.

Foi analisada a densidade preditiva para os modelos MG_k e $MGPD_k$ contra o histograma dos dados. Para fazer a comparação dos modelos, foram adotados os dois critérios de comparação vistos no nas seções anteriores, BIC e DIC.

Rio Espírito Santo

A Figura 3.27 mostra a série temporal diária e a série dos máximos quinzenais da vazão ao longo dos 35 anos. No total, há 844 observações.

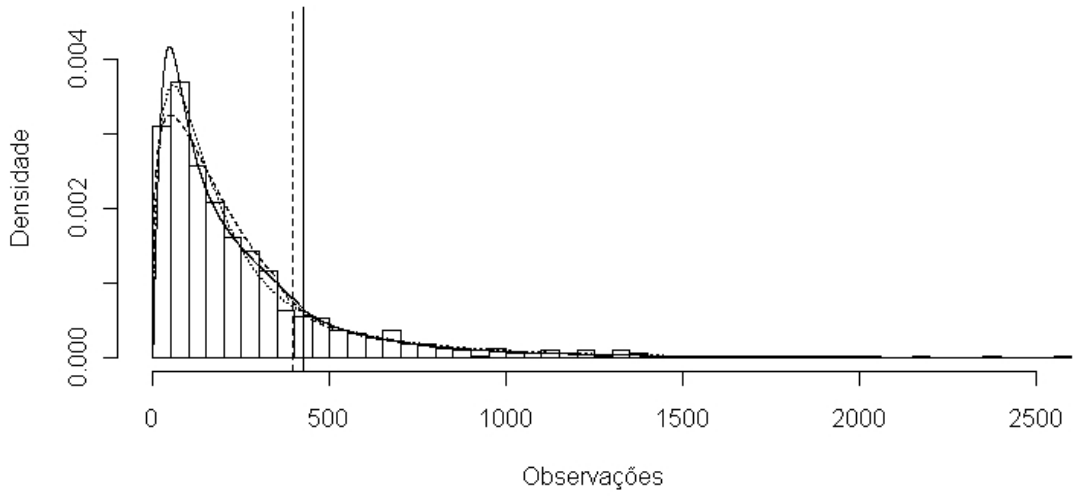
Figura 3.27: Observações do rio Espírito Santo.



Utilizando estas observações, foi feita a estimação dos parâmetros via MCMC, considerando os modelos MG_k e $MGPD_k$. A Figura 3.28 mostra o histograma das observações e a densidade preditiva para cada caso.

Quando utilizado o modelo $MGPD_k$, as densidades preditivas apresentaram resultados

Figura 3.28: Histograma e densidades preditivas os dados do Rio Espírito Santo. Linha cheia: $MGPD_3$, Linha tracejada: $MGPD_1$, Linha pontilhada: MG_2, Linhas verticais: Média a posteriori do limiar.



similares para $k = 2$ ou $k = 3$, com a curvas sendo praticamente as mesmas. A Figura 3.29 mostra que, quando é feita a modelagem com 3 componentes, um dos pesos vai para 0 e os outros dois convergem para as mesmas distribuições do que os pesos quando é feita a modelagem com duas componentes.

Pelos histogramas da distribuição a posteriori (Figura 3.30), percebe-se que o limiar parece apresentar uma distribuição bimodal, ou seja, o modelo detecta mudança de comportamento da densidade em dois valores diferentes, e pelo que foi verificado nas simulações, este é um comportamento que ocorre algumas vezes na distribuição do limiar. Um outro fator interessante neste gráfico é que os parâmetros da cauda apresentam distribuições muito parecidas, independente do número de componentes.

Para fazer uma comparação entre os modelos utilizados, além da distribuição preditiva, foram calculadas medidas de ajuste. Pela Tabela 3.5, quando é utilizado o DIC como medida de ajuste, o modelo que se ajusta melhor a estes dados foi o modelo $MGPD_3$.

Figura 3.29: Posteriori dos pesos para o modelo $MGPD_k$ para os dados do rio Espírito Santo.

A primeira linha é o modelo $MGPD_2$ e a segunda o $MGPD_3$.

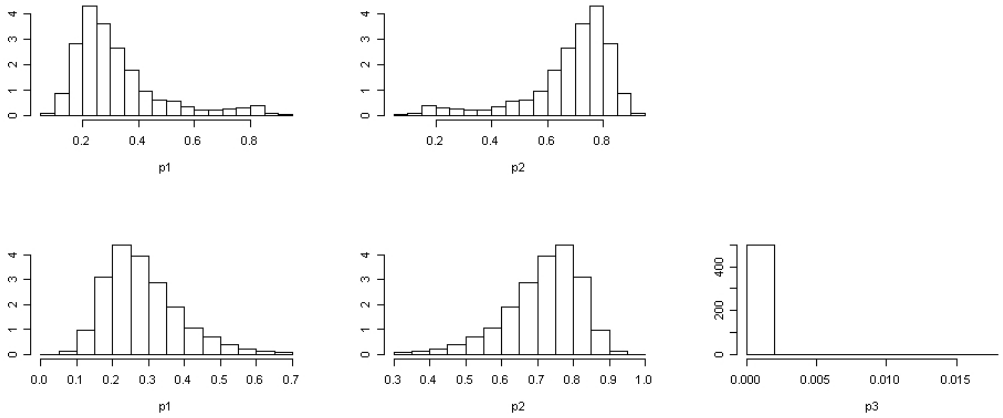
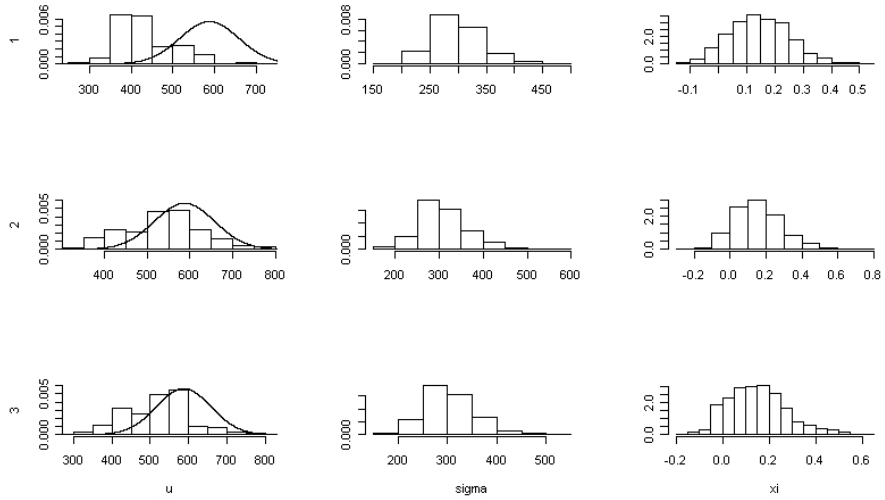


Figura 3.30: Histograma da posteriori dos parâmetros da cauda do rio Espírito Santo. As linhas representam o número de componentes k=1, 2 e 3, e as colunas são os parâmetros da cauda u, σ e ξ, nesta ordem. A curva na primeira coluna é a distribuição a priori do limiar.



Agora, em relação ao BIC, o melhor modelo foi o MG_2.

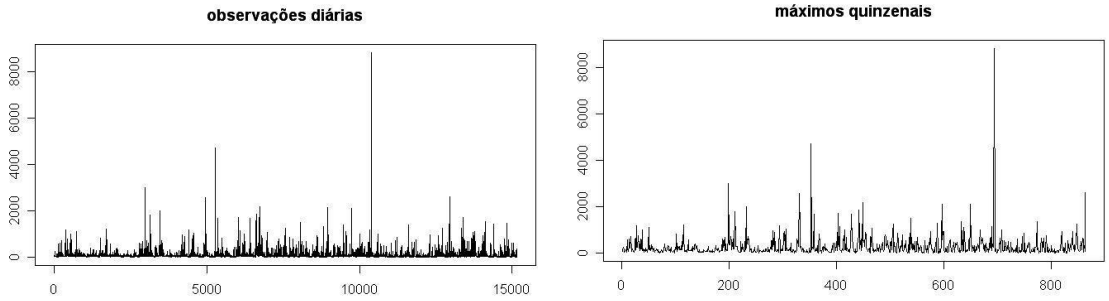
Tabela 3.5: Medidas de ajuste para os dados do rio Espírito Santo.

Modelo	MG_1	MG_2	MG_3	MG_4	MG_5	$MGPD_1$	$MGPD_2$	$MGPD_3$
Pd	2,12	3,87	3,86	4,68	11,04	0,83	4,11	6,71
DIC	11330	11275	11275	11273	11288	11299	11264	**11255**
BIC	11353	**11322**	11342	11362	11410	11336	11329	11349

Rio Fajardo

A Figura 3.31 mostra a série temporal quinzenal da vazão do Rio Fajardo ao longo de 35 anos. A Figura 3.32 mostra a densidade preditiva para cada modelo estudado.

Figura 3.31: Observações do rio Fajardo.



Quando é utilizado o modelo $MGPD_k$, há uma diferença da densidade de $k = 3$ em comparação a $k = 1$ e $k = 2$, que apresentaram densidades muito próximas. Observando os pesos do modelo $MGPD_k$, quando $k = 2$, o modelo só identifica uma componente com peso significante, mas quando são utilizadas três componentes, o modelo consegue identificar duas componentes significativas, como mostra a Figura 3.33.

Quando é feita a estimação considerando o modelo $MGPD_k$, percebe-se pelos histogramas da distribuição a posteriori (Figura 3.34) que a estimação dos parâmetros da

Figura 3.32: Histograma e densidades preditivas para os dados do Rio Fajardo. Linha cheia: $MGPD_3$, Linha tracejada: $MGPD_1$, Linha pontilhada: MG_5, Linhas verticais: Média a posteriori do limiar.

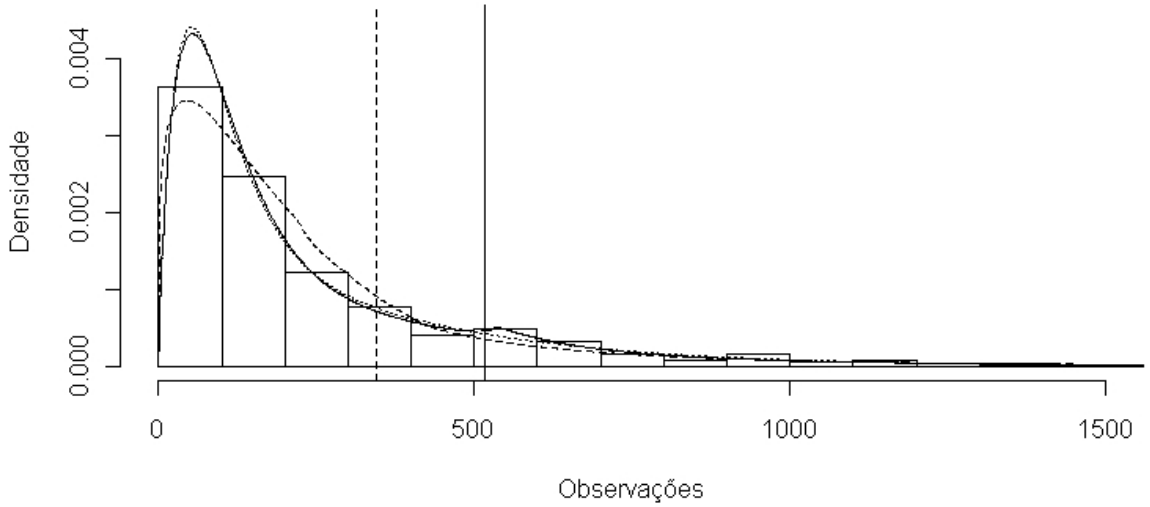
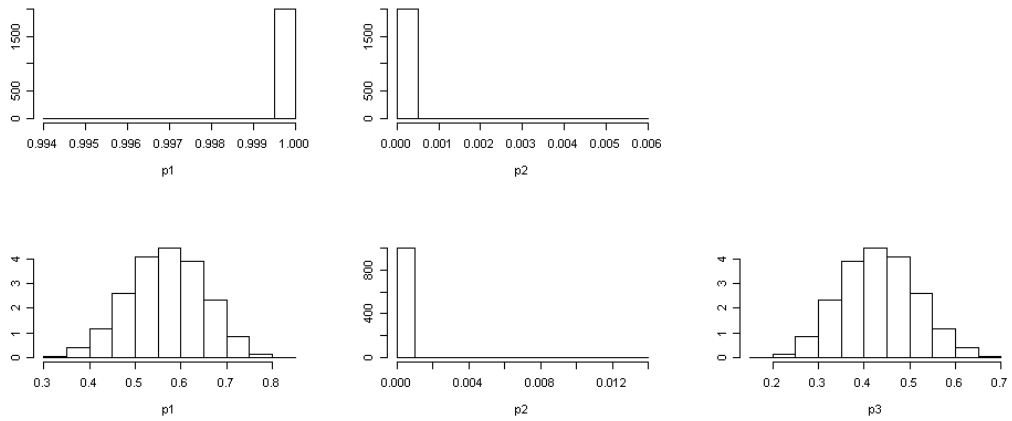
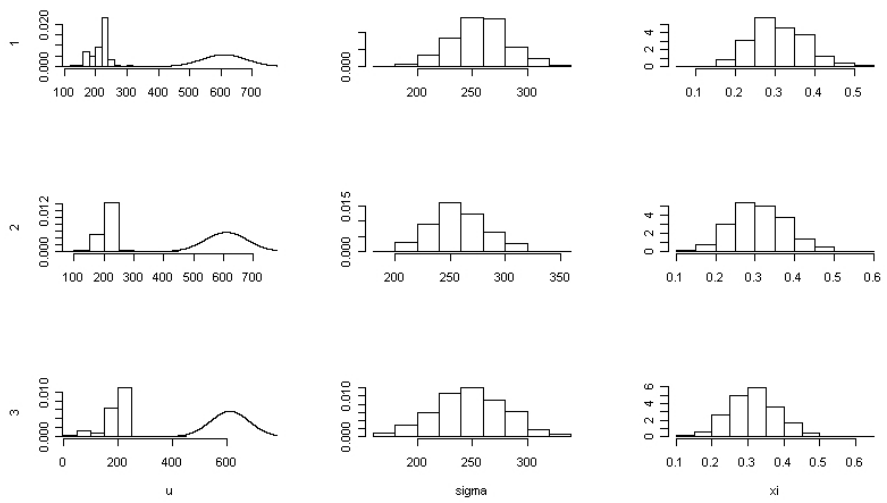


Figura 3.33: Posteriori dos pesos para o modelo $MGPD_k$ para os dados do rio Fajardo. A primeira linha é o modelo $MGPD_2$ e a segunda o $MGPD_3$.



cauda é parecida quando $k = 1$ ou $k = 2$. Entretanto, com $k = 3$, como a densidade da parte central dos dados antes do limiar tem outra forma, há uma diferença na estimação do valor do limiar, e isto altera a estimação também de σ e ξ.

Figura 3.34: Histograma da posteriori dos parâmetros da cauda do rio Fajardo.
As linhas representam o número de componentes k=1, 2 e 3, e as colunas são os parâmetros da cauda u, σ e ξ, nesta ordem. A curva na primeira coluna é a distribuição a priori do limiar.



A Tabela 3.6 mostra a comparação dos pelos métodos BIC e DIC. Pela Tabela 3.6, quando é utilizado o DIC como medida de ajuste, o modelo que se ajusta melhor a estes dados foi o $MGPD_3$. Considerando o BIC como medida de ajuste, os modelos com melhores resultados foram os modelos MG_2 e $MGPD_3$.

3.6.2 Aplicação 2 - Níveis de chuva em regiões de Portugal

Foram coletados dados de níveis pluviométricos em Portugal, que consistem em dados diários em duas estações. A Figura 3.35 mostra o nível médio pluviométrico desde 1940 a 1997. Para realizar a análise, foram selecionadas duas estações, de Barcelos, ao norte, e de Grândola, ao sul do País. Percebe-se pelo mapa que as duas estações representam

Tabela 3.6: Medidas de ajuste para os dados do rio Fajardo.

Modelo	MG_1	MG_2	MG_3	MG_4	MG_5	$MGPD_1$	$MGPD_2$	$MGPD_3$
Pd	2,01	4,65	4,53	9,31	7,22	1,05	0,53	6,28
DIC	11441	11310	11307	11293	11265	11327	11327	**11264**
BIC	11463	**11355**	11371	11391	11382	11361	11380	11357

níveis de chuvas diferentes, sendo que a estação de Barcelos historicamente recebe nível de chuva maiores que a estação de Grândola. Na estação de Barcelos, foram coletados dados diários de 1/1/1932 a 1/5/2008. Em Grândola, os dados são de 9/11/1931 a 10/05/2008. Para realizar a análise, foram tomados os máximos mensais para cada estação.

Assim como nos dados da aplicação anterior, o limiar foi tomado como tendo uma distribuição a priori Normal, com média no percentil 90 da amostra e variância 20, dando uma probabilidade de 95% entre os quantis 0,82 a 0,95.

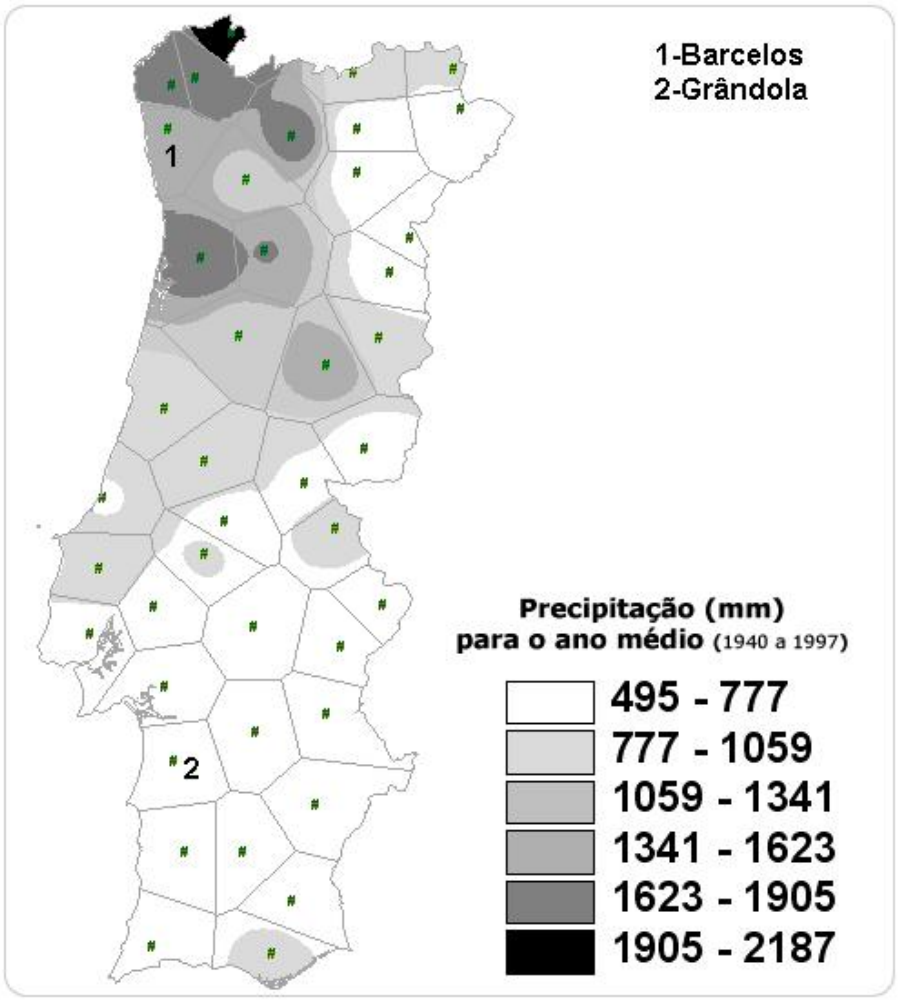
Estação Barcelos

A Figura 3.36 mostra o histograma e a densidade preditiva para cada modelo na estação de Barcelos. Pela Figura, percebe-se que, fazendo a estimação pelo modelo $MGPD_k$ com variância do limiar em 10, para $k = 1$, o valor encontrado da média do limiar foi muito menor em relação ao modelo com $k = 2$ ou $k = 3$. Isto acarretou uma estimativa da densidade preditiva muito longe do histograma dos dados para a parte central dos dados antes do limiar, e mostra a necessidade de mais de uma componente na mistura antes da cauda.

O resultado da estimação do limiar tem influência nos outros parâmetros da cauda, como pode-se ver na Figura 3.37, onde para $k = 2$ e $k = 3$, as estimativas de σ e ξ são parecidas entre si e muito diferentes em relação a $k = 1$.

Os modelos comparados graficamente para a estação de Barcelos também foram comparados através das medidas de ajuste. Pela Tabela 3.7, observa-se que o modelo com

Figura 3.35: Mapa de Portugal com índices pluviométricos.

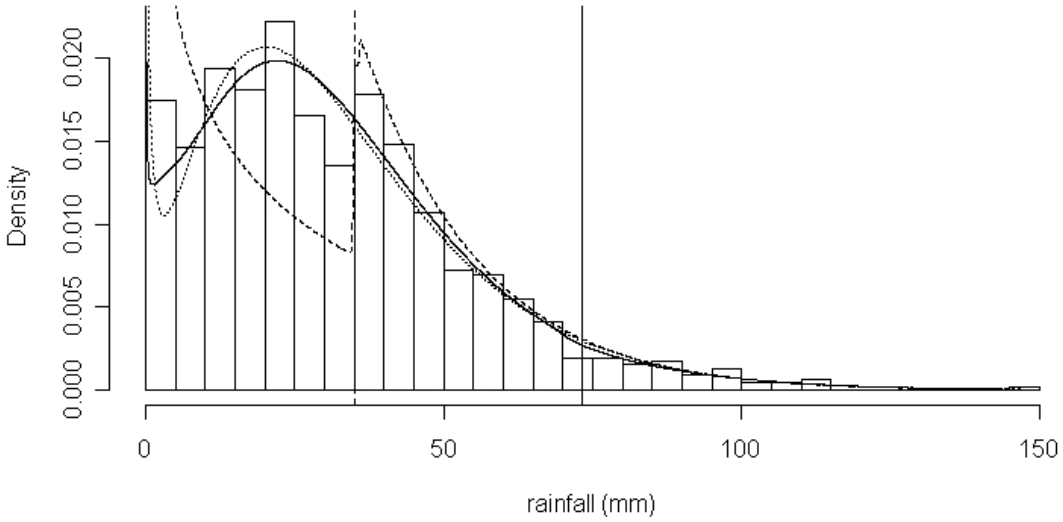


menor BIC e DIC é o modelo $MGPD_3$.

Tabela 3.7: Medidas de ajuste para a estação de Barcelos.

Modelo	MG_1	MG_2	MG_3	MG_4	MG_5	$MGPD_1$	$MGPD_2$	$MGPD_3$
Pd	2,02	3,94	4,50	3,88	128,48	5,57	5,76	7,08
DIC	8149	7951	7931	7933	8331	7998	7639	**7612**
BIC	8172	8001	8003	8022	8044	8011	7712	**7709**

Figura 3.36: Histograma com densidades preditivas da estação de Barcelos.
Linha cheia: $MGPD_3$, Linha tracejada: $MGPD_1$, Linha pontilhada: MG_3, Linhas verticais: Média a posteriori do limiar.



Estação Grândola

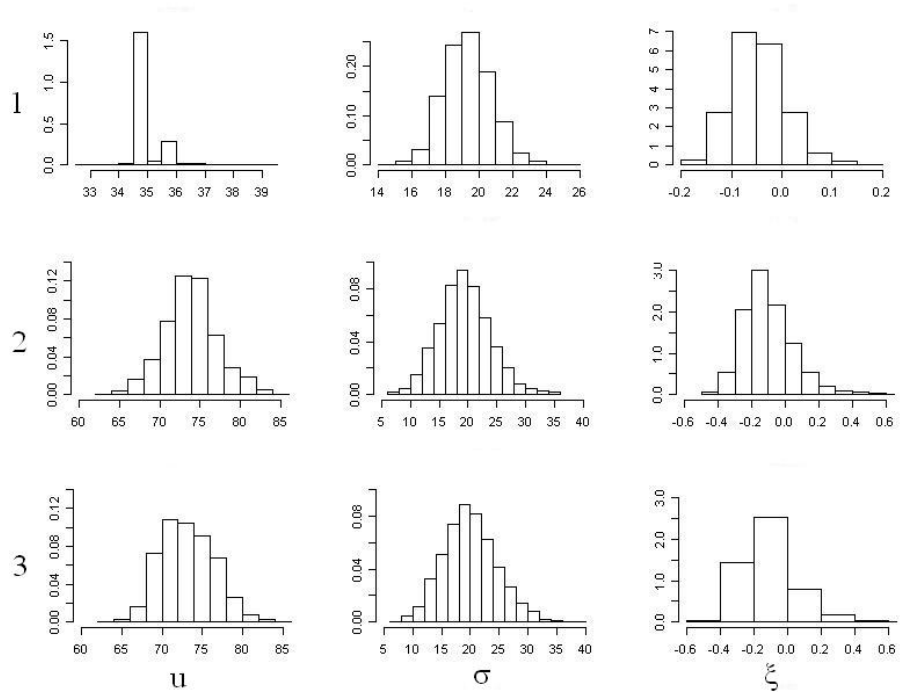
A Figura 3.38 mostra o histograma dos dados com a distribuição preditiva dos diversos modelos analisados.

Tabela 3.8: Medidas de ajuste para a estação de Grândola.

Modelo	MG_1	MG_2	MG_3	MG_4	MG_5	$MGPD_1$	$MGPD_2$	$MGPD_3$
Pd	1,98	4,21	4,18	4,26	4,05	0,20	5,31	7,01
DIC	6676	6391	6391	6391	6391	4491	4325	**4304**
BIC	6701	6442	6462	6483	6503	4522	**4397**	4400

Pela Tabela, observa-se que o menor DIC foi do modelo $MGPD_3$, enquanto que o menor BIC foi do modelo $MGPD_2$. Nesta aplicação, em todas as situações as medidas

Figura 3.37: Histograma da posteriori dos parâmetros da cauda para Barcelos.
As linhas representam o número de componentes k=1, 2 e 3, e as colunas são os parâmetros da cauda u, σ e ξ, nesta ordem.



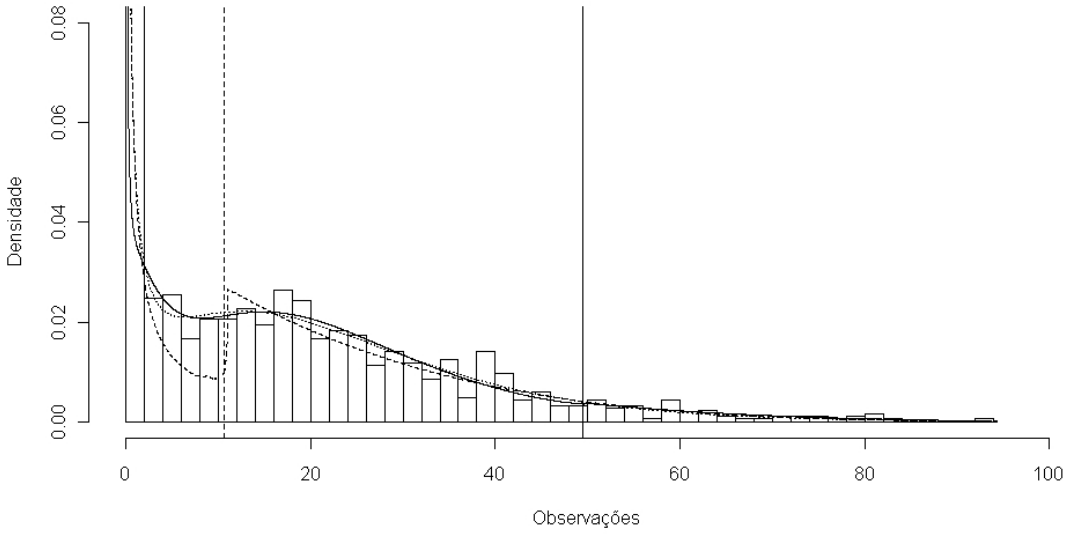
de ajuste do modelo $MGPD_k$ sempre foram melhores que os do modelo MG_k.

3.6.3 Quantis das aplicações

A determinação precisa de quantis altos é um dos principais interesses na análise de dados extremos. Estes quantis foram avaliados em ambos os conjuntos de dados. A ilustração desta análise para o Rio Espírito Santo e a Estação de Barcelos estão na Tabela 3.9.

De acordo com o modelo escolhido, no rio Espírito Santo, a quantidade 2718 ft^3/s ocorre em média com probabilidade de $0,1\%$ ou cerca de uma vez a cada 10 anos. A Figura 3.40 mostra o histograma da distribuição a posteriori do quantil $99,9\%$, denotado

Figura 3.38: Histograma com densidades preditivas da estação de Grândola.
Linha cheia: $MGPD_3$, Linha tracejada: $MGPD_1$, Linha pontilhada: MG_2, Linhas verticais: Média a posteriori do limiar.



por $q_{x;0,999}$. A distribuição é concentrada em torno do quantil empírico dos dados.

Similarmente, níveis de chuva na estação de Barcelos ultrapassa 185 mm com probabilidade $0,01\%$, ou seja, uma vez a cada século. A Figura 3.41 mostra o histograma da distribuição a posteriori para o quantil $99,99\%$, denotado por $q_{x;0,9999}$. A distribuição também é concentrada ao redor do quantil empírico.

3.6.4 Conclusões das aplicações

Após realizar a estimação para os dois conjuntos de dados, pode-se tirar as seguintes conclusões:

- há um ganho em modelar a cauda tendo distribuição GPD, ao invés de considerar somente o modelo MG_k. Aumentar o número de componentes no modelo MG_k não torna a estimação da cauda tão eficiente quanto no modelo $MGPD_k$. Isto pode ser

Figura 3.39: Histograma da posteriori dos parâmetros da cauda de Grândola.
As linhas representam o número de componentes k=1, 2 e 3, e as colunas são os parâmetros da cauda u, σ e ξ, nesta ordem.

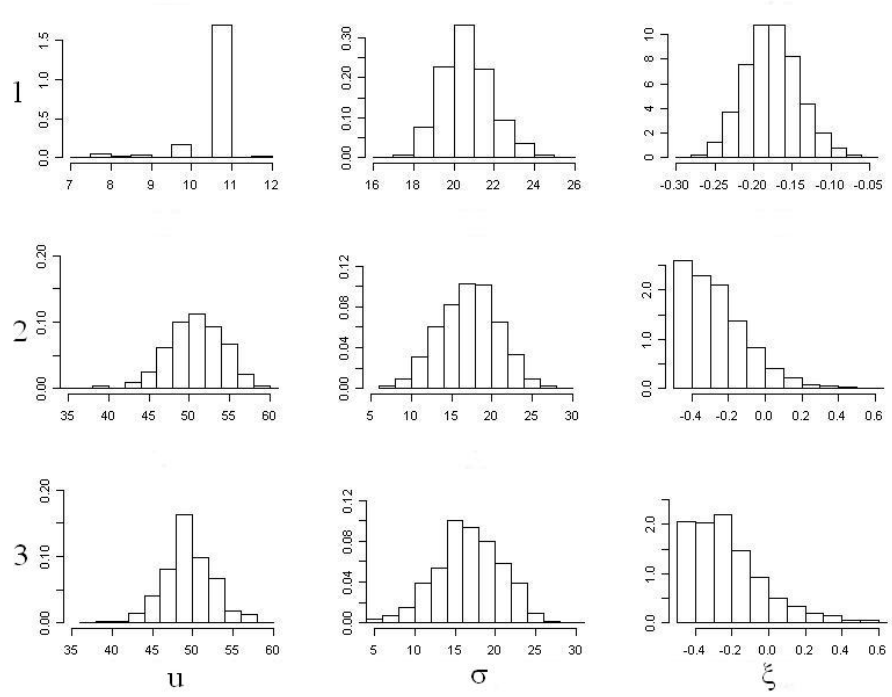


Tabela 3.9: Quantis altos do Rio Espírito Santo e da Estação de Barcelos

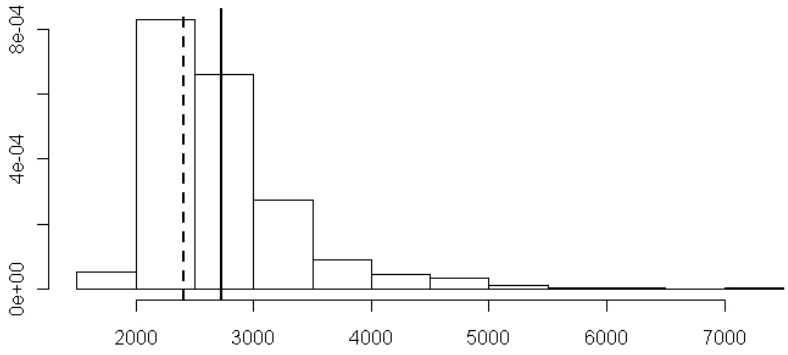
	Espírito Santo (in ft^3/s)					Barcelos (em mm)				
Prob	E	1	2	3	MG_k	E	1	2	3	MG_k
0,95	798	793,29	813,52	807,54	842,7	73,1	74,54	77,54	73,29	74,71
0,99	1360	1426,04	1450,79	1443,85	1398,8	99,4	101,73	105,38	102,24	104,09
0,999	2600	2677,56	2726,55	2718,29	2197,0	117,5	137,84	139,91	137,77	151,50
0,9999	N/A	4612,30	4734,16	4710,35	3014,0	143,5	171,41	176,12	184,54	233,00

Prob=P(X$\leq q$), $E = Empírico$, 1 =MGPD$_1$, 2 = $MGPD_2$, 3 = $MGPD_3$

MG_k se refere ao melhor modelo nesta classe segundo o DIC.

Figura 3.40: Histograma da distribuição a posteriori do quantil $99,9\%$ do Rio Espírito Santo utilizando o modelo $MGPD_3$.

Linhas verticais: cheia - média a posteriori; tracejada - empírica. O desvio padrão da distribuição a posteriori do quantil $99,9\%$ é de 560,157 ft^3/s.



visto observando as figuras das densidades preditivas e as tabelas com as medidas BIC e DIC;

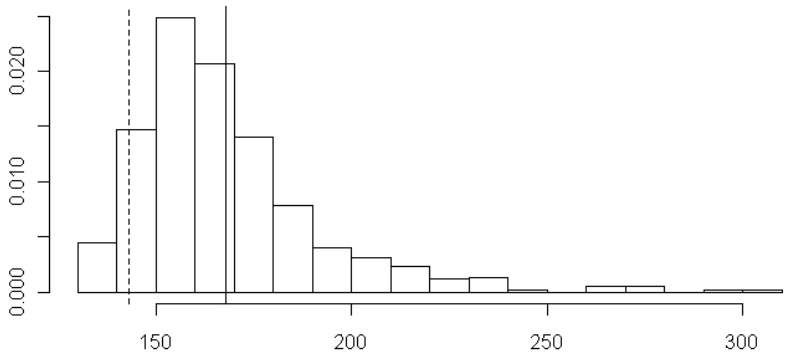
- na maioria das aplicações, os valores do BIC e DIC foram melhores quando o modelo escolhido foi o $MGPD_k$, com $k > 1$. Isto significa que, nestes casos, há a necessidade de modelar as observações abaixo da cauda por uma mistura de mais de uma Gama;

- a estimação é sensível à escolha da variância a priori do limiar, ou seja, para modelar somente a cauda após um limiar, sendo este um quantil alto dos dados, é necessário considerar como distribuição a priori uma Normal com média sendo um quantil alto da distribuição e variância relativamente pequena;

- os quantis do modelo $MGPD_k$ fornece valores mais próximos aos quantis empíricos do que o modelo MG_k, que mostrou-se pouco eficiente para estimação da cauda.

Figura 3.41: Histograma da distribuição a posteriori do quantil $99,99\%$ da estação de Barcelos, utilizando o modelo $MGPD_3$.

Linhas verticais: cheia - média a posteriori; tracejada - empírica. O desvio padrão da distribuição a posteriori do quantil $99,9\%$ é de 22,51 mm.



Apêndice 3 - Algoritmo

O Algoritmo para aplicar a técnica MCMC no modelo $MGPD_k$ é dado a seguir

Algoritmo 2

No passo s da iteração, os parâmetros são atualizados da seguinte maneira:

- Amostrando ξ: A distribuição proposta para ξ possui distribuição Normal truncada

$$\xi^* \mid \xi^{(s)} \sim N(\xi^{(s)}, V_\xi) I(-\sigma^{(s)}(M - u^{(s)}), \infty),$$

onde V_ξ é a variância da distribuição proposta e $M = \max(x_1, \ldots, x_n)$. Assim, $\xi^{(j+1)} = \xi^*$ com probabilidade α_ξ, onde

$$\alpha_\xi = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x})\Phi((\xi^{(s)} + \sigma^{(s)}/(M - u^{(s)})/\sqrt{V_\xi}))}{\pi(\tilde{\Theta}|\mathbf{x})\Phi((\xi^* + \sigma^{(s)}/(M - u^{(s)})/\sqrt{V_\xi}))}\right\},$$

$\Theta^* = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^{(s)}, \sigma^{(s)}, \xi^*)$ e $\tilde{\Theta} = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^{(s)}, \sigma^{(s)}, \xi^{(s)})$.

- Amostrando σ

  Se $\xi^{(s+1)} > 0$, então σ^* é amostrado de uma distribuição $G(\sigma^{(s)}, \sigma^{(s)^2}/V_\sigma)$, onde V_σ é a variância da distribuição proposta. Se $\xi^{(s+1)} < 0$, então σ^* é amostrado de uma $N(\sigma^{(s)}, V_\sigma)I(-\xi^{(s+1)}(M - u^{(s)}), \infty)$

  Além disso, $\sigma^{(s+1)} = \sigma^*$ com probabilidade α_σ onde, se $\xi^{(s+1)} < 0$

$$\alpha_\sigma = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x})\Phi((\sigma^{(s)} + \xi^{(s+1)}(M - u^{(s)})/\sqrt{V_\sigma}))}{\pi(\tilde{\Theta}|\mathbf{x})\Phi((\sigma^* + \xi^{(s+1)}(M - u^{(s)})/\sqrt{V_\sigma}))}\right\},$$

  e se $\xi^{(s+1)} > 0$

$$\alpha_\sigma = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x}) f_G(\sigma^{(s)}|\sigma^{(s)}, \sigma^{(s)^2}/V_\sigma)}{\pi(\tilde{\Theta}|\mathbf{x}) f_G(\sigma^*|\sigma^*, \sigma^{*2}/V_\sigma)}, \right\}$$

  onde $\Theta^* = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^{(s)}, \sigma^*, \xi^{(s+1)})$ e $\tilde{\Theta} = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^{(s)}, \sigma^{(s)}, \xi^{(s+1)})$.

- Amostrando u

  O valor do limiar u^* é amostrado de uma distribuição $N(u^{(s)}, V_u)I(a^{(s+1)}, \infty)$, onde $a^{(s+1)} = \min(x_1, \ldots, x_n)$ se $\xi^{(s+1)} \geq 0$ e se $\xi^{(s+1)} < 0$, $a^{(s+1)} = M + \sigma^{(s+1)}/\xi^{(s+1)}$. Estes valores são escolhidos para satisfazerem o espaço amostral onde a distribuição GPD em (3.2) está definida. V_u é a variância da distribuição proposta para o limiar. Aceita-se $u^{(s+1)} = u^*$ com probabilidade α_u, onde

$$\alpha_u = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x})\Phi((u^{(s)} - a^{(s+1)})/\sqrt{V_u})}{\pi(\tilde{\Theta}|\mathbf{x})\Phi((u^* - a^{(s+1)})/\sqrt{V_u})}\right\},$$

  $\Theta^* = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^*, \sigma^{(s+1)}, \xi^{(s+1)})$ e $\tilde{\Theta} = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^{(s)}, \sigma^{(s+1)}, \xi^{(s+1)})$.

- Amostrando $\mu, \eta, \mathbf{p}$

  Para amostrar os vetores μ, η e $\mathbf{p}$, o procedimento é igual ao Algoritmo 1 do Capítulo 2. Porém, o que muda no algoritmo deste capítulo é a distribuição da posteriori e utiliza-se os valores atualizados $\xi^{(s+1)}, \sigma^{(s+1)}, u^{(s+1)}$ dos parâmetros da cauda da distribuição. As amostragens são feitas da seguinte maneira

a) Amostrando μ_j e η_j

μ_j^* e η_j^* são gerados utilizando as mesmas distribuições propostas no Algoritmo 1. Os valores $\eta_j^{(s+1)} = \eta_j^*$ e $\mu_j^{(s+1)} = \mu_j^*$ são aceitos com probabilidade α_{μ_j,η_j} onde

$$\alpha_{\mu_j,\eta_j} = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x}) f_G(\mu_j^{(s)}|\mu_j^*, \mu_j^*/V_\mu) f_G(\eta_j^{(s)}|\eta_j^*, \eta_j^*/V_\eta) I_R(\underline{\mu})}{\pi(\tilde{\Theta}|\mathbf{x}) f_G(\mu_j^*|\mu_j^{(s)}, \mu_j^{(s)}/V_\mu) f_G(\eta_j^*|\eta_j^{(s)}, \eta_j^{(s)}/V_\eta) I_R(\underline{\mu})}\right\},$$

$\Theta^* = (\eta_{<j}^{(s+1)}, \eta_j^*, \eta_{>j}^{(s)}, \mu_{<j}^{(s+1)}, \mu_j^*, \mu_{>j}^{(s)}, p^{(s)}, u^{(s+1)}, \sigma^{(s+1)}, \xi^{(s+1)})$ e

$\tilde{\Theta} = (\eta_{<j}^{(s+1)}, \eta_{\geq j}^{(s)}, \mu_{<j}^{(s+1)}, \mu_{\geq j}^{(s)}, p^{(s)}, u^{(s+1)}, \sigma^{(s+1)}, \xi^{(s+1)})$.

b) Amostrando $\mathbf{p}$.

Assim como no Algoritmo 1, amostra-se $\mathbf{p}^* \sim D_k(V_p p_1^{(s)}, \ldots, V_p p_k^{(s)})$. Assim, $\mathbf{p}^{(s+1)} = \mathbf{p}^*$ com probabilidade

$$\alpha_p = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x}) f_D(\mathbf{p}^{(s)}|\mathbf{p}^*)}{\pi(\tilde{\Theta}|\mathbf{x}) f_D(\mathbf{p}^*|\mathbf{p}^{(s)})}\right\},$$

onde $\Theta^* = (\eta^{(s+1)}, \mu^{(s+1)}, \mathbf{p}^*, u^{(s+1)}, \sigma^{(s+1)}, \xi^{(s+1)})$ e $\tilde{\Theta} = (\eta^{(s+1)}, \mu^{(s+1)}, \mathbf{p}^{(s)}, u^{(s+1)}, \sigma^{(s+1)}, \xi^{(s+1)})$.

Capítulo 4

Extremos com estrutura de modelos de regressão

O comportamento da distribuição de variáveis pode estar ligado também à presença de covariáveis. Em dados de chuva, por exemplo, uma covariável sazonal, que pode ser indicadora de meses ou estações do ano pode ajudar a explicar o nível de chuva de uma região. Cabras et al. (2009) citam um exemplo onde um dos parâmetros da variável GPD está relacionado linearmente com a covariável binária onde 1 representa verão e 0 inverno. Em dados de finanças, por exemplo na variável cotação de uma moeda, a covariável taxa de juros ou índice da bolsa de valores do país pode estar diretamente correlacionada com a cotação da moeda.

Um outro fator que pode influenciar no comportamento da variável é a sua localização. Em dados meteorológicos, fatores como latitude, longitude, altitude e distância do mar podem influenciar nas condições meteorológicas de uma determinada região.

Baseado nesta motivação, a próxima meta deste trabalho está em realizar uma modelagem mais abrangente para os parâmetros da cauda dos dados, onde é utilizada a distribuição GPD. Neste capítulo, é considerado que os parâmetros da distribuição GPD estão correlacionados com covariáveis. Uma ideia mais simples seria utilizar uma relação linear nos parâmetros.

Cabras et al. (2009) desenvolve um trabalho onde é encontrado um modelo em que a priori de Jeffreys para o conjunto de parâmetros das regressões do preditor linear é proporcional a uma constante, propondo uma reparametrização da distribuição GPD, de tal forma que a distribuição a priori de Jeffreys dos parâmetros da cauda é Uniforme. Nesta reparametrização, o parâmetro de escala σ é substituído por $\nu = \sigma(1+\xi)$. Assim, a função de densidade da GPD pode ser reescrita como

$$g(x_i|\xi_i,\nu_i,u_i) = \begin{cases} \nu_i^{-1}(1+\xi_i)(1+(x_i-u_i)(\xi_i+\xi_i^2)/\nu_i)^{-(1+\xi_i)/\xi_i}, & \text{se} \quad \xi \neq 0 \\ \nu_i^{-1}\exp(-(x_i-u_i)/\nu_i), & \text{se} \quad \xi_i = 0 \end{cases}, \quad (4.1)$$

onde $x_i - u_i \geq 0$ para $\xi_i \geq 0$ e $(1+(x_i-u_i)(\xi_i+\xi_i^2)/\nu_i) > 0$, se $-0,5 < \xi_i < 0$.

4.1 Modelo de regressão para os parâmetros da GPD

Em relação a distribuição GPD escrita como em (4.1), cada um dos parâmetros da cauda pode ser escrito em função de funções de preditores lineares de um conjunto de covariáveis, de tal forma que $\xi_i = \xi(\beta_\xi \mathbf{z}'_{1,i})$, com $\beta_\xi = (\beta_{\xi,0},\ldots,\beta_{\xi,k_\xi})$, $\nu_i = \nu(\beta_\nu \mathbf{z}'_{2,i})$ com $\beta_\nu = (\beta_{\nu,0},\ldots,\beta_{\nu,k_\nu})$, e $u_i = u(\beta_u \mathbf{z}'_{3,i})$, com $\beta_u = (\beta_{u,0},\ldots,\beta_{u,k_u})$. Os vetores $\mathbf{z}_1$, $\mathbf{z}_2$ e $\mathbf{z}_3$ são as covariáveis de ξ, ν e u respectivamente, que podem ou não ter covariáveis em comum, e β_ξ, β_ν e β_u são os seus respectivos vetores de parâmetros.

Segundo Cabras et al.(2009), aplicando a regra de Jeffreys, as funções de ligação escolhidas para os preditores lineares são $\xi_i = \exp(\mathbf{z}'_{1,i}\beta_\xi) - 1$ e $\nu_i = \exp(\mathbf{z}'_{2,i}\beta_\nu)$, onde $\beta_\xi = (\beta_{\xi_0},...,\beta_{\xi_{k_\xi}})$ e $\beta_\nu = (\beta_{\nu_0},...,\beta_{\xi_{k_\nu}})$. Com estas funções, a distribuição a priori conjunta para (β_ξ,β_ν) é proporcional a uma distribuição Uniforme, o que facilita os cálculos na estimação destes parâmetros.

Com esta reparametrização, a densidade é dada em termos de parâmetros ortogonais ξ e ν de acordo com Chaves-Demoulin e Davison (2005). Nesta parametrização, os blocos de parâmetros β_ν e β_ξ são ortogonais e isto simplifica a derivação da priori conjunta $\pi(\beta_\xi,\beta_\nu)$.

Para o limiar, também é considerada uma estrutura de regressão, dada por $u_i =$

$u(\mathbf{z}'_{3,i}\beta_u)$, onde $\beta_u = (\beta_{u_0}, ..., \beta_{u_{k_u}})$. Foi considerada uma estrutura de regressão linear simples para o limiar $u_i = \mathbf{z}'_{3,i}\beta_u$. Uma alternativa seria utilizar o fato dos dados serem positivos, definindo $u_i = \exp(\mathbf{z}'_{3,i}\beta_u)$. Neste trabalho, o modelo foi desenvolvido utilizando a primeira abordagem para u_i.

4.2 Mistura de Gamas com GPD via modelos de regressão

Baseado na estrutura de modelos de regressão dos parâmetros da distribuição GPD, neste capítulo será proposto um modelo que engloba mistura de Gamas para a parte central das observações e distribuição GPD. A diferença em relação ao Capítulo 3 está na forma de analisar os parâmetros da cauda, que agora estão escritos como funções de covariáveis. Dessa maneira, o modelo, denominado por $MGPDR_k$, tem densidade similar a da equação (3.7), substituindo $g(x|\Psi)$ pela densidade dada em (4.1).

4.2.1 Distribuição a priori

A distribuição a priori para os parâmetros da distribuição antes da cauda, ou seja, da parte modelada por mistura de Gamas, é a mesma distribuição a priori do capítulo anterior.

Para os parâmetros da Distribuição GPD, a distribuição a priori de Jeffreys para o vetor de parâmetros $\pi(\beta_\xi, \beta_\eta \mid \beta_u)$ é dada por uma distribuição Uniforme, como mostra Cabras et al. (2009). Para o vetor de parâmetros do limiar β_u, utiliza-se uma priori com distribuição Normal $\beta_{u_0} \sim N(a_0, V_{\beta_{u_0}})$ e $\beta_{u_i} \sim N(0, V_{\beta_{u_i}})$, $i = 1, ..., k_u$, onde k_u é o número de covariáveis utilizados na estimação do limiar. O capítulo anterior mostrou algumas restrições para a estimação do limiar quando o tamanho da amostra é pequeno, não podendo utilizar uma priori vaga. Da mesma maneira, é necessário tomar cuidado na escolha de $V_{\beta_{u_i}}$, $i = 0, ..., k_u$, onde em alguns casos pode ser necessário impor uma

variância não muito alta quando o tamanho da amostra é pequeno.

4.2.2 Distribuição a posteriori

Dada a distribuição a priori e a função de densidade da distribuição de mistura, encontra-se a seguinte forma para o logaritmo da densidade a posteriori do modelo com mistura de Gamas com GPD, considerando estrutura de regressão para os parâmetros da cauda

$$\begin{aligned}\pi(\Theta|\mathbf{x}) \quad \propto \quad & \sum_{i:x_i<u_i} \log\left[\sum_{j=1}^{k} p_j f_G(x_i|\mu_j,\eta_j)\right] + \sum_{x_i\geq u_i} \log\left(1-\sum_{j=1}^{k} p_j F_G(u_i|\mu_j,\eta_j)\right) \\ + \quad & \sum_{x_i\geq u_i} \log[g(x_i|\xi_i,\nu_i,u_i)] + \sum_{j=1}^{k}\left[(a_j-1)\log(\eta_j) - b_j\eta_j - (c_j+1)\log(\mu_j) - \frac{d_j}{\mu_j}\right] \\ - \quad & \frac{(\beta_{u_0}-a_0)^2}{2V_{\beta_{u_0}}} - \sum_{i=1}^{k_u}\left(\frac{\beta_{u_i}^2}{2V_{\beta_{u_i}}}\right),\end{aligned} \tag{4.3}$$

onde $\Theta = (\mu_1,\ldots,\mu_k,\eta_1,\ldots,\eta_k,p_1,\ldots,p_k,\beta_u,\beta_\nu,\beta_\xi)$ é o vetor de parâmetros do modelo.

O suporte da priori e da verossimilhança em Θ dependem das covariáveis $\mathbf{z}$ e das restrições dos parâmetros do modelo $MGPDR_k$, onde Θ satisfaz as seguintes restrições:

$$\begin{cases} \exp(2\mathbf{z}'_{1,i}\beta_\xi) - \exp(2\mathbf{z}'_{1,i}\beta_\xi) > -x_i^{-1}\exp(\mathbf{z}'_{2,i}\beta_\nu) \\ \mathbf{z}'_{1,i}\beta_\xi > -\log(2). \end{cases} \tag{4.4}$$

A primeira restrição é imposta para garantir a existência da função de verossimilhança. A segunda restrição é necessária para a existência da Informação de Fisher (Smith, 1984). Segundo Cabras et al. (2009), estas restrições são verificadas apenas numericamente.

Para realizar a estimação dos parâmetros, são utilizadas técnicas MCMC. Para os valores menores que o limiar, a estimação utilizando mistura de distribuições é feita da mesma maneira como no Capítulo 3. Para os parâmetros da cauda, cada vetor de parâmetros β_u,

β_ν e β_ξ são estimados como sendo um bloco. O algoritmo para a estimação dos parâmetros é dado no Apêndice 4.

4.3 Simulações

Simulações do modelo proposto foram feitas em diferentes configurações de parâmetros. Através da estimação dos parâmetros e quantis baseados nestas simulações, pode-se ter uma evidência empírica da precisão do método em recuperar os verdadeiros valores dos parâmetros. Com isso, tem-se uma base para realizar a estimação pelo modelo proposto em dados reais, que será apresentado na seção a seguir.

O exercício de simulação foi realizado considerando amostras de tamanhos $n = 1000$ e $n = 10000$. Foram geradas covariáveis utilizando distribuições Uniformes $z_{1,i} \sim U(0, 2)$ e $z_{2,i} \sim U(0, 4)$, $i = 1, ..., n$. Considerando n pontos de uma distribuição de mistura de duas Gamas com os mesmos parâmetros utilizados no capítulo anterior. Para cada observação gerada, se o valor desta observação for maior que um limiar $u_i = \beta_{0,u} + \beta_{1,u} z_{1,i}$, então é gerada um ponto da distribuição GPD com densidade $g(x_i \mid \xi_i, \nu_i, u_i)$, onde $\xi_i = \exp(\beta_{0,\xi} + \beta_{1,\xi} z_{1,i}) - 1$ e $\nu_i = \exp(\beta_{0,\nu} + \beta_{1,\nu} z_{2,i})$. As simulações foram feitas com os valores $\beta_{0,u}$=6 e 9, $\beta_{1,u}$=0,5 e -0,5, $\beta_{0,\xi}$=0,2, $\beta_{1,\xi}$=0,3 e -0,3, $\beta_{0,\nu}$=3 e $\beta_{1,\nu}$=0,5.

Para $n = 1000$, $\beta_{0,u} \sim N(7, 10)$ e $\beta_{1,u} \sim N(0, 5)$. Para $n = 10000$, $\beta_{0,u} \sim N(7, 100)$ e $\beta_{1,u} \sim N(0, 100)$. Como visto no capítulo anterior, é necessário impor uma priori com maior informação quando o tamanho da amostra não é muito grande.

4.3.1 Simulações para amostra de tamanho 1000

As Figuras 4.1 e 4.2 mostram o histograma da distribuição a posteriori dos parâmetros da cauda para $\beta_{1,\xi} = 0,3$ e $\beta_{1,\xi} = -0,3$, ambas com $\beta_{1,u} = 0,5$ e $\beta_{0,u} = 6$. Pela Figura 4.1, observa-se que há uma dificuldade de estimar o efeito da covariável sobre o limiar, como mostra o histograma de $\beta_{1,u}$. Para os outros parâmetros, a estimação da distribuição parece estar próxima aos valores verdadeiros utilizados na simulação. Pela Figura 4.2,

todos os parâmetros parecem estar bem estimados em relação ao verdadeiro valor. Em ambas as figuras, o valor 0 está dentro da distribuição a posteriori com uma probabilidade não muito baixa, ou seja, em algumas situações está havendo dificuldade de dizer se realmente o efeito da covariável é significativo ou se existe intercepto.

Figura 4.1: Histograma da cauda para $n = 1000$, $\beta_{0,u} = 6$, $\beta_{1,\xi} = 0,3$ e $\beta_{1,u} = 0,5$. As linhas verticais representam os valores verdadeiros.

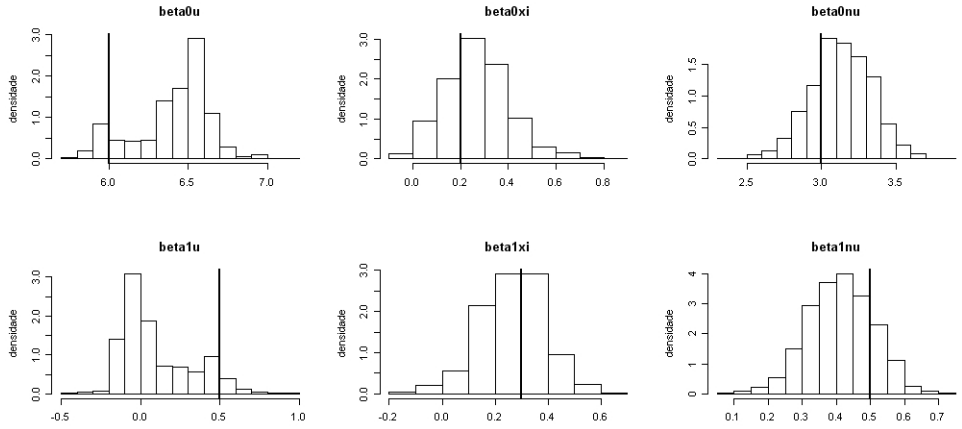
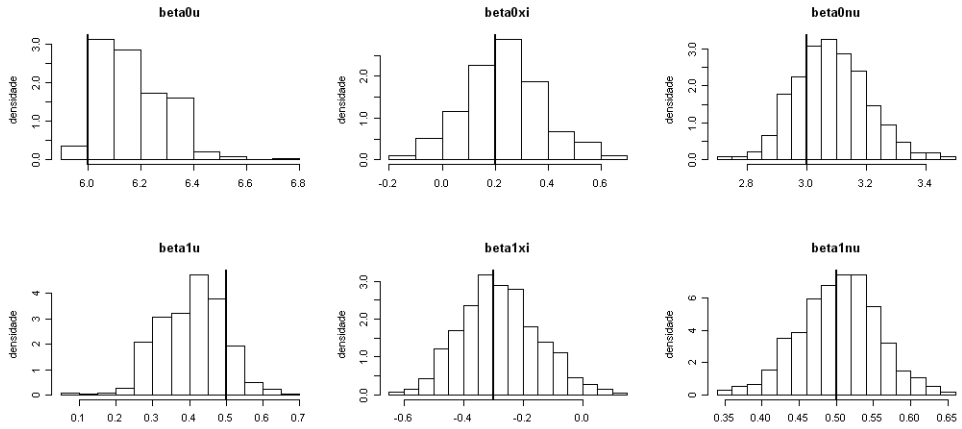


Figura 4.2: Histograma da cauda para $n = 1000$, $\beta_{0,u} = 6$, $\beta_{1,\xi} = -0,3$ e $\beta_{1,u} = 0,5$. As linhas verticais representam os valores verdadeiros.

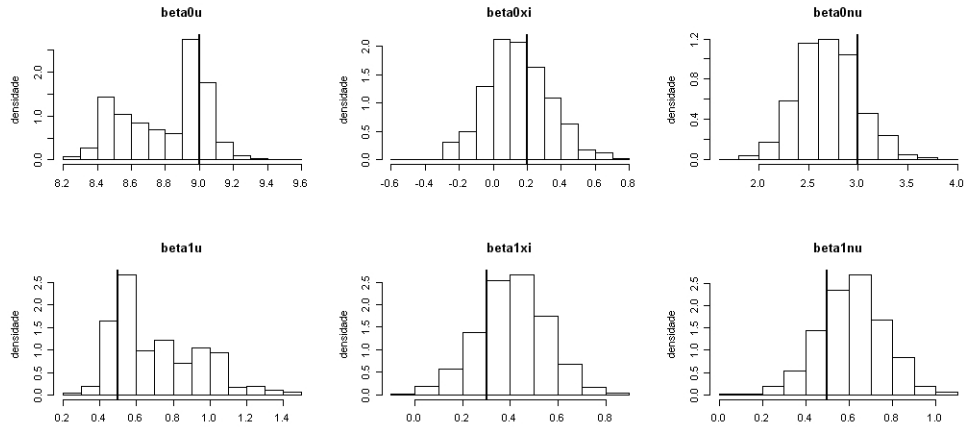


As Figuras 4.3 e 4.4 mostram o histograma da distribuição a posteriori dos parâmetros

da cauda para $\beta_{1,\xi} = 0,3$ e $\beta_{1,\xi} = -0,3$, agora com ambas as Figuras com $\beta_{1,u} = 0,5$ e $\beta_{0,u} = 9$. As figuras mostram que a estimação é eficiente para a maioria dos parâmetros da cauda, embora a amplitude da distribuição parece ser alta, não dando uma informação muito precisa sobre os parâmetros. Em comparação com as figuras onde $\beta_{0,u} = 6$, os histogramas com $\beta_{0,u} = 9$ tem uma amplitude maior para os outros parâmetros da cauda, pois com um limiar num quantil mais alto, há menos observações na cauda, o que aumenta a incerteza e diminui a precisão na estimação dos parâmetros da cauda.

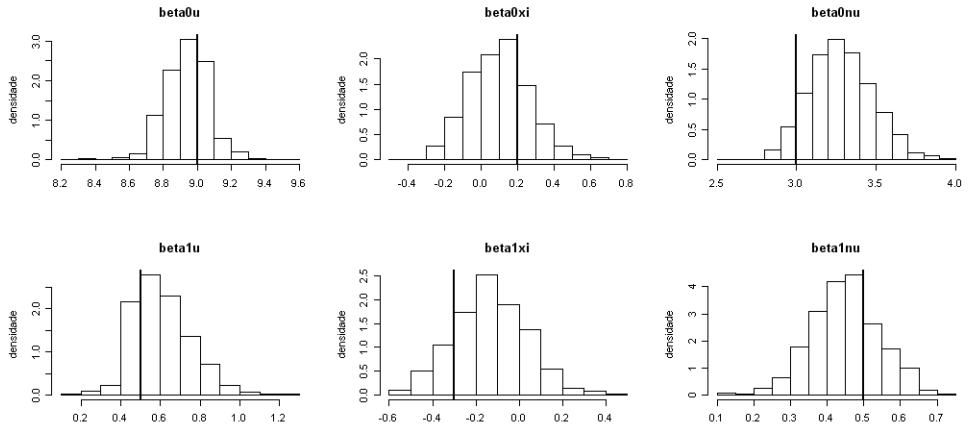
A maior dificuldade foi na estimação dos parâmetros relacionados ao parâmetro de forma ξ. Em ambas as Figuras 4.3 e 4.4, a probabilidade de $\beta_{0,\xi} = 0$ é alta, ou seja, a estimação indica que pode não haver intercepto no preditor linear de ξ. Na Figura 4.4, o efeito de $\beta_{1,\xi}$ também não parece ser significativo, ou seja, neste caso, o modelo indica que o parâmetro ξ poderia ser estimado como sendo fixo e igual a 0, quando na verdade os valores simulados foram de um modelo com $\xi \neq 0$.

Figura 4.3: Histograma da cauda para $n = 1000$, $\beta_{0,u} = 9$, $\beta_{1,\xi} = 0,3$ e $\beta_{1,u} = 0,5$. As linhas verticais representam os valores verdadeiros.



A Tabela 4.1 apresenta o intervalo de credibilidade de 95% dos parâmetros da cauda para $n = 1000$. Observando a tabela, em quase todas as estimações, o verdadeiro valor do parâmetro se encontra dentro do intervalo de credibilidade, porém a variabilidade de

Figura 4.4: Histograma da cauda para $n = 1000$, $\beta_{0,u} = 9$, $\beta_{1,\xi} = -0,3$ e $\beta_{1,u} = 0,5$. As linhas verticais representam os valores verdadeiros.



algumas estimações é muito alta, dando pouca informação sobre a distribuição a posteriori dos parâmetros. Com exceção de $\beta_{0,u}$ e $\beta_{0,\nu}$, a estimação dos parâmetros não é precisa, em alguns casos não dando para concluir se o efeito da covariável é significativo, principalmente nos parâmetros que determinam a forma da distribuição GPD $\beta_{0,\xi}$ e $\beta_{1,\xi}$, onde em quase todos os casos o valor 0 apareceu dentro do intervalo de credibilidade. Portanto, há uma restrição em apontar efeitos de covariáveis em uma amostra de tamanho $n = 1000$.

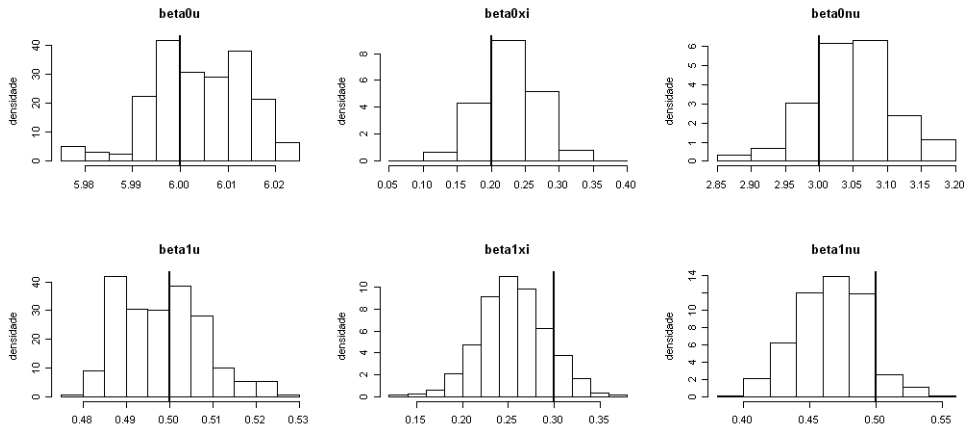
Tabela 4.1: Intervalos de credibilidade para simulações com $n = 1000$, $\beta_{0,\nu} = 3$, $\beta_{1,\nu} = 0,5$ e $\beta_{0,\xi} = 0,2$.

	$\beta_{1,\xi} = 0.3$				$\beta_{1,\xi} = -0.3$			
	$\beta_{0,u} = 6$		$\beta_{0,u} = 9$		$\beta_{0,u} = 6$		$\beta_{0,u} = 9$	
	$\beta_{1,u} = 0,5$	$\beta_{1,u} = -0,5$	$\beta_{1,u} = 0,5$	$\beta_{1,u} = -0,5$	$\beta_{1,u} = 0,5$	$\beta_{1,u} = -0,5$	$\beta_{1,u} = 0,5$	$\beta_{1,u} = -0,5$
$\beta_{0,u}$	(5,93;6,75)	(5,48;6,27)	(8,38;9,14)	(8,30;9,17)	(5,98;6,45)	(5,92;6,32)	(8,69;9,20)	(8,72;9,61)
$\beta_{1,u}$	(-0,17;0,58)	(-0,71;0,05)	(0,41;1,27)	(-0,58;-0,01)	(0,26;0,56)	(-0,71;-0,39)	(0,38;0,93)	(-0,76;-0,13)
$\beta_{0,\nu}$	(2,73;3,53)	(2,85;3,66)	(2,13;3,33)	(3,02;4,05)	(2,85;3,34)	(2,90;3,33)	(2,91;3,67)	(2,72;3,32)
$\beta_{1,\nu}$	(0,23;0,58)	(0,23;0,58)	(0,31;0,89)	(0,07;0,50)	(0,39;0,61)	(0,34;0,52)	(0,27;0,63)	(0,42;0,66)
$\beta_{0,\xi}$	(0,03;0,58)	(0,17;0,70)	(-0,21;0,53)	(-0,01;0,63)	(-0,05;0,54)	(-0,34;0,17)	(-0,20;0,46)	(-0,07;0,68)
$\beta_{1,\xi}$	(-0,02;0,49)	(-0,06;0,37)	(0,12;0,69)	(-0,20;0,37)	(-0,51;-0,01)	(-0,27;0,16)	(-0,46;0,18)	(-0,64;-0.06)

4.3.2 Simulações para amostra de tamanho 10000

As Figuras 4.5 e 4.6 mostram o histograma da distribuição a posteriori dos parâmetros da cauda para $\beta_{1,\xi} = 0,3$ e $\beta_{1,\xi} = -0,3$, ambas as Figuras com $\beta_{1,u} = 0,5$ e $\beta_{0,u} = 6$. Pode-se observar que a estimação dos parâmetros é muito mais precisa que as simulações com os mesmos parâmetros no caso onde $n = 1000$. Em todas as situações, a média a posteriori é muito próxima ao valor verdadeiro simulado. Esta informação é toda fornecida pelos dados, pois a distribuição a priori para todos os parâmetros da cauda foi dada com menor informação quando $n = 10000$. Portanto a estimação é eficiente em detectar o verdadeiro valor do vetor de parâmetros para o modelo proposto neste capítulo, dando indícios que ele pode ser bem utilizado em dados reais, como será visto na próxima seção.

Figura 4.5: Histograma da cauda para $n = 10000$, $\beta_{0,u} = 6$, $\beta_{1,\xi} = 0,3$ e $\beta_{1,u} = 0,5$. As linhas verticais representam os valores verdadeiros.



As Figuras 4.7 e 4.8 mostram o histograma da distribuição a posteriori dos parâmetros da cauda para $\beta_{1,\xi} = 0,3$ e $\beta_{1,\xi} = -0,3$, considerando $\beta_{1,u} = 0,5$ e $\beta_{0,u} = 9$. Percebe-se a eficiência da estimação em recuperar os valores verdadeiros.

A Tabela 4.2 apresenta o intervalo de credibilidade de 95% dos parâmetros da cauda para $n = 10000$. Observa-se que em todas as configurações, os valores verdadeiros dos parâmetros estão dentro do intervalo de credibilidade, sendo que estes intervalos são

Figura 4.6: Histograma da cauda para $n = 10000$, $\beta_{0,u} = 6$, $\beta_{1,\xi} = -0,3$ e $\beta_{1,u} = 0,5$. As linhas verticais representam os valores verdadeiros.

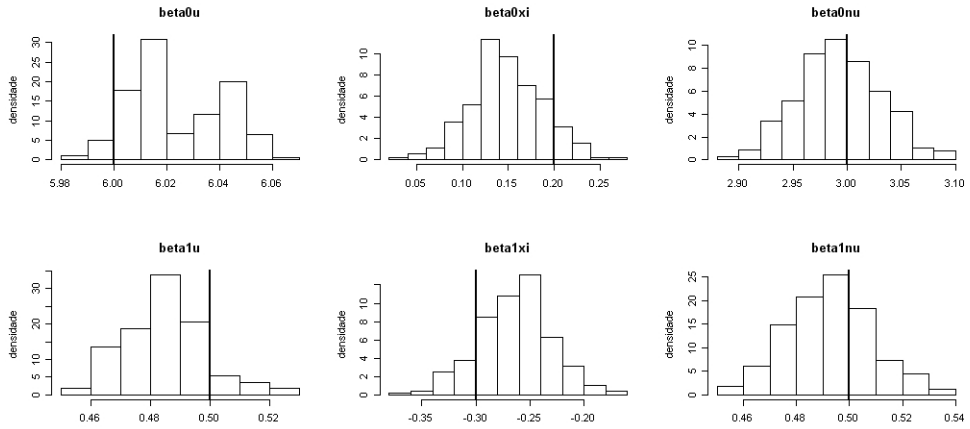
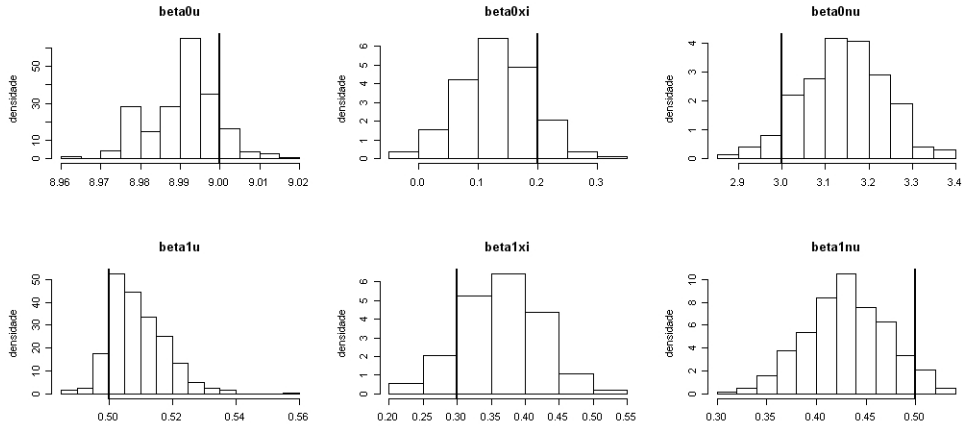


Figura 4.7: Histograma da cauda para $n = 10000$, $\beta_{0,u} = 9$, $\beta_{1,\xi} = 0,3$ e $\beta_{1,u} = 0,5$. As linhas verticais representam os valores verdadeiros.



bem mais precisos do que para $n = 1000$, podendo-se afirmar que a estimação recupera o verdadeiro valor utilizado para realizar as simulações. Portanto, para tamanhos de amostras maiores, em torno de $n = 10000$, pode-se utilizar o modelo considerando o efeito de covariáveis para estimar valores extremos.

Figura 4.8: Histograma da cauda para $n = 10000$, $\beta_{0,u} = 9$, $\beta_{1,\xi} = -0,3$ e $\beta_{1,u} = 0,5$. As linhas verticais representam os valores verdadeiros.

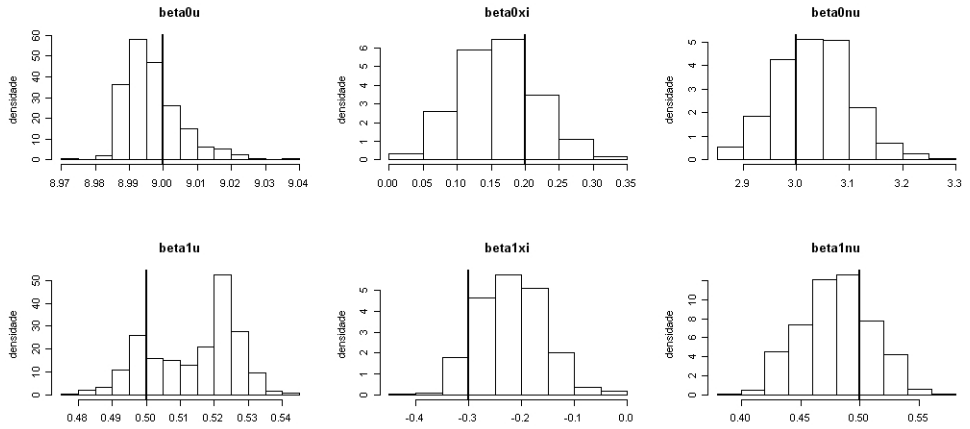


Tabela 4.2: Intervalos de credibilidade para simulações com $n = 10000$, $\beta_{0,\nu} = 3$, $\beta_{1,\nu} = 0,5$ e $\beta_{0,\xi} = 0,2$.

	$\beta_{1,\xi} = 0,3$				$\beta_{1,\xi} = -0,3$			
	$\beta_{0,u} = 6$		$\beta_{0,u} = 9$		$\beta_{0,u} = 6$		$\beta_{0,u} = 9$	
	$\beta_{1,u} = 0,5$	$\beta_{1,u} = -0,5$	$\beta_{1,u} = 0,5$	$\beta_{1,u} = -0,5$	$\beta_{1,u} = 0,5$	$\beta_{1,u} = -0,5$	$\beta_{1,u} = 0,5$	$\beta_{1,u} = -0,5$
$\beta_{0,u}$	(5,98;6,02)	(5,96;6,02)	(8,97;9,01)	(8,97;9,03)	(5,99;6,06)	(5,98;6,01)	(8,98;9,02)	(8,99;9,04)
$\beta_{1,u}$	(0,48;0,53)	(-0,51;-0,45)	(0,49;0,53)	(-0,51;-0,47)	(0,46;0,52)	(-0,51;-0,48)	(0,49;0,53)	(-0,53;-0,48)
$\beta_{0,\nu}$	(2,93;3,16)	(2,93;3,16)	(2,93;3,32)	(2,93;3,27)	(2,92;3,07)	(2,90;3,03)	(2,89;3,16)	(2,89;3,06)
$\beta_{1,\nu}$	(0,41;0,52)	(0,44;0,54)	(0,35;0,51)	(0,41;0,56)	(0,46;0,53)	(0,47;0,53)	(0,42;0,54)	(0,46;0,54)
$\beta_{0,\xi}$	(0,14;0,31)	(0,14;0,31)	(0,05;0,26)	(0,12;0,34)	(0,07;0,23)	(0,11;0,27)	(0,05;0,28)	(0,05;0,31)
$\beta_{1,\xi}$	(0,18;0,33)	(0,22;0,36)	(0,25;0,47)	(0,15;0,35)	(-0,33;-0,20)	(-0,35;-0,22)	(-0,33;-0,10)	(-0,41;-0,19)

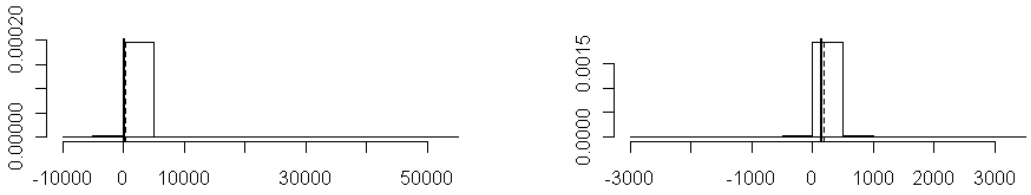
4.3.3 Cálculo de máximos e quantis altos

Em alguns casos, quando ξ_i é negativo, a distribuição GPD possui um limite superior, que é dado por $u_i - \nu_i/(\xi_i(1 + \xi_i))$. Em algumas configurações de covariáveis, pode-se calcular a distribuição a posteriori do máximo. Para as simulações onde $n = 1000$, há uma restrição em encontrar os máximos, pois a estimação dos parâmetros que determinam ξ

não foram boas, e mesmo se a média a posteriori é negativa, alguns valores gerados nas iterações do MCMC são positivos, dificultando a estimação do máximo, pois a distribuição GPD é ilimitada para $\xi > 0$.

A Figura 4.9 mostra a dificuldade de estimação dos máximos no par de covariáveis ($z_\xi = 1$,$z_\nu = 1$) e ($z_\xi = 2$,$z_\nu = 1$) em uma das simulações quando $n = 1000$. O valor de ξ desta simulação nestas covariáveis é de $-0,09$ e $-0,32$ respectivamente, mas como a amplitude dos parâmetros que determinam ξ é muito grande, como apresentado na Tabela 4.1, muitos valores de ξ na distribuição acabam sendo positivos, o que entra numa situação onde não há máximo, e quando ξ é positivo e muito próximo de 0, a equação $u_i - \nu_i/(\xi_i(1+\xi_i))$ tende a ir para valores negativos, enquanto que se ξ é muito próximo de 0 e negativo, a equação tende a ir para um valor muito alto. A consequência disso é mostrada na Figura 4.9, onde há valores situados entre -10000 e 10000 na distribuição do máximo, que não fornece nenhuma informação sobre esta quantidade.

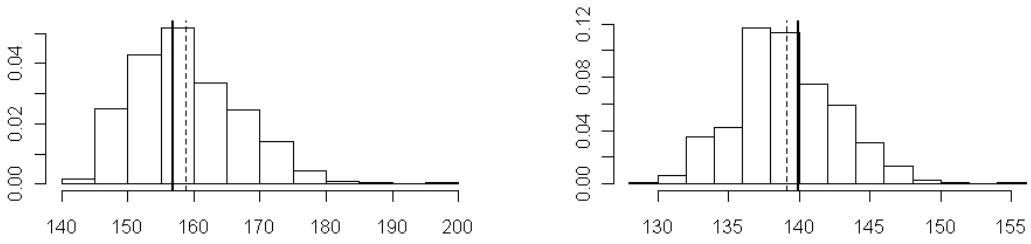
Figura 4.9: Histograma do máximo para $n = 1000$.
$\beta_{0,u} = 6$, $\beta_{1,\xi} = -0,3$ e $\beta_{1,u} = 0,5$. Painel à esquerda: $z_\xi = 1, z_\nu = 1$. Painel à direita: $z_\xi = 2, z_\nu = 1$. As linhas verticais são os máximos verdadeiros e as tracejadas as médias a posteriori dos máximos.



A Figura 4.10 mostra a estimação do máximo dos mesmos parâmetros, nas mesmas covariáveis utilizadas na Figura 4.9, mas agora com $n = 10000$. Observa-se que para $n = 10000$, a estimação do máximo é precisa e a média a posteriori do máximo é muito próxima ao máximo verdadeiro. Portanto, para tamanho de amostras grandes, o método

é eficiente em encontrar os máximos quando $\xi < 0$.

Figura 4.10: Histograma do máximo para $n = 10000$.
$\beta_{0,u} = 6$, $\beta_{1,\xi} = -0,3$ e $\beta_{1,u} = 0,5$. Painel à esquerda: $z_\xi = 1, z_\nu = 1$. Painel à direita: $z_\xi = 2, z_\nu = 1$. Linha vertical cheia: Máximo verdadeiro. Linha vertical tracejada: Média a posteriori do máximo.



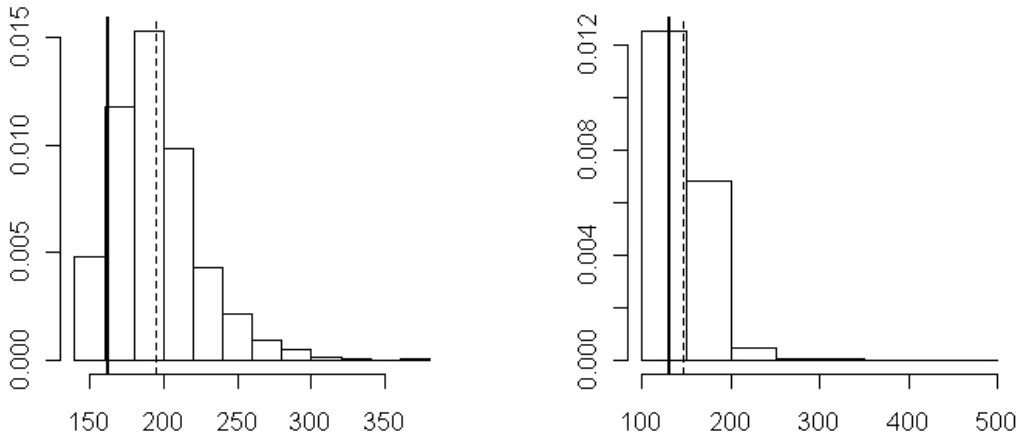
Assim como feito no capítulo anterior, este modelo permite a estimação de quantis altos da distribuição. A Figura 4.11 mostra a distribuição no quantil 99,9% para uma simulação com $n = 1000$. Assim como nos máximos, para $n = 1000$, a estimação do quantil é não é precisa, fornecendo pouca informação sobre o verdadeiro valor do quantil. Quando $z_\xi = 3$, a informação é um pouco mais precisa e a média a posteriori do quantil está próxima ao quantil verdadeiro.

A Figura 4.12 mostra a distribuição no quantil 99,9% para uma simulação com $n = 10000$. Pela Figura, em ambas as configurações de covariáveis a estimação do quantil foi eficiente, com distribuição em torno do máximo verdadeiro. Portanto, este modelo também é eficiente em encontrar os quantis altos da distribuição.

4.3.4 Identificação dos parâmetros da cauda

Assim como no capítulo anterior, na Seção 3.5, é necessário verificar se os parâmetros da cauda da distribuição GPD podem ser identificados, independente da correlação que existe entre os parâmetros u e $\sigma = \nu/(1 + \xi)$. Nas simulações, foram verificadas as

Figura 4.11: Histograma do quantil 99,9% para $n = 1000$.
$\beta_{0,u} = 6$, $\beta_{1,\xi} = -0,3$ e $\beta_{1,u} = 0,5$. Painel à esquerda: $z_\xi = 1, z_\nu = 1$. Painel à direita: $z_\xi = 3, z_\nu = 1$. Linha vertical cheia: Quantil verdadeiro. Linha vertical tracejada: Média a posteriori do quantil.

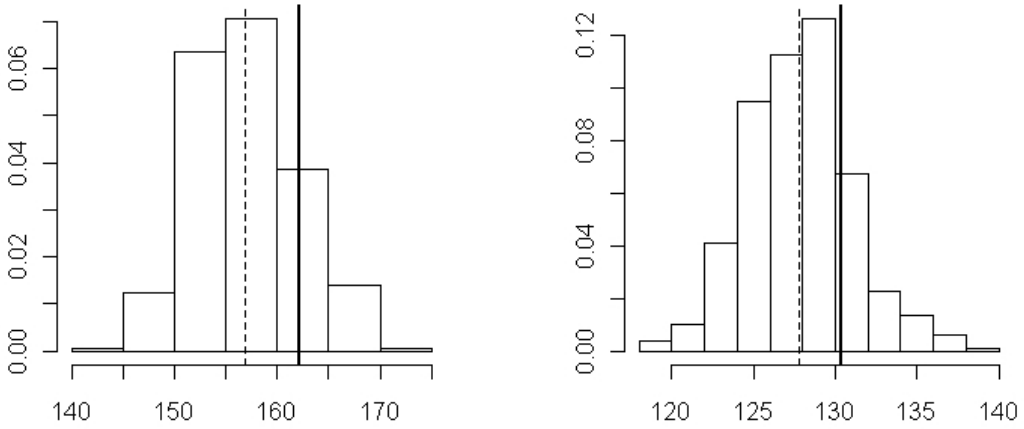


correlações entre os parâmetros $u = \beta_{0,u} + \beta_{1,u} z_u$ e σ, que pode ser escrito por $\exp(\beta_{0,\nu} + \beta_{1,\nu} z_\nu) / \exp(\beta_{0,\xi} + \beta_{1,\xi} z_\xi)$.

A Figura 4.13 mostra o gráfico de dispersão entre os parâmetros para as simulações com $n = 1000$. Pela figura, nota-se que que há pouca correlação entre os parâmetros, não sendo possível visualizar esta correlação nos gráficos de dispersão. Além disso, as linhas dos valores verdadeiros estão dentro da maior moda. Portanto, embora nas simulações com $n = 1000$ a estimação dos parâmetros não tenha sido precisa, isto não ocorreu pela falta de identificabilidade entre os parâmetros da cauda.

A Figura 4.14 mostra o gráfico de dispersão entre os parâmetros para as simulações com $n = 10000$. Pela figura, nota-se que não parece haver correlação entre os parâmetros, sendo estes bem estimados próximos aos valores verdadeiros. No caso onde $\beta_{1,\xi} = -0,3$ e $z_\xi = 0,5$ há duas massa de dados, indicando a distribuição bimodal do limiar. O valor

Figura 4.12: Histograma do quantil 99,9% para $n = 10000$.
$\beta_{0,u} = 6$, $\beta_{1,\xi} = -0,3$ e $\beta_{1,u} = 0,5$. Painel à esquerda: $z_\xi = 1, z_\nu = 1$. Painel à direita: $z_\xi = 3, z_\nu = 1$. Linha vertical cheia: Quantil verdadeiro. Linha vertical tracejada: Média a posteriori do quantil.



verdadeiro do limiar esteve situado na menor moda, porém ainda sim a média a posteriori do limiar ficou próxima do verdadeiro valor.

4.3.5 Conclusões das simulações

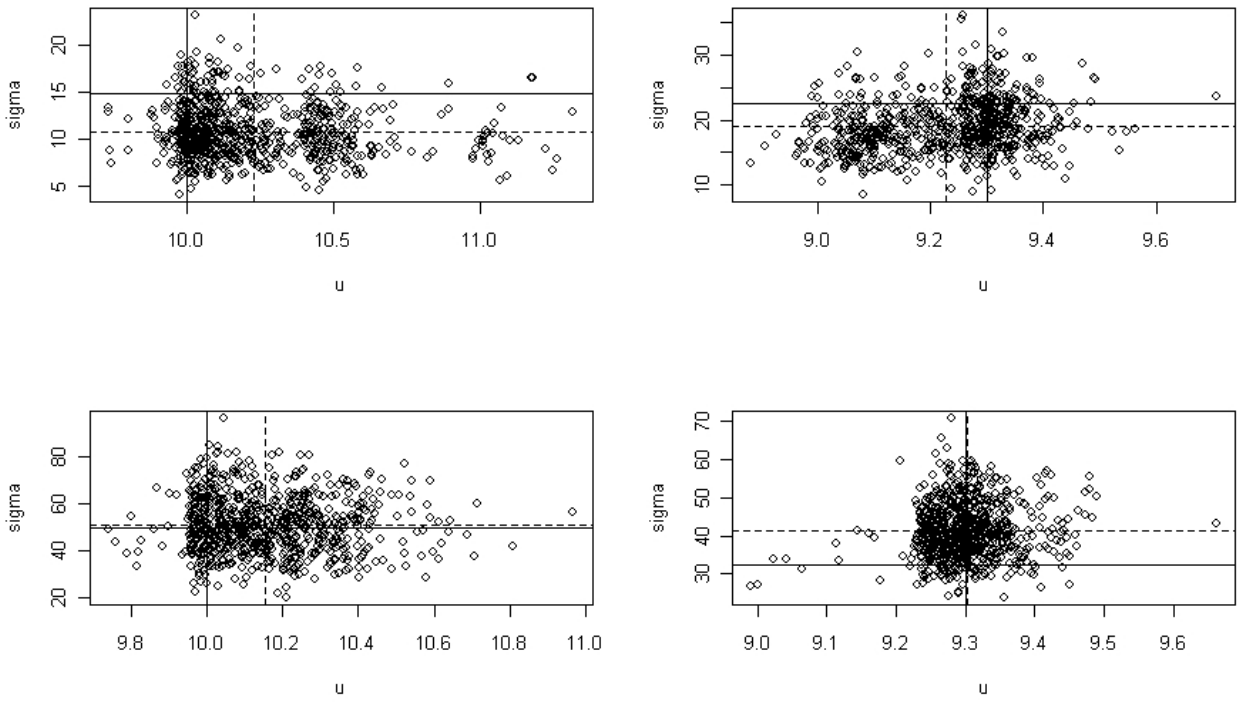
Após apresentar o modelo $MGPDR_k$ e realizar um estudo de simulações deste modelo, pode-se tirar as seguintes conclusões:

- A estimação não conseguiu ser precisa para amostras de tamanho $n = 1000$. Embora em quase todas as situações estudadas o verdadeiro valor do parâmetro estivesse dentro de um intervalo de credibilidade de 95% da distribuição a posteriori, esta distribuição foi muito vaga e deu pouca informação a respeito dos parâmetros;

- para amostras com $n = 10000$, a estimação dos parâmetros foi muito eficiente, onde

Figura 4.13: Relação dos parâmetros da cauda para $n = 1000$

Simulações com $\beta_{0,u} = 9$, $\beta_{1,u} = 0,5$ e $z_\nu = 1$. Primeira linha: $\beta_{1,\xi} = 0,3$. Segunda linha:$\beta_{1,\xi} = -0,3$. Primeira coluna: $z_\xi = 0,5$. Segunda coluna: $z_\xi = 2$. As linhas cheias são os valores verdadeiros e as tracejadas a média a posteriori.



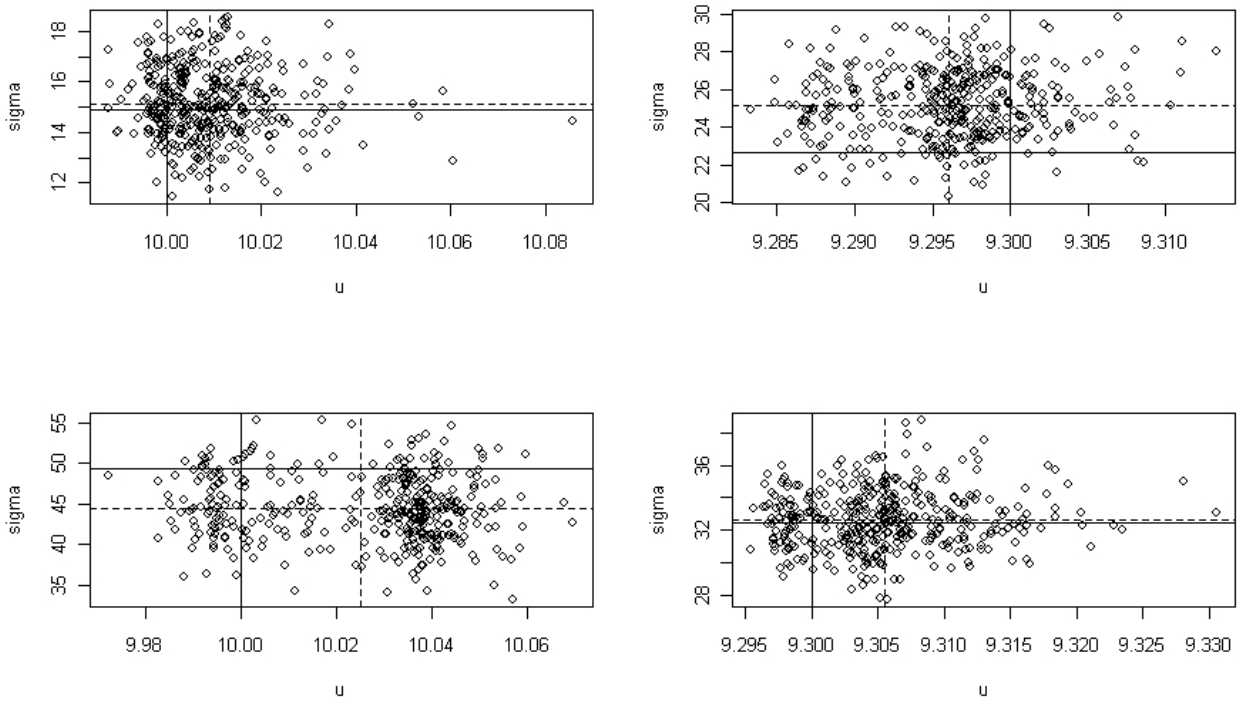
em todas as situações a distribuição a posteriori do parâmetro teve uma variabilidade muito pequena e próxima do valor verdadeiro simulado;

- também para amostras grandes, o modelo mostrou-se muito eficiente na estimação do máximo da distribuição nos casos onde $\xi < 0$ e no cálculo dos quantis extremos das distribuições, que são duas das mais importantes quantidades quando se analisa dados extremos;

- para as simulações estudadas neste capítulo, não houve problema de correlação entre os parâmetros da cauda, mostrando que neste caso não houve problema de falta de

Figura 4.14: Relação dos parâmetros da cauda para $n = 10000$

Simulações com $\beta_{0,u} = 9$, $\beta_{1,u} = 0,5$ e $z_\nu = 1$. Primeira linha: $\beta_{1,\xi} = 0,3$. Segunda linha:$\beta_{1,\xi} = -0,3$. Primeira coluna: $z_\xi = 0,5$. Segunda coluna: $z_\xi = 2$. As linhas cheias são os valores verdadeiros e as tracejadas a média a posteriori.



identificabilidade entre estes parâmetros.

4.4 Aplicações

Esta seção irá mostrar resultados de dados reais em duas aplicações de dados extremos de temperatura. Temperaturas máximas em cidades dos Estados Unidos e temperaturas mínimas em cidades do Estado do Rio de Janeiro. O modelo proposto será comparado com outros modelos propostos na literatura, como o MG_k de Wiper et al. (2001), o $MGPD_1$ de Behrens et al. (2004) e o modelo $MGPD_k$ proposto no capítulo anterior,

segundo os critérios de comparação BIC e DIC.

4.4.1 Aplicação 1 - Temperaturas máximas nos EUA

O conjunto de dados coletados é de temperatura média diária, em ^oF, de 1995 à 2008 em 84 cidades dos Estados Unidos. Os dados estão disponíveis no site `www.engr.udayton.edu/weather`. A análise foi feita utilizando os máximos mensais. No total foram analisadas 14356 observações. Vários fatores podem influenciar a temperatura. Entre os principais estão a época do ano e a localização da cidade. Os Estados Unidos, por ser um país de dimensões continentais, possui diferentes faixas de latitude, o que causa uma forte influência nas temperaturas das cidades. Existem latitudes que ficam próximas do Trópico de Câncer, como Miami com latitude de $25^o47'$, e cidades que se situam entre o Trópico de Câncer e o o Círculo Polar Ártico, como Seattle, com latitude de $47^o37'$. A Figura 4.15 mostra o mapa dos Estados Unidos com a localização de suas principais cidades.

Em relação aos meses do ano, como a temperatura possui ciclos sazonais, as covariáveis que representam os meses podem ser dadas em termos de funções trigonométricas. Portanto, o vetor de covariáveis para este modelo pode ser escrito da seguinte maneira

$\mathbf{z}_i = (z_{0,i}, z_{1,i}, z_{2,i}, z_{3,i}, z_{4,i})$,

onde:

- $z_{0,i} = 1$ é o intercepto.
- $z_{1,i} = \cos\left(\frac{2\pi m_i}{12}\right)$ e $z_{2,i} = \sin\left(\frac{2\pi m_i}{12}\right)$, onde m_i é o mês da observação i,
- $z_{3,i} = (l_i - m_l)/10$, onde l_i é a latitude da observação i e m_l é o valor médio da latitude de todas as observações.
- $z_{4,i} = z_{3,i}z_{2,i}$ representa a interação entre o mês do ano e a latitude.

Foi feita a estimação pelo modelo $MGPDR_k$ proposto neste trabalho, para diferentes valores de k, onde cada parâmetro da cauda é escrito em função do vetor de covariáveis

Figura 4.15: Mapa dos Estados Unidos com os estados e as maiores cidades.

z. A distribuição a priori para o vetor de parâmetros β_u foram prioris vagas Normais com $\beta_{u,0} \sim N(40, 10000)$ e $\beta_{u,i} \sim N(0, 1000)$, com $i = 1, ..., 4$.

A Tabela 4.3 mostra os resultados das medidas de ajuste BIC e DIC para os diferentes modelos comparados. Observando a tabela, percebe-se que ao utilizar modelos de regressão nos parâmetros da cauda há um ganho significativo no ajuste, pois o modelo $MGPDR_k$ é melhor do que se os parâmetros da cauda fossem fixos (modelo $MGPD_k$), ou se fosse utilizada uma abordagem totalmente não-paramétrica via mistura de Gamas (modelo MG_k). A Tabela 4.4 apresenta os resultados da estimação dos parâmetros da cauda para o modelo $MGPDR_3$, apontado pela Tabela 4.3 como o melhor modelo.

Em relação ao modelo $MGPDR_k$, outras reparametrizações também foram analisadas para verificar se poderiam ser melhores do que a proposta neste trabalho. A primeira foi considerar $u_i = \exp(\mathbf{z}\beta_u)$. Com esta parametrização, o melhor BIC e DIC dos modelos $MGPDR_k$ foi para $k = 3$ com respectivos valores 0,9333 $\times 10^5$ e 0,9309 $\times 10^5$. Outra

Tabela 4.3: Medidas de ajuste para as temperaturas dos EUA

Modelo	Pd	DIC $\times 10^5$	BIC $\times 10^5$
$MGPDR_1$	19,30	0,9352	0,9371
$MGPDR_2$	11,54	0,9169	0,9192
$MGPDR_3$	5,16	**0,9141**	**0,9163**
$MGPDR_4$	8,55	0,9151	0,9177
MG_3*	6,86	1,1321	1,1331
$MGPD_2*$	5,99	1,1325	1,1335

MG_k* e $MGPD_k*$ são os melhores modelos de cada classe de acordo com o DIC.

reparametrização estudada foi considerar a parametrização da distribuição GPD usual, como na equação (3.2). A parametrização proposta para $\sigma = \exp(\mathbf{z}\beta_\sigma)$, considerando uma distribuição Uniforme para o vetor de parâmetros β_σ, sendo esta não informativa, porém não satisfazendo a regra de Jeffreys. Com esta parametrização o melhor BIC e DIC foram respectivamente 0,9243 $\times 10^5$ e 0,9265 $\times 10^5$. Portanto, de acordo com os dados, a parametrização original proposta na Seção 4.1, além de ter uma vantagem teórica dos parâmetros da cauda (β_ξ, β_ν) possuírem distribuição a priori de Jeffreys proporcional Uniforme, também tiveram a vantagem prática de se ajustar melhor aos dados, tendo menores medidas de ajuste do que outras reparametrizações.

A Figura 4.16 apresenta o histograma da distribuição a posteriori dos parâmetros da cauda para o melhor modelo $MGPDR_k$, segundo a Tabela 4.3. A Tabela 4.4 apresenta o intervalo de credibilidade para estes parâmetros. Em todos os parâmetros, a amplitude da distribuição é pequena, e em quase todas as situações o efeito da covariável é significativo, pois o valor 0 aparece apenas em $\beta_{4,\xi}$, que é o efeito da interação entre os meses do ano e a latitude em ξ. Em relação a latitude, esta tem uma influência negativa no limiar, ou seja, quanto maior a latitude menor o limiar, porém maiores são os valores de ξ e ν.

Além de identificar o melhor modelo para a análise de temperatura e estimar os seus

Figura 4.16: Histograma da cauda para o modelo $MGPDR_3$ dos dados dos EUA.

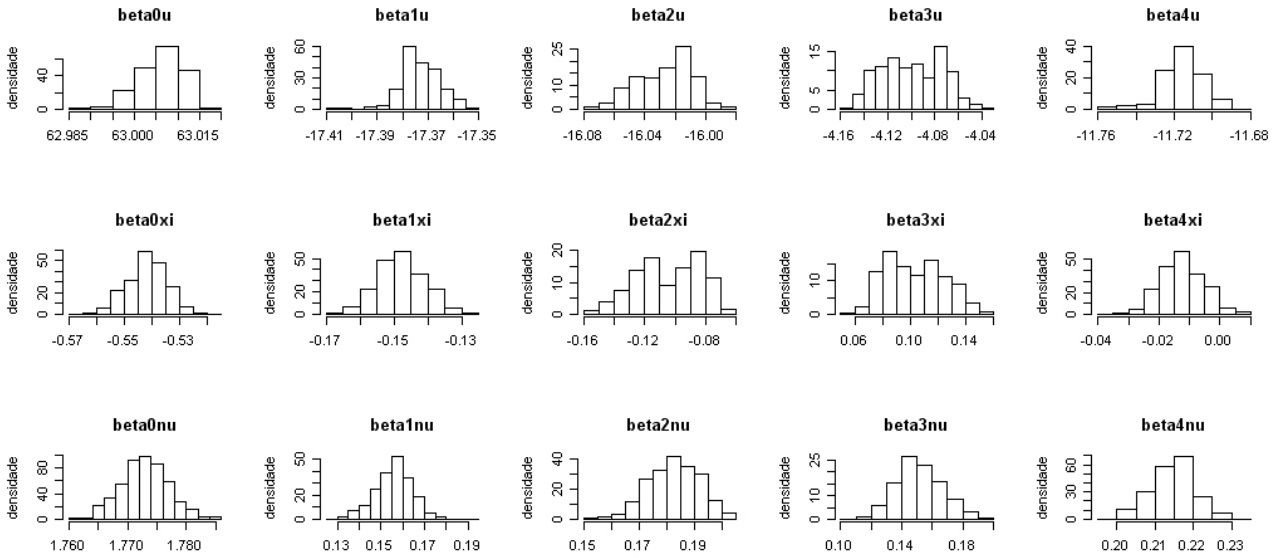


Tabela 4.4: Média a posteriori dos parâmetros para o modelo $MGPDR_3$ para os dados dos EUA.

Em parênteses os intervalos de 95% de credibilidade

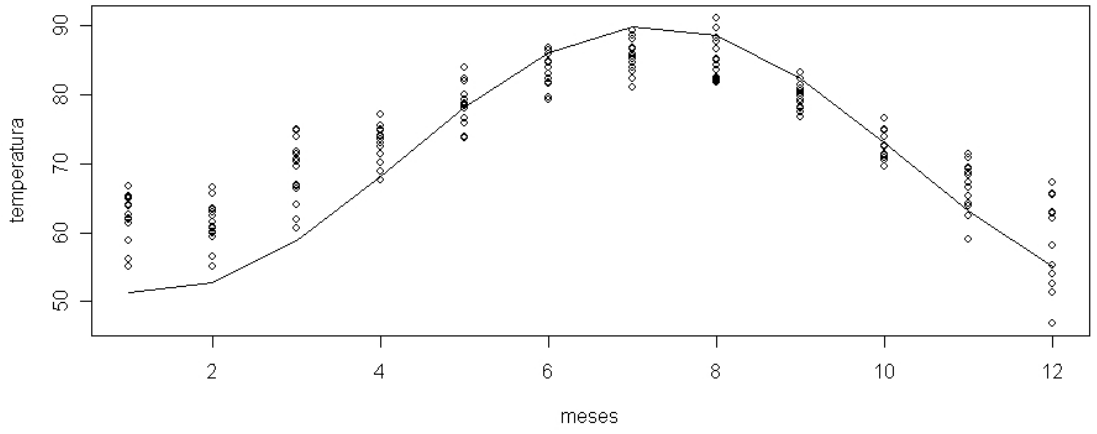
$\beta_{0,\xi}$	$\beta_{1,\xi}$	$\beta_{2,\xi}$	$\beta_{3,\xi}$	$\beta_{4,\xi}$
-0,54 (-0,55;-0,53)	-0,15 (-0,16;-0,13)	-0,10 (-0,14;-0,07)	0,10 (0,07;0,14)	-0,01 (-0,02;0,01)
$\beta_{0,\nu}$	$\beta_{1,\nu}$	$\beta_{2,\nu}$	$\beta_{3,\nu}$	$\beta_{4,\nu}$
1,77 (1,76;1,78)	0,15 (0,14;0,17)	0,18 (0,16;0,20)	0,15 (0,12;0,18)	0,21 (0,20;0,23)
$\beta_{0,u}$	$\beta_{1,u}$	$\beta_{2,u}$	$\beta_{3,u}$	$\beta_{4,u}$
63,01 (62,99;63,02)	-17,37 (-17,39;-17,35)	-16,03 (-16,06;-15,99)	-4,10 (-4,14; -4,06)	-11,71 (-11,74;-11,69)

parâmetros, pode-se fazer uma análise detalhada sobre o comportamento da temperatura em cada estação do ano ou em cada cidade. Por exemplo, como o limiar também é formulado em função das covariáveis, cada cidade vai ter um limiar diferente e em relação a uma única cidade os valores do limiar irão se alterar de acordo com os meses do ano.

As Figuras 4.17 e 4.18 apresentam a evolução do limiar em relação aos meses do ano para as cidades de Atlanta, Georgia, e Portland, Maine, respectivamente com latitudes

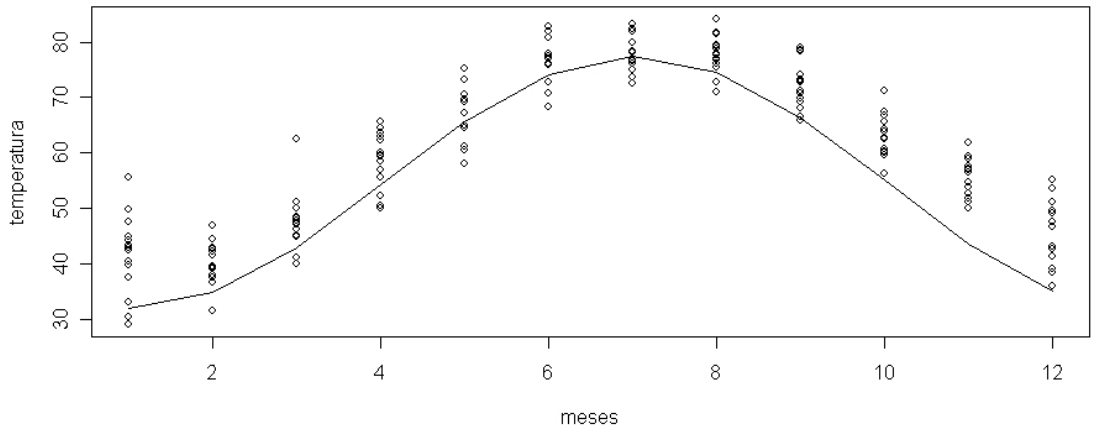
de $33^o45'$ e $43^o39'$. Observando as figuras, percebe-se que o limiar acompanha a curva dos dados ao longo dos anos, ou seja, as temperaturas são menores nos primeiros e últimos meses do ano, e vão aumentando conforme se aproxima o meio do ano, que é a época do verão no hemisfério norte. A cidade de Portland, por ser uma cidade com latitude maior que Atlanta, apresenta durante quase todo o ano temperaturas mais baixas. Um outro fator importante de se observar é que a maioria dos dados está acima do limiar, ou seja, no melhor modelo estudado, a grande maioria das observações foi considerada um valor extremo pertencente a cauda da distribuição. Como se trabalha com máximos mensais, a maioria destes valores é considerada uma observação extrema, fazendo parte da cauda. Como nem todo máximo mensal necessariamente é uma observação extrema, alguns dos valores não fazem parte da cauda e são modelados pela mistura de Gamas. O mais importante em relação ao limiar é que a distribuição a priori para este parâmetro é vaga, deixando apenas os dados escolherem o valor do limiar que fornecerá a melhor distribuição a posteriori dos parâmetros.

Figura 4.17: Observações ao longo do ano para os dados de Atlanta.
A linha cheia representa a média a posteriori do limiar para cada mês.



Além de estudar o comportamento do limiar ao longo do ano para cada cidade, outro

Figura 4.18: Observações ao longo do ano para os dados de Portland.
A linha cheia representa a média a posteriori do limiar para cada mês.



estudo que pode ser feito é um estudo do comportamento do limiar entre todas as cidades para cada mês do ano, ou seja, verificar como as latitudes diferentes influenciam na estimação do limiar. As Figuras 4.19 e 4.20 mostram os valores da média a posteriori do limiar para os meses de Janeiro, no inverno, e Julho, no verão. Observando os gráficos, nota-se que as temperaturas máximas no verão são bem maiores do que no inverno, e o limiar capta esta diferença, sendo este muito maior durante o verão. Em relação as latitudes, observando os gráficos de dispersão a tendência linear parece pequena, porém os resultados de estimação apresentados na Tabela 4.4 mostraram que o efeito das covariáveis foi significativo e assim foi mantida a estrutura de tendência linear, que mostrou ser a melhor, comparando-se com outras estruturas, como por exemplo uma tendência exponencial entre temperatura e latitude. Percebe-se nas tabelas que, de acordo com o valor estimado de $\beta_{3,u}$, quanto maior a latitude, menor e a temperatura. Esta relação parece ser mais acentuada no inverno.

Para fazer inferência sobre valores extremos, uma outra medida importante que pode ser encontrada baseado na estimação dos parâmetros do modelo é o cálculo dos máximos,

Figura 4.19: Observações com as latitudes para o mês de Janeiro nos EUA.
A linha cheia representa a média a posteriori do limiar para cada cidade.

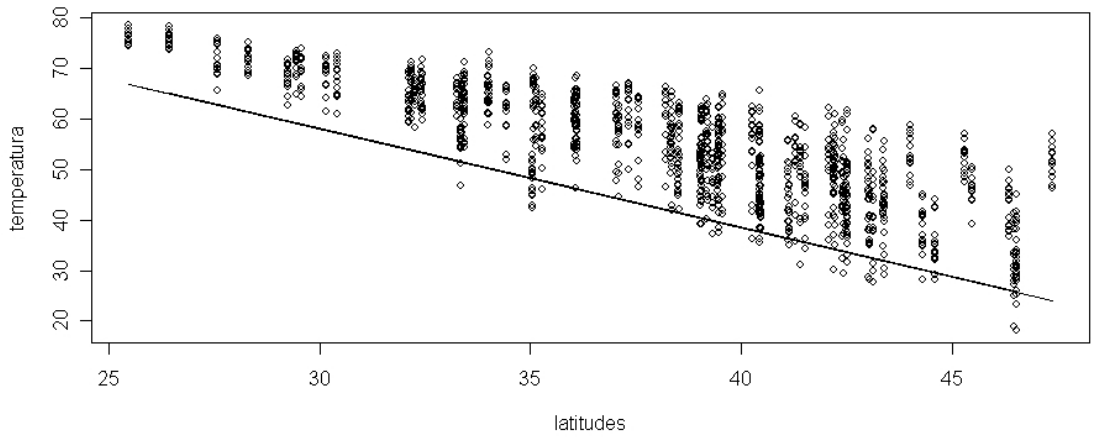
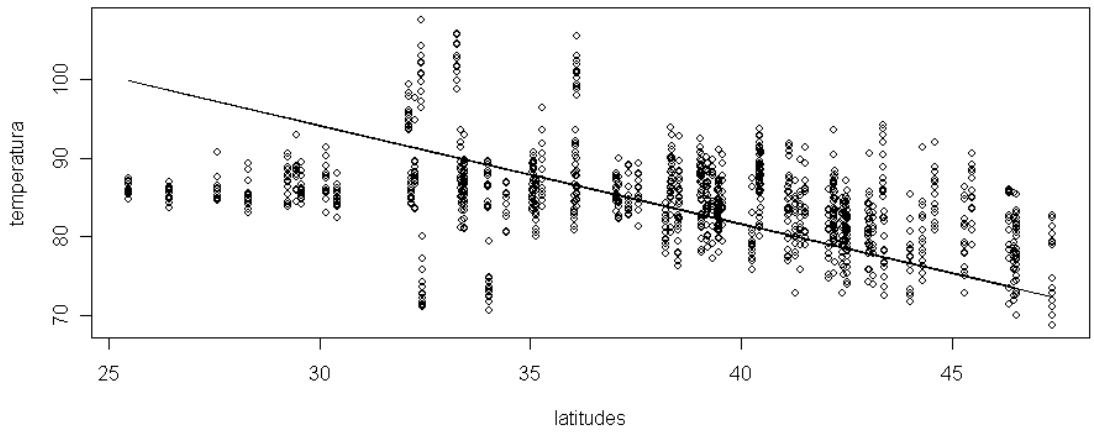


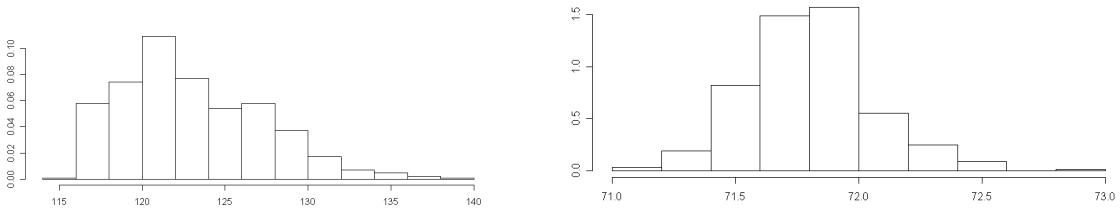
Figura 4.20: Observações com as latitudes para o mês de Julho nos EUA.
A linha cheia representa a média a posteriori do limiar para cada cidade.



que é finito nos casos onde o parâmetro de forma da distribuição GPD é negativo. A Figura 4.21 mostra o histograma da distribuição a posteriori dos máximos para duas cidades diferentes, em duas diferentes estações do ano. Pode-se dizer que a temperatura

máxima que a cidade de Nova Orleans pode alcançar em algum dia de Junho está em torno de 115^oF e 135^oF, enquanto que em Denver, no mês de Dezembro, o máximo está em torno de 71^oF e 73^oF.

Figura 4.21: Histograma dos máximos para os dados dos EUA.

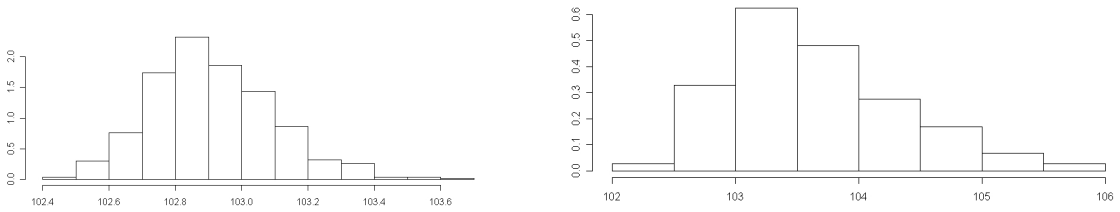


a) Nova Orleans, em Junho.

b) Denver, em Dezembro.

Outra medida importante que se pode encontrar quando são analisados dados extremos é a distribuição de quantis altos da distribuição. A Figura 4.22 mostra a distribuição a posteriori do quantil $99,9\%$ em duas cidades em dois meses do ano. Assim, por exemplo, baseado na média a posteriori do quantil, a probabilidade de no mês de março, a temperatura da cidade de Orlando ser maior do que $103,4^oF$ é de $0,1\%$.

Figura 4.22: Histograma do quantil $99,9\%$ para os dados dos EUA.



a) Austin, em Setembro.

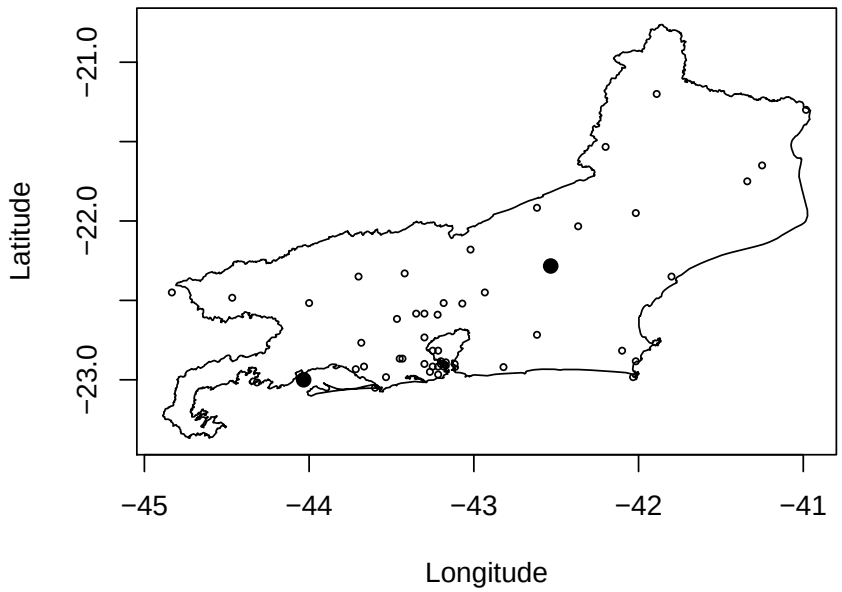
b) Orlando, em Março.

4.4.2 Aplicação 2 - Temperaturas mínimas no estado do Rio de Janeiro

Este segundo conjunto de dados consiste em temperaturas mínimas diárias, em ^oC em postos de monitoramento espalhados pelo estado do Rio de Janeiro, num período de 1961 a 2000. Para realizar a análise, foram utilizadas as temperaturas mínimas mensais, resultando num total de 11336 observações para a análise.

A Figura 4.23 mostra como as estações de monitoramento estão espalhadas pelo estado.

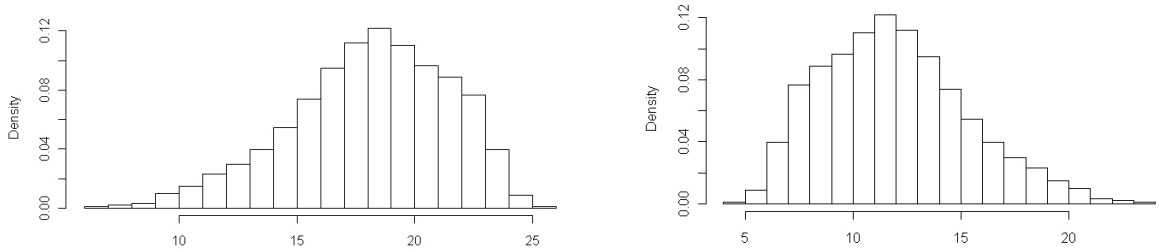
Figura 4.23: Mapa do estado do Rio de Janeiro com as estações de monitoramento. Os pontos cheios são as cidades de Ilha Guaíba, no litoral oeste, e Nova Friburgo, no centro.



Neste conjunto de dados, como são utilizadas temperaturas mínimas, a cauda da distribuição está à esquerda. Para fazer a análise segundo o modelo proposto, foi feita a transformação $x = 30,0 - y$, onde y são os dados originais. Como a maior temperatura das observações foi de 25,6, fazendo a transformação, há a garantia que as observações são

positivas e com cauda à direita. Assim, o modelo também permite calcular a probabilidade de observações mínimas maiores que 25,6. A Figura 4.24 mostra o histograma dos dados originais e dos dados transformados.

Figura 4.24: Histograma das observações dos dados do Rio de Janeiro.
Painel à esquerda: Dados originais. Painel à direita: Dados transformados.



Observando no mapa na Figura 4.23, as diferenças de Latitudes do Estado são pequenas, entre $20^o76'$ e $23^o36'$. Uma covariável que irá ter mais relevância na análise dos dados é a altitude. O estado do Rio de Janeiro tem um relevo diversificado, com regiões no nível do mar, regiões serranas e planaltos. Assim, uma covariável escolhida para a análise destes dados foi a altitude. Os meses do ano certamente é outra covariável que interfere nos níveis de temperatura e será analisada por meio de transformações trigonométricas.

O vetor de covariáveis para este modelo pode ser escrito da seguinte maneira

$\mathbf{z}_i = (z_{0,i} z_{1,i} z_{2,i} z_{3,i} z_{4,i})$,

onde:

- $z_{0,i} = 1$ é o intercepto
- $z_{1,i} = \cos\left(\frac{2\pi m_i}{12}\right)$ e $z_{2,i} = \sin\left(\frac{2\pi m_i}{12}\right)$, onde m_i é o mês da observação i,
- $z_{3,i} = (l_i - m_l)/100$, onde l_i é a altitude da observação i e m_l é o valor médio da altitude de todas as observações.
- $z_{4,i} = z_{3,i} z_{2,i}$ representa a interação entre o mês do ano e a altitude

As distribuições a priori dos parâmetros do limiar do modelo $MGPDR_k$ são dadas por $\beta_{u,0} \sim N(10, 50)$ e $\beta_{u,i} \sim N(0, 30)$, $i = 1, ..., 4$.

A Tabela 4.5 mostra os resultados das medidas de Ajuste BIC e DIC para os diferentes modelos comparados.

Tabela 4.5: Medidas de ajuste para as temperaturas do estado do Rio de Janeiro

Modelo	Pd	DIC $\times 10^5$	BIC $\times 10^5$
$MGPDR_1$	6,54	0,4253	0,4249
$MGPDR_2$	9,78	0,4230	0,4250
$MGPDR_3$	18,26	**0,4222**	**0,4248**
$MGPDR_4$	11,03	0,4227	0,4253
MG_4*	8,13	0,5863	0,5876
$MGPD_2*$	1,70	0,5866	0,5874

MG_k* e $MGPD_k*$ são os melhores modelos de cada classe de acordo com o DIC.

Assim como nos dados da aplicação anterior, os modelos $MGPDR_k$ tem uma ampla vantagem na estimação em relação aos outros modelos, ou seja, o modelo é melhor ajustado quando é considerado o efeito de covariáveis na cauda da distribuição.

A Figura 4.25 apresenta os histogramas das distribuições a posteriori dos parâmetros da cauda para o modelo $MGPDR_3$, o melhor segundo a Tabela 4.5.

A Tabela 4.6 mostra a média a posteriori e o intervalo de credibilidade de 95% parâmetros da cauda, onde nota-se que quase todos os parâmetros apontam um efeito significativo das covariáveis e interações nos três parâmetros da distribuição GPD, com exceção do efeito da interação entre altitude e estação do ano para a determinação dos parâmetros ξ e ν.

Assim como na aplicação anterior, pode-se tirar medidas importantes que podem ajudar a avaliar o comportamento de observações extremas na temperatura. Os dados foram transformados com a finalidade de se ter uma cauda à direita, que é como se usualmente

Figura 4.25: Histograma da cauda para o modelo $MGPDR_3$ dos dados do Rio de Janeiro.

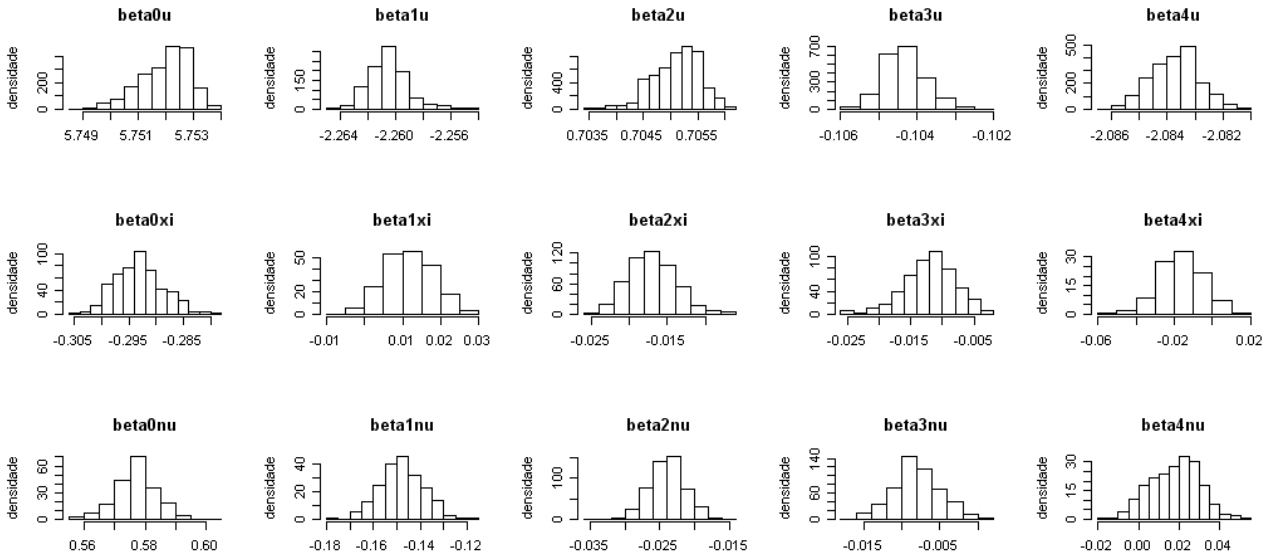


Tabela 4.6: Média a posteriori dos parâmetros para o modelo $MGPDR_3$ dos dados do Rio de Janeiro.

Em parênteses os intervalos de 95% de credibilidade

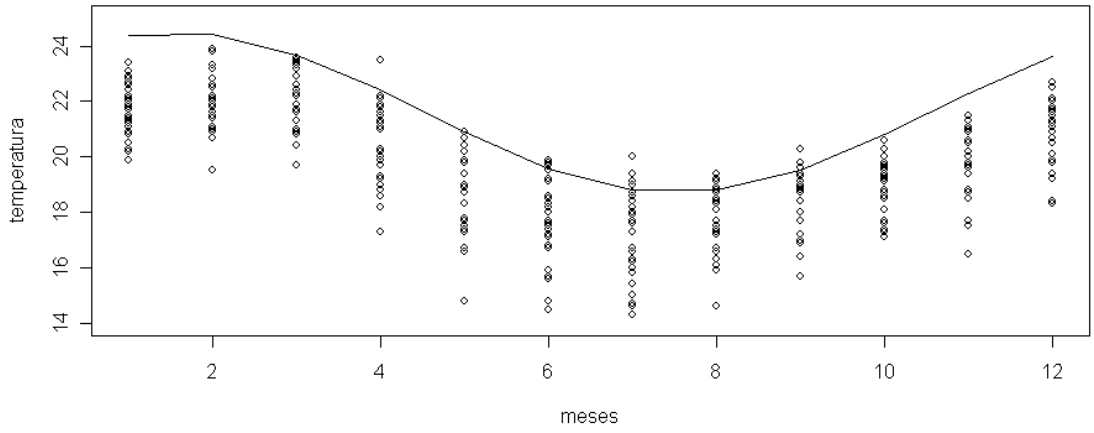
$\beta_{0,\xi}$	$\beta_{1,\xi}$	$\beta_{2,\xi}$	$\beta_{3,\xi}$	$\beta_{4,\xi}$
-0,29 (-0,30;-0,28)	0,01 (0,00;-0,02)	-0,02 (-0,02;-0,01)	-0,01 (-0,02;-0,01)	-0,02 (-0,04;0,01)
$\beta_{0,\nu}$	$\beta_{1,\nu}$	$\beta_{2,\nu}$	$\beta_{3,\nu}$	$\beta_{4,\nu}$
0,58 (0,56;0,59)	-0,15 (-0,17;-0,13)	-0,02 (-0,03;-0,02)	-0,008 (-0,014;-0,002)	0,02 (-0,01;0,04)
$\beta_{0,u}$	$\beta_{1,u}$	$\beta_{2,u}$	$\beta_{3,u}$	$\beta_{4,u}$
5,752 (5,750;5,753)	-2,260 (-2,262;-2,256)	0,705 (0,704;0,706)	-0,104 (-0,105; -0,103)	-2,083 (-2,085;-2,082)

trabalha na distribuição GPD, para coletar medidas de temperaturas ao longo do ano, para as diferentes estações, além de estatísticas extremas como máximos e quantis. Nesta aplicação, como o interesse principal está em analisar temperaturas mínimas, é necessário fazer a transformação inversa para se fazer a análise baseado nos dados na escala original. Assim, são analisados os mínimos e quantis baixos.

As Figuras 4.26 e 4.27 mostram as temperaturas mínimas mensais ao longo do ano

para as cidades de Cabo Frio, ao nível do mar, e Nova Friburgo, na região serrana do Estado. Pelas figuras, percebe-se que o limiar acompanha bem o comportamento das observações, que ao contrário dos EUA, cai de temperatura no meio do ano, que é o inverno no hemisfério sul. Por estar situada na região serrana, a temperatura na cidade de Nova Friburgo é sempre menor do que Cabo Frio. Pode-se perceber pelas figuras que o limiar escolhido pelas observações foi situado num quantil baixo, ficando a maioria das observações acima da média a posteriori do limiar.

Figura 4.26: Observações ao longo do ano para os dados de Cabo Frio.
A linha cheia representa a média a posteriori do limiar para cada mês.



Pode-se verificar também o comportamento do limiar em relação as altitudes em diferentes meses do ano. As Figuras 4.28 e 4.29 mostram os valores da média a posteriori do limiar para os meses de Janeiro, no verão, e Julho, no inverno. Observando os gráficos nota-se que, em relação as altitudes, as temperaturas mínimas são menores em regiões com altitudes mais altas.

Outra medida que se pode encontrar no modelo $MGPDR_k$ é a distribuição dos máximos, quando ξ é negativo. Como nos dados do Rio de Janeiro, trabalha-se com os dados transformados para a cauda da distribuição estar à direita, encontrando a dis-

Figura 4.27: Observações ao longo do ano para os dados de Nova Friburgo.
A linha cheia representa a média a posteriori do limiar para cada mês.

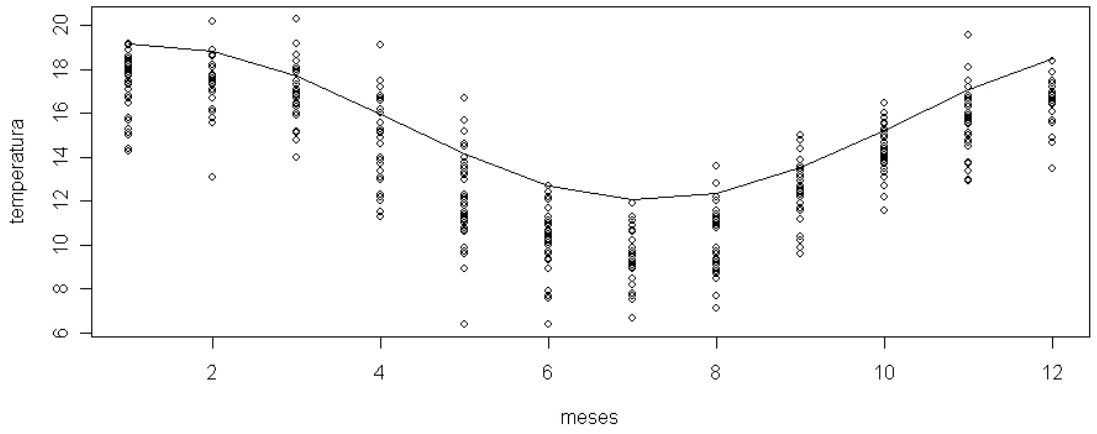
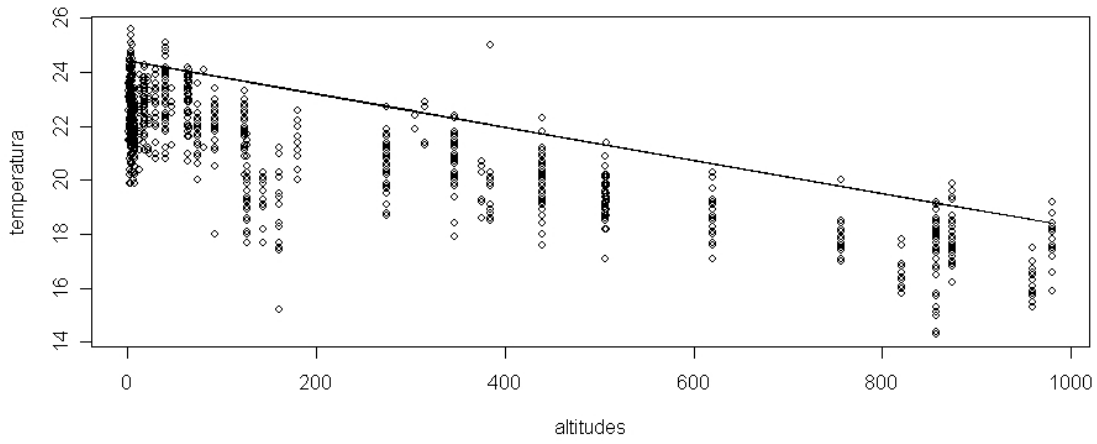
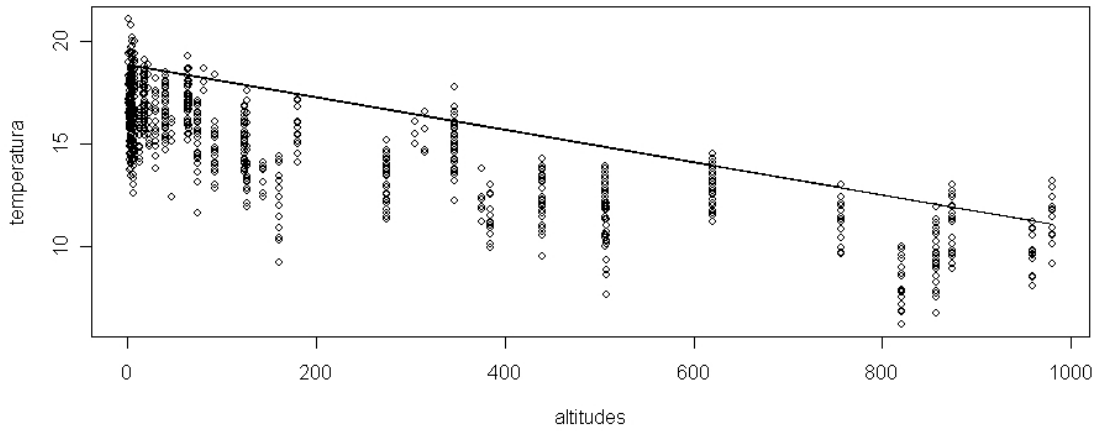


Figura 4.28: Observações com as latitudes para o mês de Janeiro no Rio de Janeiro .
A linha cheia representa a média a posteriori do limiar para cada cidade.



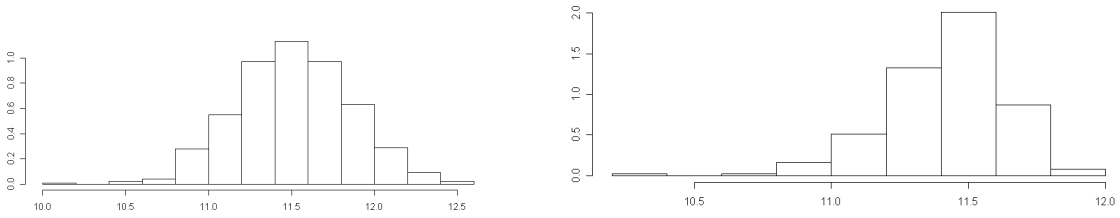
tribuição dos máximos dos dados transformados e voltando esta distribuição na escala original, o resultado será a distribuição do mínimo das temperaturas. A Figura 4.30 mostra a distribuição do mínimo em duas cidades do Rio de Janeiro em dois diferentes

Figura 4.29: Observações com as latitudes para o mês de Julho no Rio de Janeiro.
A linha cheia representa a média a posteriori do limiar para cada cidade.



meses. Assim, pode-se dizer que a temperatura mínima em Angra dos Reis no mês de abril está entre $10,5^oC$ e $12,5^oC$.

Figura 4.30: Histograma dos mínimos para os dados do Rio de Janeiro.



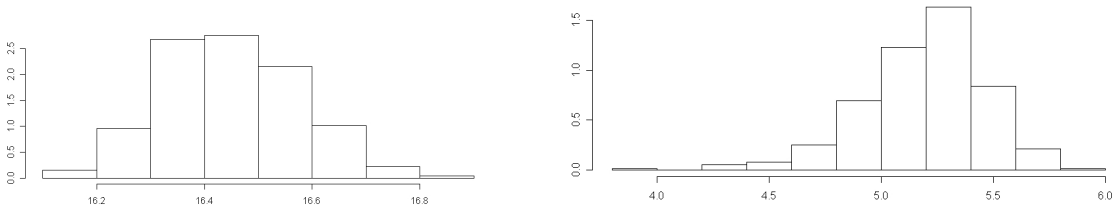
a) Angra dos Reis, em Abril.

b) Ilha Rasa, em Novembro.

Uma outra medida que pode ser encontrada são os quantis baixos. Ao contrário de dados máximos, quando se analisa observações mínimas, o maior interesse é encontrar a probabilidade de se obter uma observação menor do que um determinado valor. Para encontrar o quantil $0,1\%$ de alguma cidade em um determinado mês do ano, para os dados do Rio de Janeiro, encontra-se a distribuição do quantil $99,9\%$ dos dados transformados e

retorna esta distribuição para a escala dos dados originais, obtendo assim a distribuição do quantil $0,1\%$ das temperaturas mínimas. A Figura 4.31 apresenta a distribuição do quantil $0,1\%$ para duas cidades diferentes em dois meses do ano. Baseado na média a posteriori dos quantis, pode-se dizer que a probabilidade da temperatura em Niterói no mês de Janeiro ser menor do que $16,4^oC$ é de $0,1\%$, enquanto que esta mesma probabilidade se equivale a temperatura de apenas $5,3^oC$ para a cidade de Teresópolis no mês de Julho.

Figura 4.31: Histograma do quantil $0,1\%$ para os dados do Rio de Janeiro.



a) Niterói, em Janeiro.

b) Teresópolis, em Julho.

4.4.3 Conclusões das aplicações

Após realizar a estimação dos parâmetros nas duas aplicações de temperaturas, pode-se tirar como principais conclusões:

- O modelo $MGPDR_k$ é superior ao modelo que considera a cauda fixa ($MGPD_k$) e o modelo que considera uma abordagem totalmente não-paramétrica (MG_k), mostrando que quando há a presença de covariáveis que ajudam a explicar a variável resposta, elas têm que ser incorporadas no modelo;

- com a distribuição a priori do limiar pouco informativa, em ambas as aplicações o valor que os dados escolheram para o limiar situou-se em um quantil baixo das observações, em torno de 15%. Como os dados que foram analisados foram máximos

mensais para a primeira e mínimos mensais para a segunda aplicação, a maioria destes valores já são observações extremas, fazendo parte da cauda;

- o modelo $MGPDR_k$ permite o cálculo de medidas importantes para cada configuração de covariável, sendo possível observar o comportamento do limiar ao longo do ano para uma particular cidade, além da mudança do limiar de acordo com latitude e altitude. Além disso, o modelo fornece uma distribuição de medidas extremas importantes, como o máximo e quantis altos, para particulares cidades em determinados meses do ano;

- estimar todas as cidades em diferentes meses do ano pelo modelo $MGPDR_k$ é mais eficiente do que estimar valores de uma única cidade em um único mês do ano pelo modelo $MPGD_k$. Na aplicação dos Estados Unidos, como as observações coletadas são de 13 anos, haveria apenas 13 observações por cidade e mês, mesmo agrupando por estações do ano haveria apenas 39 observações, o que é insuficiente para realizar a estimação pelo modelo $MPGD_k$.

Apêndice 4 - Algoritmo

O Algoritmo MCMC para a estimação dos parâmetros do modelo $MGPDR_k$ é dado por:

Algoritmo 3

Na s-ésima iteração, as cadeias dos parâmetros se movimentam para o passo $s+1$ da seguinte maneira:

- Amostrando β_ξ

  Para cada β_{ξ_i}, $i = 0, \ldots, k_\xi$, amostra-se $\beta^*_{\xi_i}$ da $N(\beta^{(s)}_{\xi_i}, V_{\beta_{\xi_i}})$. Se o vetor amostrado β^*_ξ satisfaz a restrição em (4.4) em conjunto com $\beta^{(s)}_\nu, \beta^{(s)}_u$. Caso não satisfaça, amostra-se o vetor β^*_ξ novamente até satisfazer a restrição.

Atualiza $\beta_\xi^{(s+1)} = \beta_\xi^*$ com probabilidade α_{β_ξ} onde

$$\alpha_{\beta_\xi} = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x})}{\pi(\tilde{\Theta}|\mathbf{x})},\right.$$

$\Theta^* = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, \beta_u^{(s)}, \beta_\nu^{(s)}, \beta_\xi^*)$ e $\tilde{\Theta} = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, \beta_u^{(s)}, \beta_\nu^{(s)}, \beta_\xi^{(s)})$.

- Amostrando β_ν

  Para cada β_{ν_i}, $i = 0, \ldots, k_\nu$, amostra-se $\beta_{\nu_i}^*$ da $N(\beta_{\nu_i}^{(s)}, V_{\beta_{\nu_i}})$. Se o vetor amostrado β_ν^* satisfaz a restrição em (4.4) em conjunto com $\beta_\xi^{(s+1)}, \beta_u^{(s)}$. Caso não satisfaça, amostra-se o vetor β_ν^* novamente até satisfazer a restrição.

  Atualiza $\beta_\nu^{(s+1)} = \beta_\nu^*$ com probabilidade α_{β_ν} onde

$$\alpha_{\beta_\nu} = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x})}{\pi(\tilde{\Theta}|\mathbf{x})},\right.$$

  $\Theta^* = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, \beta_u^{(s)}, \beta_\nu^*, \beta_\xi^{(s+1)})$ e $\tilde{\Theta} = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, \beta_u^{(s)}, \beta_\nu^{(s)}, \beta_\xi^{(s+1)})$.

- Amostrando β_u

  Para cada β_{u_i}, $i = 0, \ldots, k_u$, amostra-se $\beta_{u_i}^*$ da $N(\beta_{u_i}^{(s)}, V_{\beta_{u_i}})$. Se o vetor amostrado β_u^* satisfaz a restrição em (4.4) em conjunto com $\beta_\xi^{(s+1)}, \beta_\nu^{(s+1)}$. Caso não satisfaça, amostra-se o vetor β_u^* novamente até satisfazer a restrição.

  Atualiza $\beta_u^{(s+1)} = \beta_u^*$ com probabilidade α_{β_u} onde

$$\alpha_{\beta_u} = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x})}{\pi(\tilde{\Theta}|\mathbf{x})},\right.$$

  $\Theta^* = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, \beta_u^*, \beta_\nu^{(s+1)}, \beta_\xi^{(s+1)})$ e $\tilde{\Theta} = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, \beta_u^{(s)}, \beta_\nu^{(s+1)}, \beta_\xi^{(s+1)})$.

- Amostrando μ, η e $\mathbf{p}$

  Para aplicar o algoritmo de Metropolis para os parâmetros das componentes de mistura da parte central dos dados, o procedimento é igual ao do Algoritmo 1 do Capítulo 2.

Capítulo 5

Extremos com estrutura de modelos dinâmicos

Em séries temporais é estudado como o comportamento dos dados pode se alterar com o tempo. Este tipo de alteração é comum para dados de valores extremos.

Em dados ambientais, por exemplo, em chuva, vento e temperatura, seus níveis podem estar correlacionados com a sazonalidade, além de apresentar uma tendência de aumento ao longo dos anos, devido mudanças climáticas no planeta.

Em dados de finanças, um período de recessão altera todos os índices econômicos. Assim, investidores procuram mercados mais sólidos para aplicar seus investimentos.

Um trabalho recente, sendo outra extensão do trabalho do Behrens et al. (2004), foi desenvolvido por Lopes et al. (2009), onde o modelo utilizado é uma distribuição Gama com GPD, levando em consideração o fato dos dados serem provenientes de uma série temporal, onde os parâmetros da parte da cauda do modelo com distribuição GPD varia no tempo. Ao contrário do modelo do capítulo anterior, aqui, o modelo é feito por equações de atualização de um modelo linear dinâmico, ou *Dynamic Linear Model* (DLM). Huerta e Sansó (2007) usaram uma ideia similar quando modelaram níveis diários de ozônio por uma distribuição GEV variando no tempo e espaço.

No modelo proposto por Lopes et al. (2009), cada observação $x_t, t = 1, \ldots, n$, que

está abaixo de um limiar u, segue uma distribuição Gama, e cada observação x_t que está acima do limiar possui distribuição GPD. Neste modelo, dois dos parâmetros da GPD muda ao longo do tempo, enquanto os outros parâmetros permanecem fixos.

Neste modelo, a função distribuição de mistura é dada por

$$F_t(x_t|\mu,\eta,u_t,\xi_t,\sigma_t) = \begin{cases} F_G(x_t|\mu,\eta) & \text{se} \quad x_t < u, \\ F_G(u,\mu,\eta) + [1 - F_G(u,\mu,\eta)]G(x_t|u,\xi_t,\sigma_t) & \text{se} \quad x_t > u \end{cases}, \quad (5.1)$$

onde $F_G(.|\mu,\eta)$ é a função de distribuição da Gama e G a função de distribuição da GPD.

5.1 Dinâmica na cauda

O modelo DLM, visto em West e Harrison (1997), é usado aqui para modelar os parâmetros da GPD, que muda ao longo do tempo pela seguinte estrutura

$$\begin{pmatrix} \xi_t \\ \sigma_t \end{pmatrix} = F_t'\beta_t + v_t \text{ e}$$
$$\beta_t = G_t\beta_{t-1} + w_t. \quad (5.2)$$

Num caso particular da equação (5.2), pode considerar que os parâmetros da cauda ξ e σ sigam um modelo dinâmico de primeira ordem,

$$\begin{aligned} \xi_t &= \theta_{\xi,t} + v_{\xi,t} \qquad v_{\xi,t} \sim N(0, 1/V_\xi), \\ \theta_{\xi,t} &= \theta_{\xi,t-1} + w_{\xi,t} \quad w_{\xi,t} \sim N(0, 1/W_\xi), \\ \sigma_t &= \theta_{\sigma,t} + v_{\sigma,t} \qquad v_{\sigma,t} \sim N(0, 1/V_\sigma), \\ \theta_{\sigma,t} &= \theta_{\sigma,t-1} + w_{\sigma,t} \quad w_{\sigma,t} \sim N(0, 1/W_\sigma), \end{aligned} \quad (5.3)$$

onde $\theta_{\xi,t}, \theta_{\sigma,t}, t = 0, \ldots, n$, V_ξ, W_ξ, V_σ e W_σ são hiperparâmetros do modelo.

Porém, sabe-se que σ necessariamente é um valor positivo e Smith (1985) mostrou que estimadores de máxima verossimilhança para ξ não existem para $\xi < -1$. A forma dinâmica proposta em (5.3) permite que σ seja negativo e ξ seja menor do que -1. Foi

proposta uma reparametrização na cauda, com a dinâmica em $\sigma_t = \exp(l\sigma_t)$ e $\xi_t = \exp(l\xi_t) - 1$. Assim, o modelo dinâmico para os parâmetros da cauda são dados por

$$\begin{aligned} l\xi_t &= \theta_{\xi,t} + v_{\xi,t} \qquad v_{\xi,t} \sim N(0, 1/V_\xi), \\ \theta_{\xi,t} &= \theta_{\xi,t-1} + w_{\xi,t} \quad w_{\xi,t} \sim N(0, 1/W_\xi), \\ l\sigma_t &= \theta_{\sigma,t} + v_{\sigma,t} \qquad v_{\sigma,t} \sim N(0, 1/V_\sigma), \\ \theta_{\sigma,t} &= \theta_{\sigma,t-1} + w_{\sigma,t} \quad w_{\sigma,t} \sim N(0, 1/W_\sigma), \end{aligned} \tag{5.4}$$

5.2 Mistura de Gamas com GPD usando DLM

Uma outra extensão que pode ser feita consiste em utilizar um modelo que trabalha com mistura de k distribuições Gama para valores abaixo de um limiar u e uma distribuição GPD com estrutura DLM em ξ_t para valores acima de um limiar u. A função de distribuição do modelo, denominado por $MGPDLM_k$ é dada por

$$F_t(x_t|\theta, \Psi_t) = \begin{cases} \sum_{j=1}^{k} p_j F_G(x_t|\mu_j, \eta_j), & \text{se} \quad x_t < u \\ \sum_{j=1}^{k} p_j F_G(u \mid \mu_j, \eta_j) + \left[1 - \sum_{j=1}^{k} F_G(u, \mu_j, \eta_j)\right] G(x_t|\xi_t, \sigma_t, u), & \text{se} \quad x_t > u, \end{cases}$$

Em relação aos parâmetros da cauda, ao invés de estimar diretamente ξ_t e σ_t, os parâmetros estimados são $\sigma_t = \exp(l\sigma_t)$ e $\xi_t = \exp(l\xi_t) - 1$, considerando a estrutura dinâmica como em (5.4). Considerando independência entre os parâmetros da mistura e da cauda, e na cauda considerando apenas a dependência dos parâmetros do modelo dinâmico em (5.3), a distribuição a priori pode ser escrita da seguinte maneira:

$$\pi(\mu, \eta, \mathbf{p}, l\sigma, l\xi, u, \theta_\sigma, \theta_\xi, V_\sigma, V_\xi, W_\sigma, W_\xi) = \pi(\mu, \eta, \mathbf{p})\pi(u)\pi(l\sigma, \theta_\sigma, V_\sigma, W_\sigma)\pi(l\xi, \theta_\xi, V_\xi, W_\xi),$$

onde $\pi(\mu, \eta, \mathbf{p})$ e $\pi(u)$ possuem a mesma forma da distribuição a priori do Capítulo 3. A distribuição a priori dos parâmetros $\pi(l\xi, \theta_\xi, V_\xi, W_\xi)$ e $\pi(l\sigma, \theta_\sigma, V_\sigma, W_\sigma)$ pode ser escrita como

$$\pi(l\xi, \theta_\xi, V_\xi, W_\xi) = \prod_{t=1}^{n} \left(\pi(l\xi_t|\theta_{\xi,t}, V_\xi)\pi(\theta_{\xi,t}|\theta_{\xi,t-1}, W_\xi)\right) \pi(\theta_{\xi,0})\pi(V_\xi)\pi(W_\xi)$$

$$\pi(l\sigma, \theta_\sigma, V_\sigma, W_\sigma) \;=\; \prod_{t=1}^{n} \left(\pi(l\sigma_t|\theta_{\sigma,t}, V_\sigma)\pi(\theta_{\sigma,t}|\theta_{\sigma,t-1}, W_\sigma)\right) \pi(\theta_{0,\sigma})\pi(V_\sigma)\pi(W_\sigma)$$

onde $\xi_t|\theta_{\xi,t}, V_\xi \sim N(\theta_{\xi,t}, 1/V_\xi)$, $\theta_{\xi,t}|\theta_{\xi,t-1}, W_\xi \sim N(\theta_{\xi,t-1}, 1/W_\xi)$, $\theta_{\xi,0} \sim N(m_{\xi,0}, C_{\xi,0})$, $V_\xi \sim G(f_\xi/o_\xi, f_\xi)$, $W_\xi \sim G(l_\xi/m_\xi, l_\xi)$, $\sigma_t|\theta_{\sigma,t}, V_\sigma \sim N(\theta_{\sigma,t}, 1/V_\sigma)$, $\theta_{\sigma,t}|\theta_{\sigma,t-1}, W_\sigma \sim N(\theta_{\sigma,t-1}, 1/W_\sigma)$, $\theta_{\sigma,0} \sim N(m_{\sigma,0}, C_{\sigma,0})$, $V_\sigma \sim G(f_\sigma/o_\sigma, f_\sigma)$ e $W_\sigma \sim G(l_\sigma/m_\sigma, l_\sigma)$.

A escolha destas distribuições a priori, de acordo com a forma dinâmica em (5.4), tem como objetivo facilitar os cálculos das distribuições a posteriori para os parâmetros, sendo possível encontrar distribuições condicionais completas conhecidas para a maioria dos parâmetros, e realizar a estimação pelo amostrador de Gibbs.

Assim, baseado na função de verossimilhança e na distribuição a priori dos parâmetros, a função de densidade a posteriori pode ser escrita da seguinte maneira:

$$\begin{aligned}
\pi(\Theta|\mathbf{x}) \;\propto\; & \prod_{t:x_t<u} \left[\sum_{j=1}^{k} p_j f_G(x_t|\mu_j, \eta_j)\right] \prod_{x_t\geq u} \left[\left(1 - \sum_{j=1}^{k} p_j F_G(u|\mu_j, \eta_j)\right) g(x_t|\xi_t, \sigma_t, u)\right] \\
\times\; & \prod_{j=1}^{k} \left[\eta_j^{a_j-1} e^{-b_j\eta_j} \beta_j^{-(c_j+1)} e^{-d_j/\mu_j}\right] \exp\left(-\frac{(u-\mu_u)^2}{2\sigma_u^2}\right) \\
\times\; & V_\xi^{n/2+f_\xi-1} \exp\left(-\frac{V_\xi}{2}\sum_{t=1}^{n}(l\xi_t - \theta_{\xi,t})^2 - o_\xi V_\xi\right) \exp\left(-\frac{1}{2C_{\xi,0}}(\theta_{\xi,0} - m_{\xi,0})^2\right) \\
\times\; & W_\xi^{n/2+l_\xi-1} \exp\left(-\frac{W_\xi}{2}\sum_{t=1}^{n}(\theta_{\xi,t} - \theta_{\xi,t-1})^2 - m_\xi W_\xi\right) \\
\times\; & V_\sigma^{n/2+f_\sigma-1} \exp\left(-\frac{V_\sigma}{2}\sum_{t=1}^{n}(l\sigma_t - \theta_{\sigma,t})^2 - o_\sigma V_\sigma\right) \exp\left(-\frac{1}{2C_{\sigma,0}}(\theta_{\sigma,0} - m_{\sigma,0})^2\right) \\
\times\; & W_\sigma^{n/2+l_\sigma-1} \exp\left(-\frac{W_\sigma}{2}\sum_{t=1}^{n}(\theta_{\sigma,t} - \theta_{\sigma,t-1})^2 - m_\sigma W_\sigma\right). \qquad (5.5)
\end{aligned}$$

É possível encontrar uma forma conhecida para a distribuição condicional completa para a maioria dos parâmetros definidos no modelo linear dinâmico em (5.4). A única exceção é para $l\xi_t$ e $l\sigma_t$ quando $x_t > u$. Os cálculos das distribuições condicionais completas e o algoritmo para estimação dos parâmetros por MCMC estão no Apêndice 5.

5.3 Simulações

Foram realizadas simulações do modelo $MGPDLM_k$. Para cada simulação, foram gerados n pontos de uma distribuição $Gama(\eta, \eta/\mu)$ com parâmetros $\mu = 5$ e $\eta = 1$. O limiar foi tomado no quantil 80 dos valores simulados. Para os valores das observações maiores que o limiar, os valores foram substituídos por pontos gerados de uma distribuição GPD, cujos parâmetros foram gerados de acordo com uma estrutura dinâmica como na equação (5.4), onde $\theta_{\xi,0} = 0,2$, $\theta_{\sigma,0} = 2$. Em relação as precisões do modelo dinâmico, as simulações foram feitas em duas configurações. Na Configuração 1, $V_\xi = 200$, $W_\xi = 1000$, $V_\sigma = 200$ e $W_\sigma = 1000$. Na Configuração 2, $V_\xi = 2000$, $W_\xi = 10000$, $V_\sigma = 2000$ e $W_\sigma = 10000$

Os valores das distribuições a priori foram dados por $m_{\xi,0} = \theta_{\xi,0}$, $m_{\sigma,0} = \theta_{\sigma,0}$, $C_{\xi,0} = 100$, $C_{\sigma,0} = 100$. Para V_ξ, V_σ, W_ξ e W_σ foram dadas distribuições a priori Gama com média no valor verdadeiro e variância 10000 para a Configuração 1 e 1000000 para a Configuração 2.

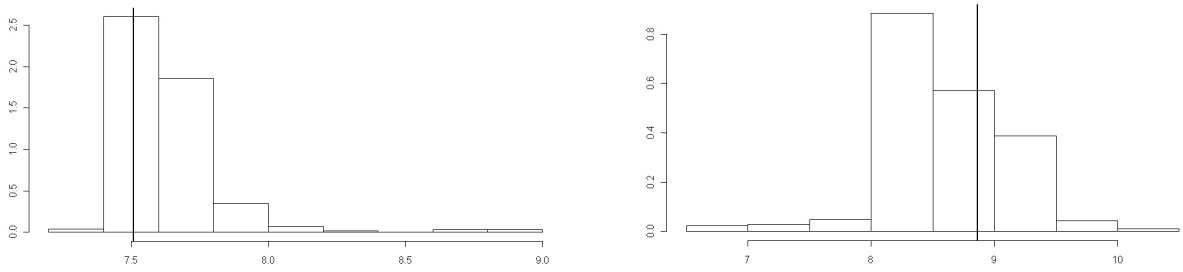
Para o limiar foi dada a priori uma distribuição Normal com média no valor verdadeiro e variância 10. Em relação ao tamanho da amostra, todas as simulações foram feitas utilizando três tamanhos diferentes: $n = 1000, 2500$ e 10000.

5.3.1 Simulações com amostras de tamanho 1000

A Figura 5.1 mostra a distribuição do limiar para $n = 1000$ nas duas configurações. Pela figura, percebe-se que o limiar está bem estimado, com a distribuição em torno do verdadeiro valor utilizado na simulação. A distribuição parece estar bem concentrada entre $7,2$ e $8,0$ na Configuração 1 e entre 7 e 10 na Configuração 2, bem mais concentrada que a distribuição a priori que tem variância 10.

A Figura 5.2 mostra o histograma da distribuição a posteriori de $\theta_{0,\xi}$ e $\theta_{0,\sigma}$ para $n = 1000$ nas duas configurações simuladas. Mesmo com a variância a priori sendo vaga ($C_{0,\xi}$ e $C_{0,\xi} = 100$), a estimação destes parâmetros foi precisa, com média próxima aos

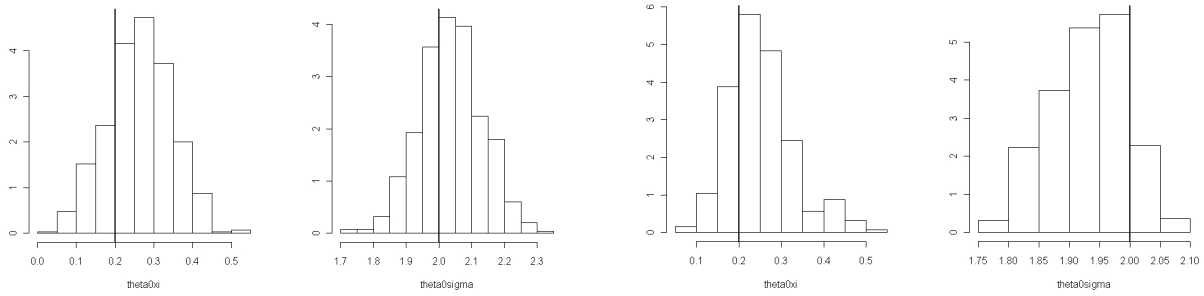
Figura 5.1: Histograma do limiar para simulações com $n = 1000$.
linha cheia: limiar verdadeiro



a) Configuração 1. b) Configuração 2.

valores verdadeiros.

Figura 5.2: Histograma para $\theta_{0,\xi}$ e $\theta_{0,\sigma}$ para $n = 1000$.
linha cheia: valor verdadeiro

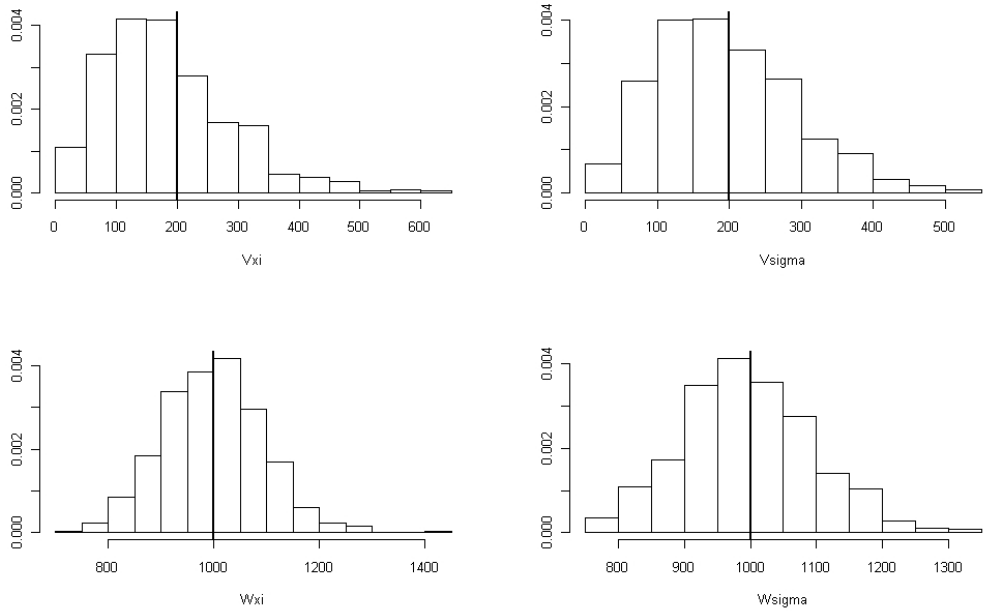


a) Configuração 1. b) Configuração 2.

As Figuras 5.3 e 5.4 mostram os histogramas da distribuição a posteriori de V_ξ, V_σ, W_ξ e W_σ para $n = 1000$ nas duas configurações simuladas. Pelas Figuras, a princípio percebe-se que a distribuição a posteriori tem média próxima ao verdadeiro valor utilizado na simulação. Porém, em ambos os casos a variância parece ser alta. Na Configuração 1, a variância a priori dos parâmetros é de 10000, ou seja, o desvio padrão é de 100 e um intervalo de credibilidade de 95% na priori está entre o valor verdadeiro mais ou menos 200.

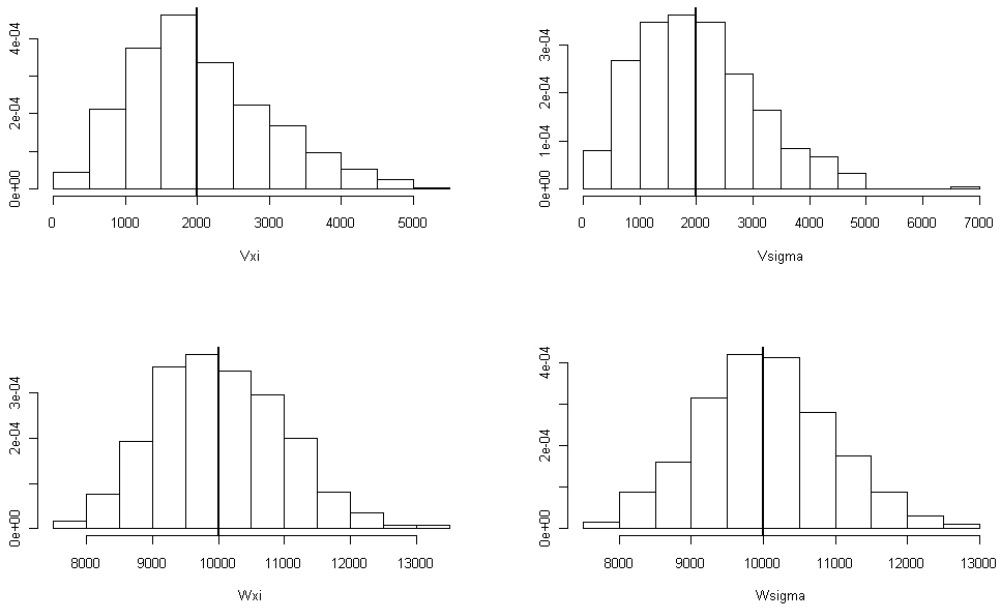
Pela Figura 5.3, o intervalo de credibilidade da distribuição a posteriori parece ter uma amplitude muito próxima da priori. Portanto, os dados não estão dando muita informação nestes parâmetros e quase toda a informação sobre estes parâmetros está compreendida na distribuição a priori. O mesmo caso acontece na Configuração 2, onde a variância a priori é de 1000000, com desvio padrão de 1000. Pela Figura 5.4, observa-se que a distribuição a posteriori está compreendida no valor verdadeiro mais ou menos 2000. Portanto, para $n = 1000$ os dados não trazem informação a respeito das variâncias do modelo dinâmico.

Figura 5.3: Histograma de V_ξ, V_σ, W_ξ e W_σ na Configuração 1 para $n = 1000$.
linha cheia: valor verdadeiro.



As Figuras 5.5 e 5.6 mostram o histograma da média a posteriori com as bandas de credibilidade de 95% para ξ, σ, θ_ξ e θ_σ para $n = 1000$ nas duas configurações simuladas. Pela Figura 5.5, percebe-se que, na Configuração 1, a estimação de ξ e $l\xi$ acompanha bem o verdadeiro valor dos parâmetros ao longo do tempo, estando quase todos os valores dentro do intervalo de credibilidade. Em relação a σ e $l\sigma$, a estimação não é eficiente, onde na maior parte do tempo, os parâmetros verdadeiros estão abaixo do intervalo de

Figura 5.4: Histograma de V_ξ, V_σ, W_ξ e W_σ na Configuração 2 para $n = 1000$.
linha cheia: valor verdadeiro.

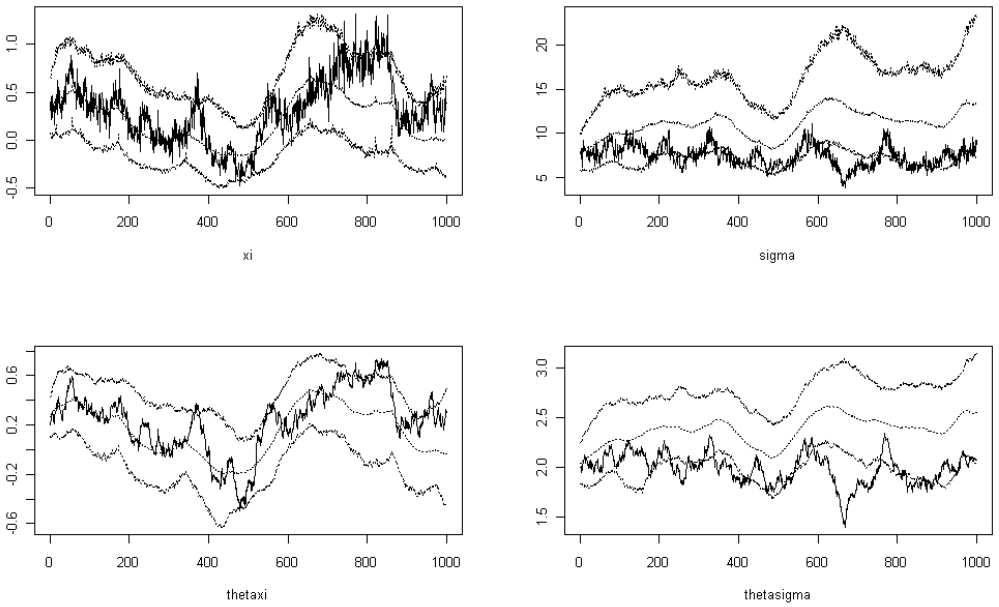


credibilidade estimado. Na Configuração 2, pela Figura 5.6, em todas as situações a estimação não consegue estimar as diferenças ao longo do tempo, ou seja, embora há grandes altas e baixas nos parâmetros ao longo do tempo, a estimação sobe ou desce de valor de uma maneira muito mais lenta que o valor verdadeiro. Portanto, para amostras de tamanho $n = 1000$, há uma dificuldade em estimar os parâmetros que evoluem no tempo.

5.3.2 Simulações com amostras de tamanho 2500

A Figura 5.7 mostra a distribuição do limiar para $n = 2500$ nas duas configurações. Pela Figura, percebe-se que a estimação é eficiente em detectar o verdadeiro valor. Em ambas as configurações, a distribuição a posteriori do limiar é mais concentrada do que quando $n = 1000$. Na Configuração 1, simulando com menor precisão o modelo linear dinâmico, a distribuição do limiar mostrou-se mais concentrada do que na Configuração

Figura 5.5: Estimação de ξ, σ, θ_ξ e θ_σ na Configuração 1 para $n = 1000$.
Linha cheia: valores verdadeiros. Linhas tracejadas: Média a posteriori e intervalo de credibilidade de 95%.



2, que considera menor variabilidade no modelo linear dinâmico.

A Figura 5.8 mostra o histograma da distribuição a posteriori de $\theta_{0,\xi}$ e $\theta_{0,\sigma}$ para $n = 2500$ nas duas configurações simuladas. Assim como para $n = 1000$, as distribuições a posteriori dos parâmetros estão centradas próximas ao verdadeiro valor utilizado na simulação. A amplitude da distribuição a posteriori é a mesma do que no caso em que $n = 1000$.

As Figuras 5.9 e 5.10 mostram o histograma da distribuição a posteriori de V_ξ, V_σ, W_ξ e W_σ para $n = 2500$ nas duas configurações simuladas. Pela Figura, são observados resultados muito próximos aos obtidos para quando $n = 1000$, ou seja, a variância a posteriori do limiar é praticamente a mesma da variância a priori. Assim quase toda a informação sobre estes quatro parâmetros é fornecida pela distribuição a priori, com os dados praticamente não dando informação alguma sobre estes parâmetros.

Figura 5.6: Estimação de ξ, σ, θ_ξ e θ_σ na Configuração 2 para $n = 1000$.
Linha cheia: valores verdadeiros. Linhas tracejadas: Média a posteriori e intervalo de credibilidade de 95%.

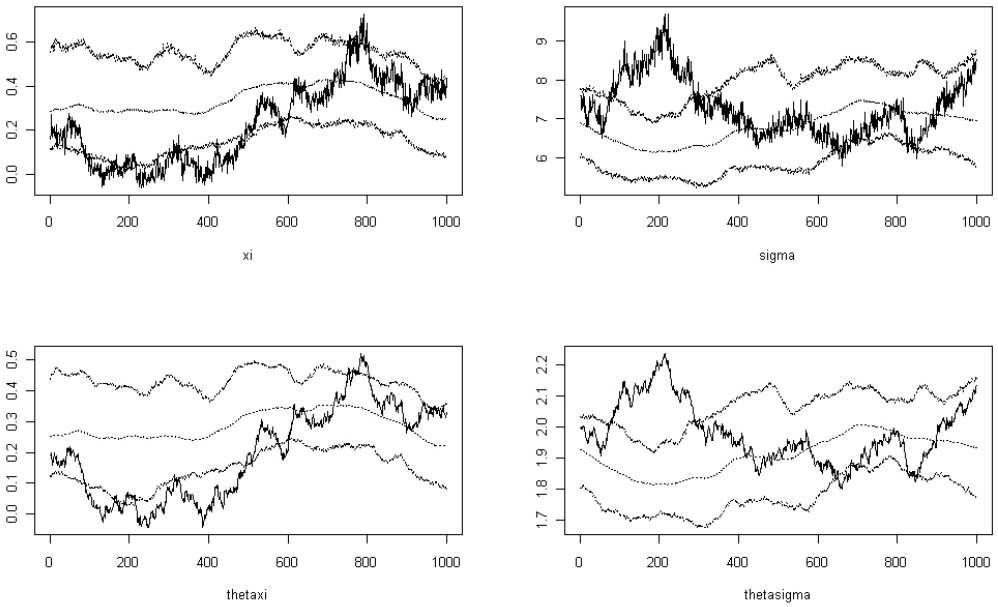
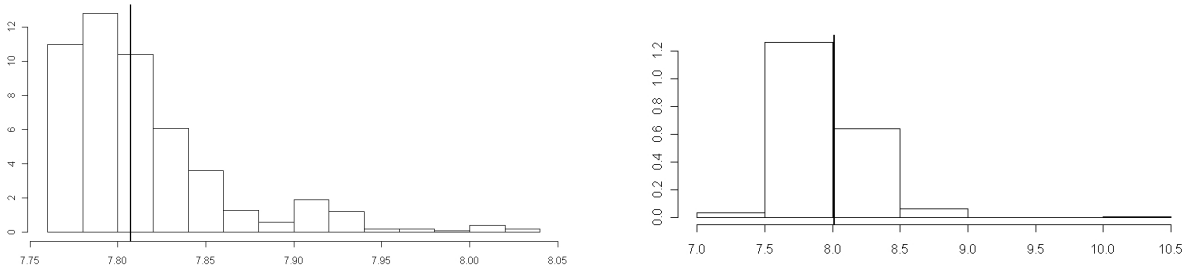


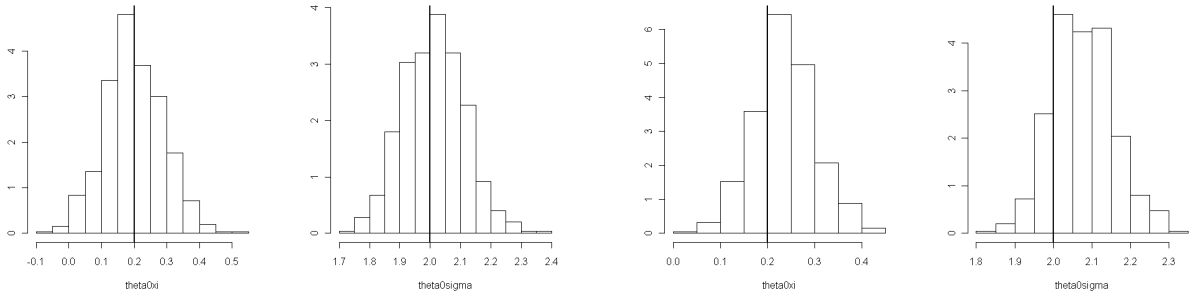
Figura 5.7: Histograma do limiar para simulações com $n = 2500$.
linha cheia: limiar verdadeiro



a) Configuração 1. b) Configuração 2.

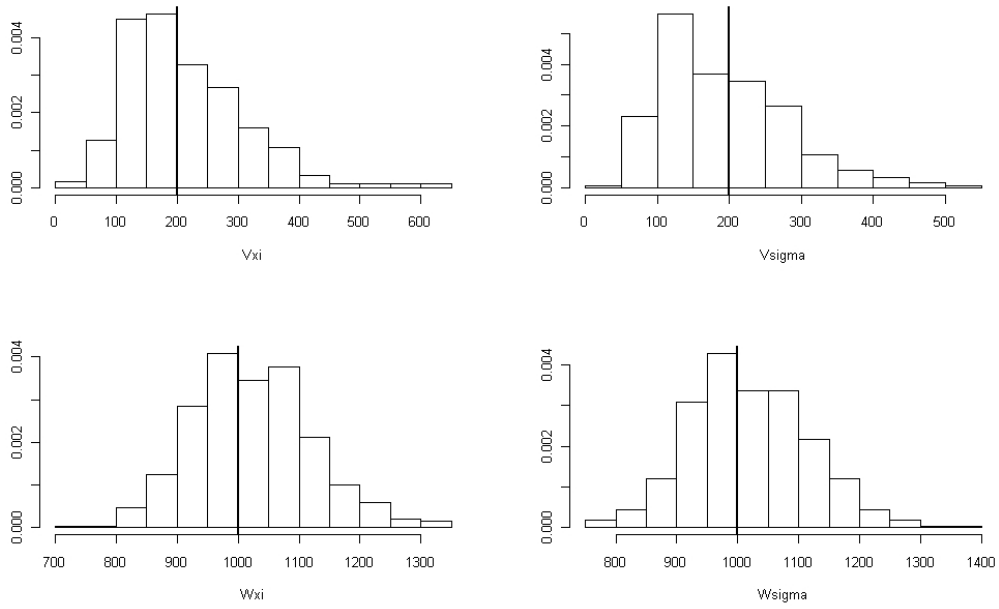
As Figuras 5.11 e 5.12 mostram o histograma da média a posteriori com as bandas de credibilidade de 95% para ξ, σ, θ_ξ e θ_σ para $n = 2500$ nas duas configurações simuladas.

Figura 5.8: Histograma para $\theta_{0,\xi}$ e $\theta_{0,\sigma}$ para $n = 2500$.
linha cheia: valor verdadeiro



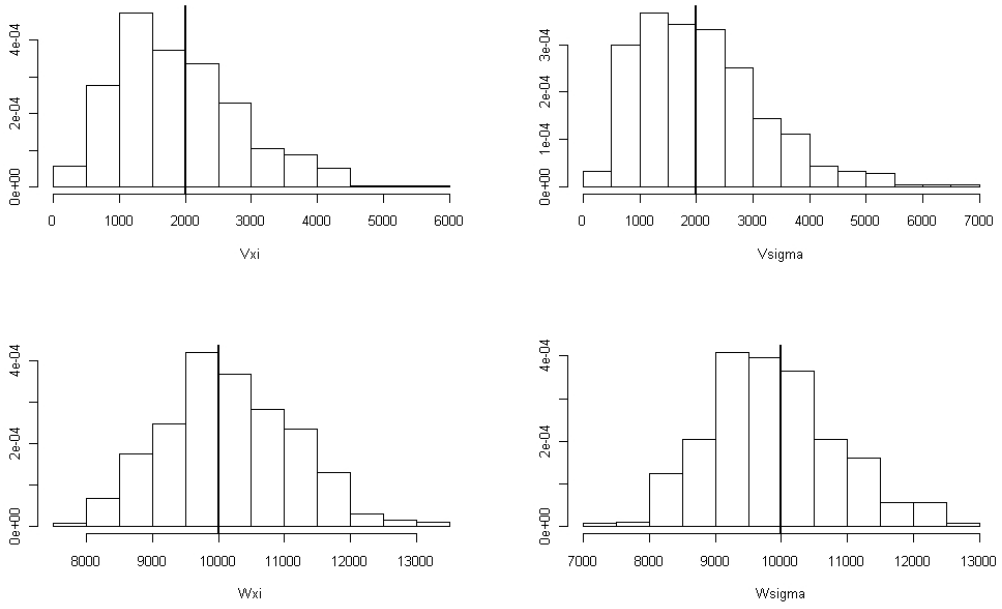
a) Configuração 1. b) Configuração 2.

Figura 5.9: Histograma de V_ξ, V_σ, W_ξ e W_σ na Configuração 1 para $n = 2500$.
linha cheia: valor verdadeiro.



Pelas Figuras, pode-se observar que houve uma melhora significativa da estimação dos parâmetros que evoluem no tempo em relação a estimação realizada nas simulações com $n = 1000$. Pela Figura 5.11, observa-se que na Configuração 1, quase todos os valores

Figura 5.10: Histograma de V_ξ, V_σ, W_ξ e W_σ na Configuração 2 para $n = 2500$.
linha cheia: valor verdadeiro.

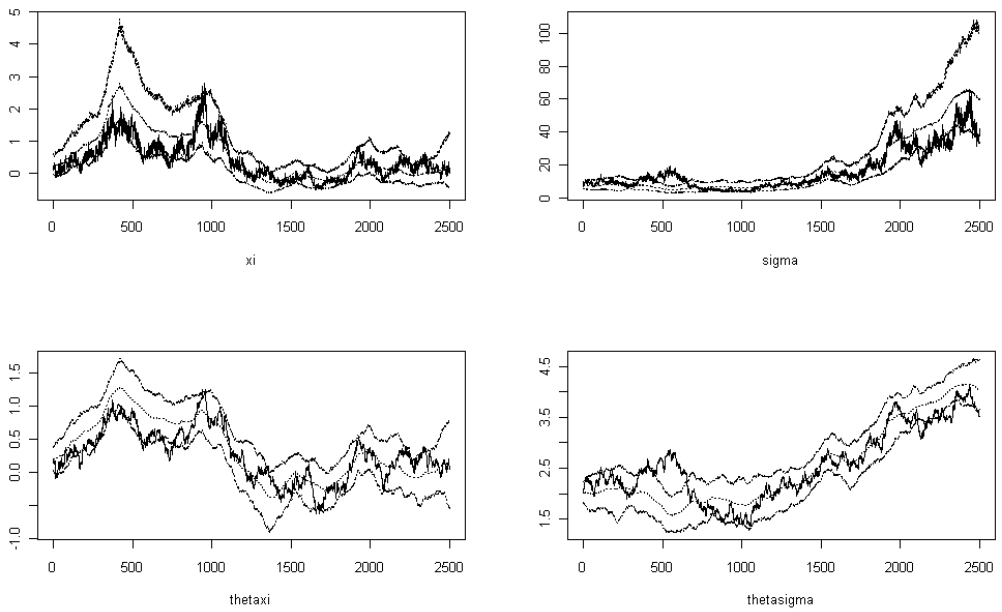


verdadeiros evoluindo no tempo está dentro do intervalo de credibilidade estimado, e a estimação acompanha o comportamento dos parâmetros. Na estimação de σ, por exemplo, os parâmetros parecem estar concentrados em valores menores do que 20 até a observação de 1700 e depois tem um aumento. A estimação consegue detectar este tipo de comportamento. Portanto, aumentando o tamanho da amostra, o maior número de informações nas observações ajuda a estimar o comportamento dos parâmetros que evoluem no tempo.

5.3.3 Simulações com amostras de tamanho 10000

A Figura 5.13 mostra a distribuição do limiar para $n = 10000$ nas duas configurações. Assim como houve uma melhora na precisão da estimação do limiar aumentando o tamanho da amostra de $n = 1000$ para $n = 2500$, aumentando de $n = 2500$ para $n = 10000$ a estimação do limiar se torna mais eficiente, com a distribuição se tornando muito precisa, bem concentrada em torno do valor verdadeiro utilizado nas simulações. Assim como

Figura 5.11: Estimação de ξ, σ, θ_ξ e θ_σ na Configuração 1 para $n = 2500$.
Linha cheia: valores verdadeiros. Linhas tracejadas: Média a posteriori e intervalo de credibilidade de 95%.



nas amostras anteriores, a precisão da distribuição é maior na Configuração 1 do que na Configuração 2.

A Figura 5.14 mostra o histograma da distribuição a posteriori de $\theta_{0,\xi}$ e $\theta_{0,\sigma}$ para $n = 10000$ nas duas configurações simuladas. Percebe-se pela Figura que a distribuição a posteriori para os parâmetros $\theta_{0,\xi}$ e $\theta_{0,\sigma}$ é muito próxima a distribuição a posteriori para as simulações com $n = 1000$ e $n = 2500$ nas duas configurações, com exceção da distribuição de $\theta_{0,\xi}$ na Configuração 2, onde a distribuição está centrada num valor um pouco menor que o verdadeiro, embora o verdadeiro valor parece estar dentro de um intervalo de credibilidade de 95% da distribuição.

As Figuras 5.15 e 5.16 mostram o histograma da média a posteriori com as bandas de credibilidade de 95% para ξ, σ, θ_ξ e θ_σ para $n = 10000$ nas duas configurações simuladas. Em ambas as Figuras, pode ser observado que a estimação acompanha bem o

Figura 5.12: Estimação de ξ, σ, θ_ξ e θ_σ na Configuração 2 para $n = 2500$.
Linha cheia: valores verdadeiros. Linhas tracejadas: Média a posteriori e intervalo de credibilidade de 95%.

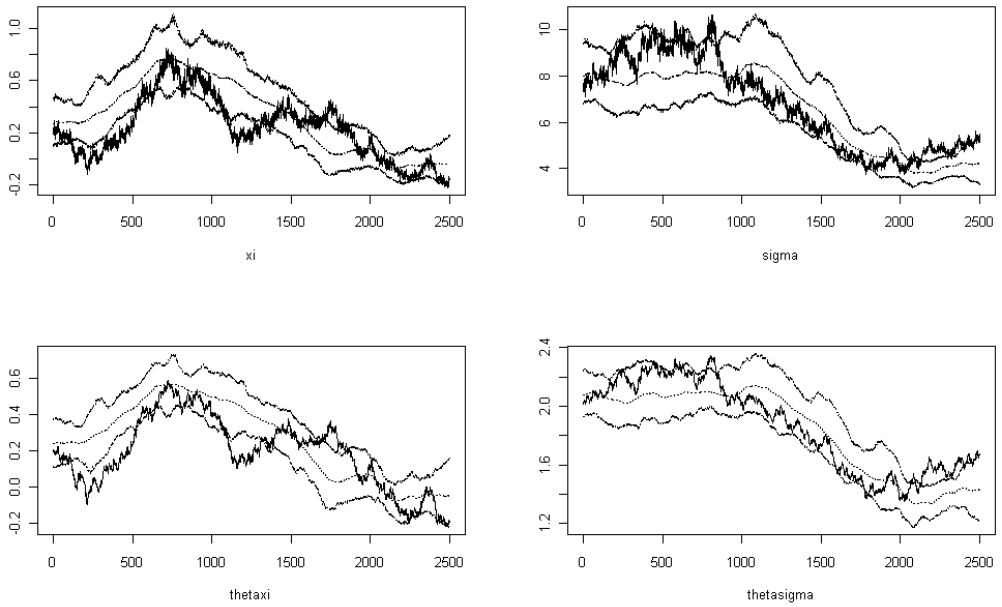
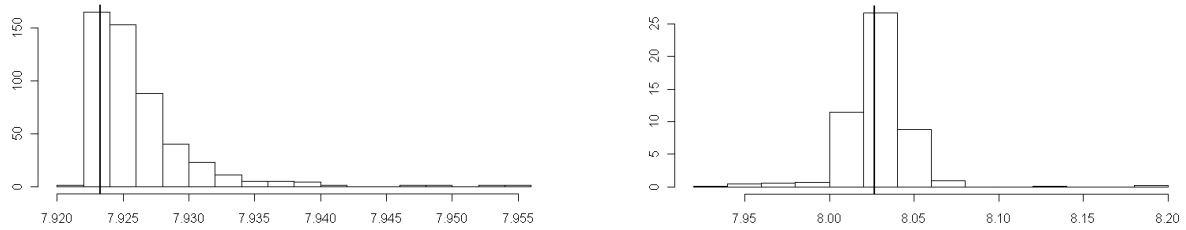


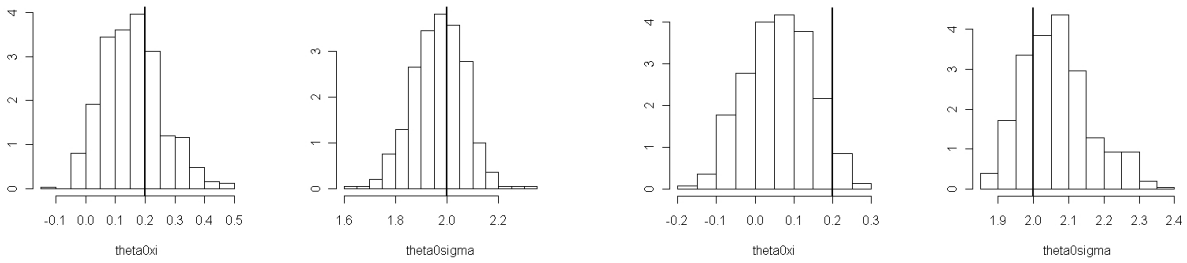
Figura 5.13: Histograma do limiar para simulações com $n = 10000$.
linha cheia: limiar verdadeiro



a) Configuração 1. b) Configuração 2.

comportamento dos valores verdadeiros ao longo do tempo, com intervalos concentrados próximos ao verdadeiro valor. Comparando as duas configurações, percebe-se que na

Figura 5.14: Histograma para $\theta_{0,\xi}$ e $\theta_{0,\sigma}$ para $n = 10000$.
linha cheia: valor verdadeiro



a) Configuração 1. b) Configuração 2.

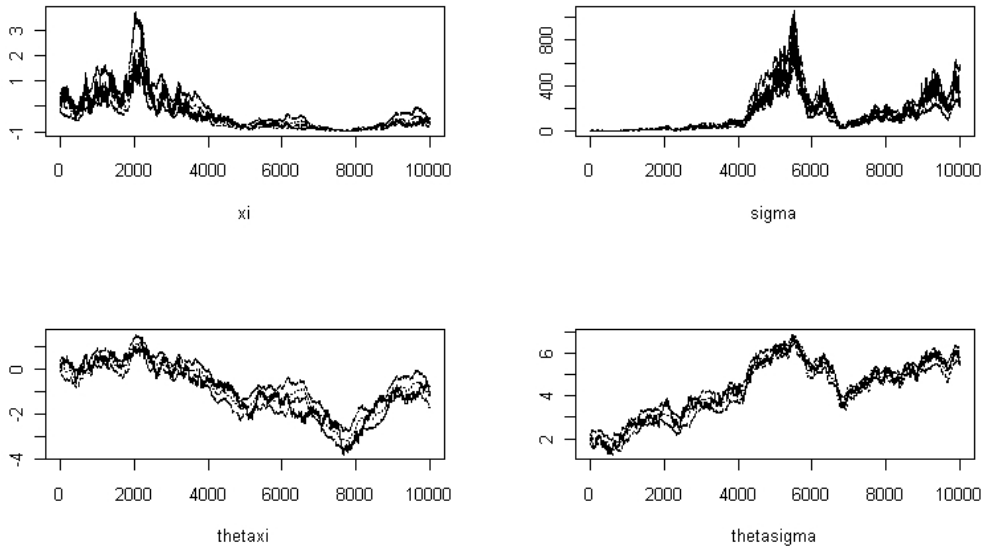
Configuração 1 (Figura 5.15), como a variância utilizada no modelo dinâmico foi maior, a simulação gerou valores muito altos para σ, com valores maiores do que 800. A estimação acompanhou bem esta grande mudança deste parâmetro ao longo do tempo. Portanto, para $n = 10000$, o modelo consegue identificar bem as mudanças dos parâmetros da cauda evoluindo no tempo.

5.3.4 Cálculo de máximos e quantis altos

Nesta seção, será mostrada a estimação de estatísticas importantes em valores extremos, como quantis altos, além dos máximos quando $\xi < 0$. A distribuição dos quantis pode ser obtida da mesma forma do que na equação (3.9) do Capítulo 3. A maior diferença é que neste capítulo os parâmetros da cauda ξ e σ evoluem no tempo, e assim para cada tempo $t = 1, \ldots, n$ o valor do quantil será diferente.

As Figura 5.17 mostra respectivamente os valores verdadeiros e estimados do quantil 95% para uma simulação com $n = 1000$ e de 99% para uma simulação com $n = 2500$. Pela Figura, pode-se observar que, para $n = 1000$, o valor estimado do quantil 95% é próximo ao verdadeiro valor deste quantil, em torno de 20 ao longo do tempo. Assim como ocorrido na estimação do parâmetro ξ e σ, a estimação do quantil não consegue captar

Figura 5.15: Estimação de ξ, σ, θ_ξ e θ_σ na Configuração 1 para $n = 10000$.
Linha cheia: valores verdadeiros. Linhas tracejadas: Média a posteriori e intervalo de credibilidade de 95%.



mudanças significativas ao longo do tempo, estimando de maneira suavizada a evolução deste quantil. Para $n = 2500$, na estimação do quantil 99%, pode-se notar que o quantil verdadeiro aumenta para observações entre 500 e 750 e decai entre 750 e 1200. Estas são as observações em que ocorrem os maiores valores na amostra, com muitos pontos bem acima do limiar. A estimação do quantil mostra-se eficiente, acompanhando por uma curva suavizada o comportamento de subida e descida do quantil para as observações entre 500 e 1000.

Além de calcular quantis extremos, quando ξ é negativo, a distribuição GPD possui um máximo, que é dado por $u - \sigma/\xi$. Na abordagem por modelos dinâmicos, como ξ varia no tempo, pode acontecer de ter valores onde ξ é negativo e valores onde ξ é positivo, sendo possível estimar nos máximos em observações do primeiro tipo.

A Figura 5.18 mostra a distribuição dos máximos para uma simulação com $n = 10000$ para os pontos onde $\xi_t < -0,10$, $t = 1, \ldots, 10000$. Foi observado graficamente apenas os

Figura 5.16: Estimação de ξ, σ, θ_ξ e θ_σ na Configuração 2 para $n = 10000$.
Linha cheia: valores verdadeiros. Linhas tracejadas: Média a posteriori e intervalo de credibilidade de 95%.

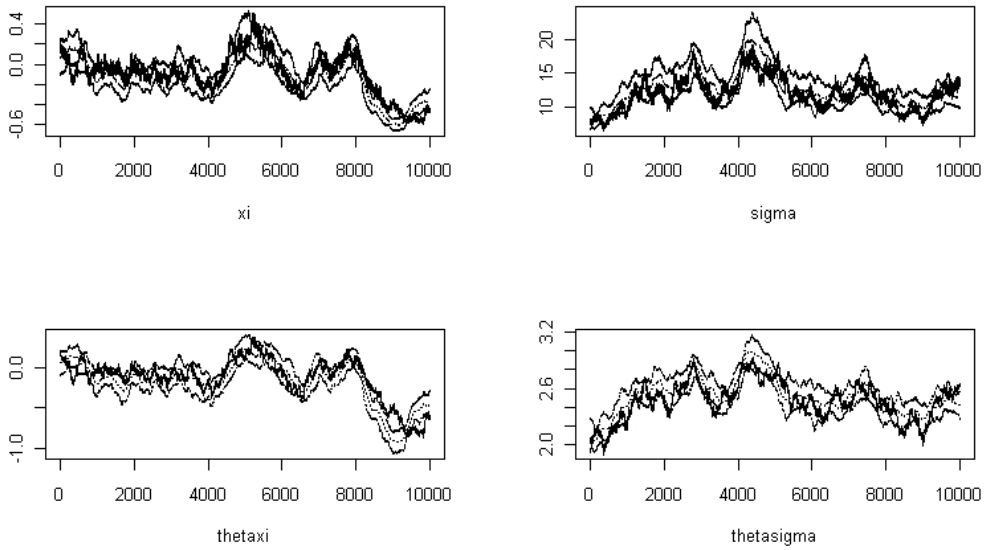
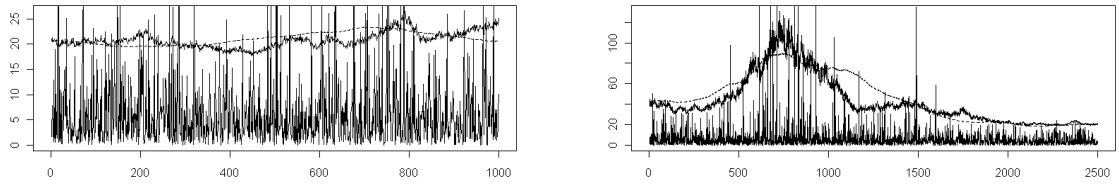


Figura 5.17: Série das observações com quantis extremos nas simulações.
linha cheia acima: quantil verdadeiro. Linha tracejada: média a posteriori do quantil.



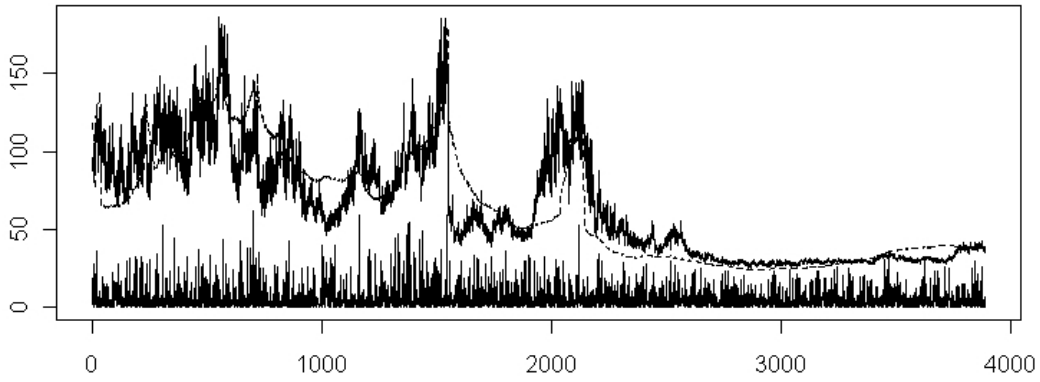
a) $n = 1000$, quantil 95%.

b) $n = 2500$, quantil 99%.

valores onde $\xi_t < -0,10$ e não $\xi_t < 0$ pois o limite do máximo à esquerda quando ξ vai à 0 é infinito e acaba gerando valores muito altos do máximo, o que fica difícil visualizar graficamente em comparação com outros valores negativos de ξ. Pela figura, nota-se que há uma oscilação muito grande do verdadeiro máximo para observações entre 0 e 2200 no

gráfico e depois o máximo se estabiliza num valor menor. Este tipo de comportamento foi bem detectado pela estimação, onde na maioria das regiões o máximo estimado está próximo do máximo verdadeiro.

Figura 5.18: Série das observações com máximos nas simulações com $n = 10000$.
Linha cheia acima: máximo verdadeiro. Linha tracejada: média a posteriori do máximo.



5.3.5 Conclusões das simulações

Após realizar as simulações para o modelo $MGPDLM_k$, baseado nos resultados obtidos, pode-se tirar as seguintes conclusões:

- O modelo é eficiente em encontrar o verdadeiro limiar, mesmo com amostras de tamanho não muito grande ($n = 1000$). Foi dada uma priori não muito vaga para este parâmetro, com variância 10. Mas em todas as situações, a distribuição a posteriori além de ter uma média muito próxima do verdadeiro valor, teve uma variância muito mais precisa do que a da priori, indicando que os dados fornecem uma boa informação sobre limiar;

- os dados forneceram pouca informação a respeito da precisão da estrutura de modelos dinâmicos, sendo que a distribuição destes parâmetros é praticamente a mesma dada na distribuição a priori. Numa situação de dados reais, será necessário escolher por algum critério qual a distribuição a priori mais plausível para estes parâmetros, baseado nos resultados das simulações;

- a estimação dos parâmetros da cauda, que evoluem no tempo, se torna mais eficiente a medida em que aumenta o tamanho da amostra. Os resultados de simulações mostraram que a estimação acompanha muito bem os parâmetros ao longo do tempo para $n = 10000$, enquanto que para $n = 1000$ há uma dificuldade maior em detectar a evolução destes parâmetros;

- o modelo mostrou-se eficiente na detecção de quantis extremos e dos máximos, principalmente para tamanho de amostras grandes, pois estas medidas estão em função dos parâmetros da cauda, que foram mais bem estimadas para $n = 2500$ e $n = 10000$.

5.4 Aplicações

Em aplicações de mercado financeiro, o nível de retorno de aplicações varia de comportamento de acordo com a situação da economia mundial. Em 2008, o mundo foi abalado por uma forte crise financeira, que provocou uma volatilidade maior nos níveis de retorno de ações, cotações de moeda, matérias primas, entre outros. O objetivo desta seção é utilizar o modelo proposto neste capítulo para estimar níveis de retorno, em porcentagem, de algumas ações considerando uma variação ao longo do tempo. Dado o valor da ação x_{t-1} no tempo $t-1$ e o valor x_t no tempo t, o retorno no tempo t é dado por

$$r_t = \log(x_t/x_{t-1}).$$

Na distribuição GPD, como trabalha-se com valores positivos, os dados que serão analisados é o valor em módulo dos retornos.

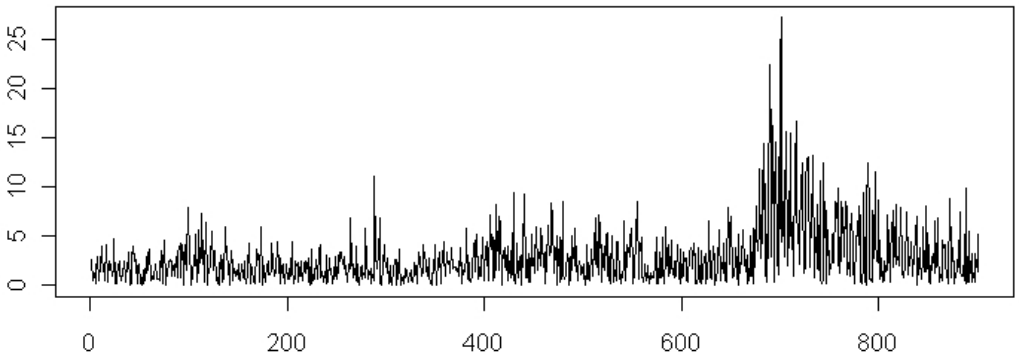
Os valores das distribuições a priori foram dados por $m_{\xi,0} = 0.2$, $m_{\sigma,0} = 0.2$, $C_{\xi,0} = 100$, $C_{\sigma,0} = 100$. Para o limiar, a distribuição a priori dada foi com média no quantil 90 dos dados de cada aplicação, com variância 10. Para V_ξ, V_σ, W_ξ e W_σ foram dadas distribuições a priori Gama com média e variância iguais a da Configuração 2, com maior precisão no modelo dinâmico. Foi escolhida a Configuração 2 pelo fato de que, nas simulações, utilizar uma precisão mais baixa pode indicar valores absurdamente grandes para os parâmetros da distribuição GPD, como por exemplo pode-se observar na Figura 5.15, onde houve valores de σ maiores do que 800 e menores que 5 numa mesma série. Na prática, espera-se uma oscilação menor destes parâmetros.

5.4.1 Aplicação 1 - Vale

Foram analisados os níveis de retorno absoluto das ações da companhia de mineração Vale do Rio Doce, num período de 1 de abril de 2006 a 8 de março de 2009, num total de 901 observações. A Figura 5.19 mostra os retornos das ações no período estudado. Observando a figura, nota-se que há uma oscilação maior no final da série, que corresponde a crise mundial que houve no segundo semestre de 2008 e abalou a economia brasileira. Em épocas de crise, espera-se uma grande oscilação nas ações, com grandes perdas seguidos por grandes ganhos em dias sucessivos. Um fator interessante que será visto a seguir é analisar como este período de crise influencia nos parâmetros da distribuição do modelo proposto.

Para verificar a adequação do modelo nestes dados, foi calculada a medida de ajuste DIC, comparando o modelo proposto neste capítulo com dois dos modelos dos capítulos anteriores. Pelo fato do número de parâmetros ser muito maior por ter vários parâmetros que evoluem no tempo, o BIC torna-se inviável como medida de comparação, pois dá um peso muito grande no número de parâmetros do modelo. No modelo $MGPDLM_k$, há $(3k + 7 + 4n)$ parâmetros, e o BIC dá um peso igual para todos estes parâmetros para

Figura 5.19: Série dos dados da Vale.



calcular a medida. O resultado do DIC para os vários modelos é dado na Tabela 5.1. Pela tabela, observa-se que o melhor modelo é o $MGPDLM_2$, mostrando que este modelo é superior a utilizar uma abordagem totalmente não-paramétrica por mistura de Gamas, ou fazer uma distribuição GPD para a cauda com parâmetros fixos.

Tabela 5.1: DIC para os dados da Vale

Modelo	$MGPDLM_1$	$MGPDLM_2$	$MGPDLM_3$	MG_2*	$MGPD_1*$
Pd	6,81	16,12	3,51	2,07	3,54
DIC	3609	**3492**	3669	3651	3637

MG_k* e $MGPD_k*$ são os melhores modelos de cada classe de acordo com o DIC.

A Tabela 5.2 mostra a média a posteriori dos parâmetros do modelo $MGPDLM_2$, com intervalo de credibilidade de 95%. Pela tabela, nota-se que para a distribuição com valores menores que um limiar, os dados são ajustados por uma distribuição Gama com média $2,76$ e variância $6,75$. Em relação a $\theta_{0,\xi}$ e $\theta_{0,\sigma}$, a distribuição encontrada está em torno de 0 com desvio padrão de aproximadamente $0,03$. Ainda pela Tabela 5.2, em

relação às variâncias da estrutura dinâmica, a média e a variância da distribuição são praticamente os mesmos dados na distribuição a priori, ou seja, os dados fornecem pouca informação a respeito dos parâmetros, como já foi mostrado anteriormente nos resultados das simulações.

Tabela 5.2: Média e intervalos de credibilidade para a aplicação da Vale.

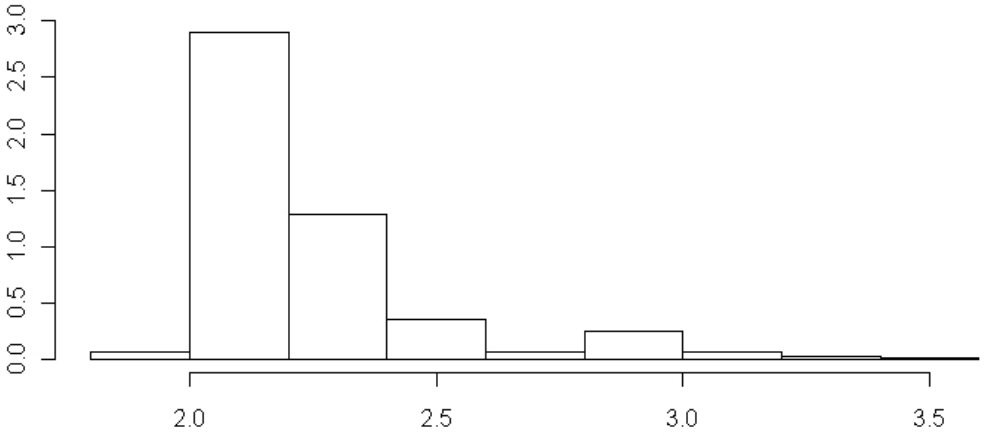
Parâmetro	μ	η	$\theta_{0,\xi}$	$\theta_{0,\sigma}$
Média	2,76	1,13	-0,01	0,02
Intervalo	(2,52;3,06)	(1,01;1,27)	(-0,07;0,05)	(-0,05;0,10)
Parâmetro	V_ξ	V_σ	W_ξ	W_σ
Média	1971	2693	9656	9791
Intervalo	(476;4317)	(725;5313)	(7750;11706)	(7943;11753)

A Figura 5.20 mostra a distribuição a posteriori do limiar. A distribuição tem uma média de $2,26$ e variância $0,05$. Nesta aplicação, a distribuição a priori do limiar foi uma Normal com média $6,05$ e variância 10, ou seja, houve pouca contribuição da priori na distribuição a posteriori.

Embora nas simulações, amostras de tamanho 1000 não foram eficientes na estimação dos parâmetros da cauda, nesta aplicação, como cerca de metade das observações estão na cauda, equivale as simulações com $n = 2500$, pois nas simulações a cauda corresponde a 20% das observações, ou seja, 500 observações na cauda, valor próximo ao número de observações na cauda desta aplicação.

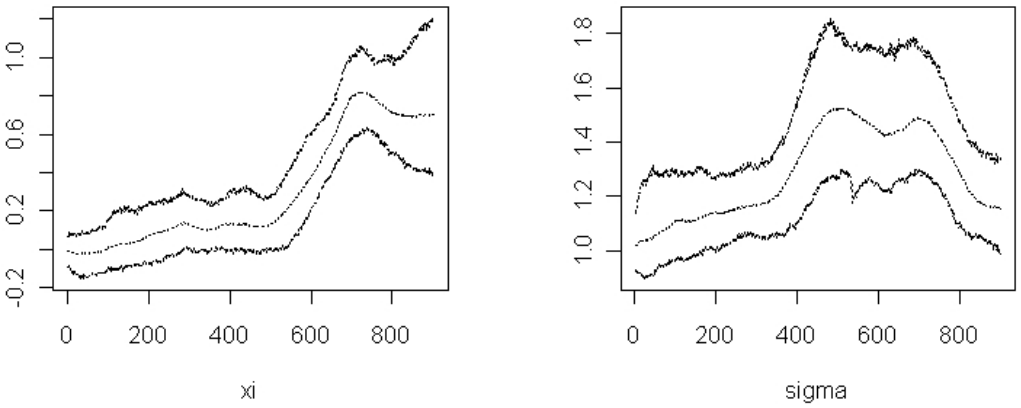
A Figura 5.21 mostra a banda de credibilidade de 95% para os parâmetros da cauda. Observando a Figura, nota-se que para ξ há um aumento no final das observações. Comparando com a Figura 5.19, é próxima a região onde há maior oscilação das ações. Portanto, no período da crise, o parâmetro ξ é maior do que em outros períodos. Um comportamento semelhante também se observa para σ, porém parece haver um aumento deste valor num período ainda antes da crise. Depois do período da crise, que são as últimas observações,

Figura 5.20: Distribuição a posteriori do limiar para os dados da Vale.



há uma diminuição no valor de σ.

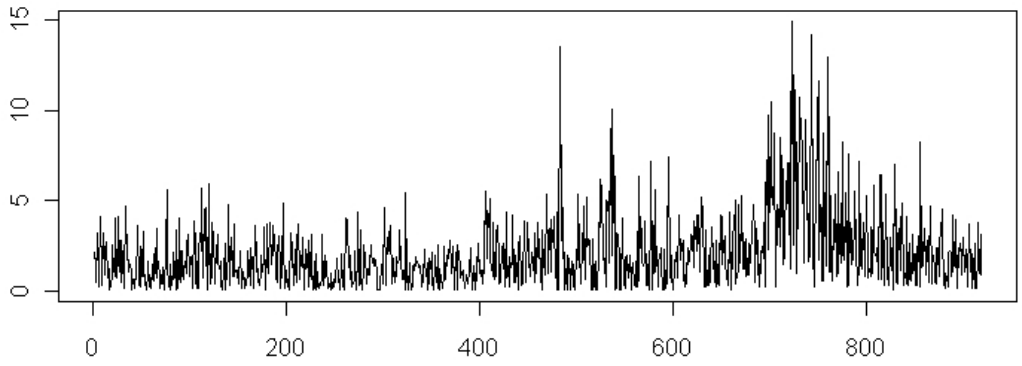
Figura 5.21: Intervalos de credibilidade de 95% da cauda dos dados da Vale.



5.4.2 Aplicação 2 - Petrobras

Assim como nos dados da aplicação anterior, foram analisados os níveis de retorno absoluto, agora da companhia Petrobras, no período de 1 de março de 2006 a 8 de março de 2009, num total de 919 observações. A Figura 5.22 mostra a série com os retornos das ações. Da mesma maneira que nas observações da Vale, há um período no final da série em que há uma oscilação maior nos retornos, que é o período da crise do segundo semestre de 2008.

Figura 5.22: Série dos dados da Petrobras.



A Tabela 5.3 mostra a medida de ajuste DIC para diferentes modelos estudados. Pelo modelo, observa-se que o melhor segundo o DIC foi o $MGPDLM_1$. Porém, há uma diferença pequena em relação ao modelo $MGPD_1$. Neste caso, estimar os parâmetros da cauda variando no tempo traz uma vantagem, embora não muito grande, em relação ao modelo que considera os parâmetros fixos.

A Tabela 5.4 mostra a média a posteriori dos parâmetros do modelo, com intervalo de credibilidade de 95%. Pela tabela, percebe-se que os parâmetros utilizados no modelo dinâmico têm valores muito parecidos com os obtidos na aplicação nas ações da Vale. Com relação aos valores antes da cauda, a distribuição Gama tem uma média de $2,39$ e

Tabela 5.3: DIC para os dados da Petrobras

Modelo	$MGPDLM_1$	$MGPDLM_2$	MG_2*	$MGPD_1*$
Pd	7,90	16,79	4,51	1,48
DIC	**2996**	3061	3106	3008

MG_k* e $MGPD_k*$ são os melhores modelos de cada classe de acordo com o DIC.

variância de $7,14$.

Tabela 5.4: Média e intervalos de credibilidade para a aplicação da Petrobras.

Parâmetro	μ	η	$\theta_{0,\xi}$	$\theta_{0,\sigma}$
Média	2,39	0,80	0,01	0,01
Intervalo	(2,10;2,72)	(0,72;0,88)	(-0,05;0,06)	(-0,05;0,08)
Parâmetro	V_ξ	V_σ	W_ξ	W_σ
Média	2016	2935	9882	9707
Intervalo	(540;4376)	(911;6426)	(7981;11848)	(7975;11672)

A Figura 5.23 mostra a distribuição a posteriori do limiar. Para os dados da Petrobras o limiar tem uma distribuição com média de $1,7$ e uma precisão maior que no limiar da Vale. Nesta aplicação, distribuição a priori do limiar foi uma Normal com média $4,46$ e variância 10.

A Figura 5.24 mostra a banda de credibilidade de 95% para os parâmetros da cauda. Observando conjuntamente as Figuras 5.22 dos dados e 5.24 dos parâmetros da cauda, percebe-se que tanto para ξ quanto para σ, os valores de ambos os parâmetros aumentam no período da crise, e diminuem novamente nas últimas observações, após a crise, já no ano de 2009. Portanto, o modelo detecta bem as alterações que ocorrem no mercado financeiro em diversos períodos de tempo.

Figura 5.23: Distribuição a posteriori do limiar para os dados da Petrobras.

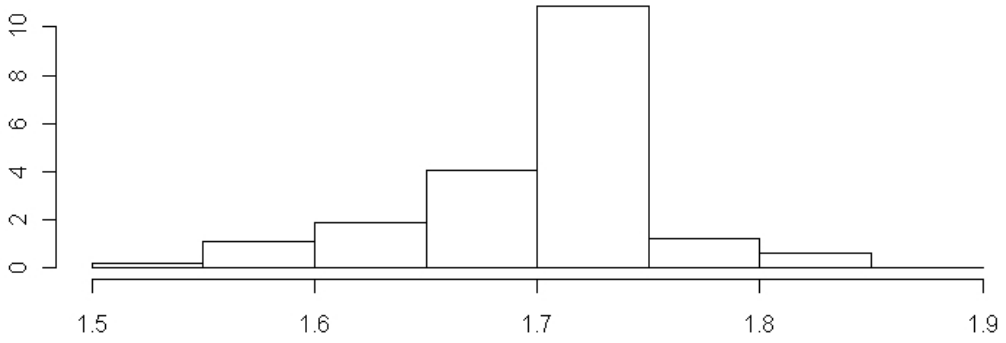
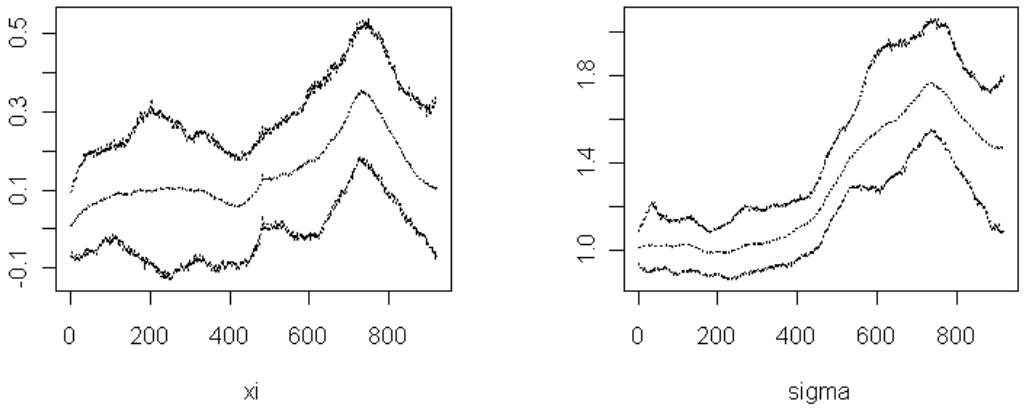


Figura 5.24: Intervalos de credibilidade de 95% da cauda dos dados da Petrobras.

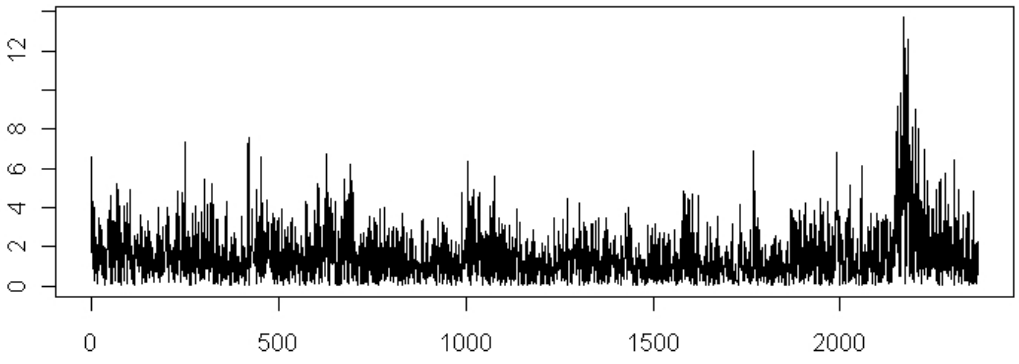


5.4.3 Aplicação 3 - BOVESPA

Nesta aplicação, foram analisados os níveis de retorno absoluto, para o índice da Bolsa de valores de São Paulo, BOVESPA, no período de 1 de abril de 2000 a 8 de março de 2009, num total de 2369 observações. A Figura 5.25 mostra os retornos das

ações, apontando também uma mudança no comportamento da série no período relativo ao segundo semestre de 2008.

Figura 5.25: Série dos dados da BOVESPA.



A Tabela 5.5 mostra o resultado do DIC para esta aplicação, pela tabela, observa-se que há um ganho em utilizar a estrutura de modelos dinâmicos na cauda da distribuição.

Tabela 5.5: DIC para os dados da BOVESPA

Modelo	$MGPDLM_1$	$MGPDLM_2$	MG_2*	$MGPD_1*$
Pd	11,92	21,77	2,11	0,81
DIC	**6495**	6591	6684	6683

MG_k* e $MGPD_k*$ são os melhores modelos de cada classe de acordo com o DIC.

A Tabela 5.6 mostra a estimação da média a posteriori e intervalo de credibilidade de 95%. Observa-se pela tabela que antes da cauda, a distribuição Gama possui uma média de $1,69$ e variância de $2,48$, significativamente menor do que nas duas aplicações anteriores. Em relação aos parâmetros do modelo dinâmico, percebe-se que a estimação de V_σ e W_σ apresentou uma precisão muito maior do que aquela apontada da distribuição

a priori, indicando que para este parâmetro, pode haver pouca mudança do parâmetro σ ao longo do tempo.

Tabela 5.6: Média e intervalos de credibilidade para a aplicação da BOVESPA.

Parâmetro	μ	η	$\theta_{0,\xi}$	$\theta_{0,\sigma}$
Média	1,69	1,15	0,01	0,01
Intervalo	(1,57;1,83)	(1,06;1,24)	(-0,05;0,06)	(-0,04;0,07)
Parâmetro	V_ξ	V_σ	W_ξ	W_σ
Média	1885	24961	9787	18620
Intervalo	(867;4317)	(14135;37800)	(7893;12083)	(14414;22986)

A Figura 5.26 mostra a distribuição a posteriori do limiar. Percebe-se que o limiar é menor do que nas duas aplicações anteriores. Como o índice BOVESPA é dado por uma média entre muitas companhias, este índice acaba sendo menos volátil em relação a uma ação de uma única companhia, principalmente em períodos de crise, onde a série da Figura 5.25 apresenta valores variando entre 0% e 12%, enquanto que na Vale (Figura 5.19) os valores variam entre 0% e 25% e na Petrobras (Figura 5.22) entre 0% e 15%. A distribuição a priori do limiar foi uma Normal com média $3,20$ e variância 10.

A Figura 5.27 mostra a banda de credibilidade de 95% para os parâmetros da cauda. Percebe-se que, para σ, por um período de tempo que corresponde a quase metade das observações, o valor deste parâmetro é praticamente constante ao longo do tempo. Isto influenciou no valor dos parâmetros V_σ e W_σ do modelo dinâmico, pois como não tem alteração em σ, o modelo dinâmico não tem evolução, ficando igual ao valor no tempo anterior com uma variação muito baixa. Em relação aos meses da crise, percebe-se que ambos os parâmetros aumentam neste período do tempo.

Figura 5.26: Distribuição a posteriori do limiar para os dados da BOVESPA.

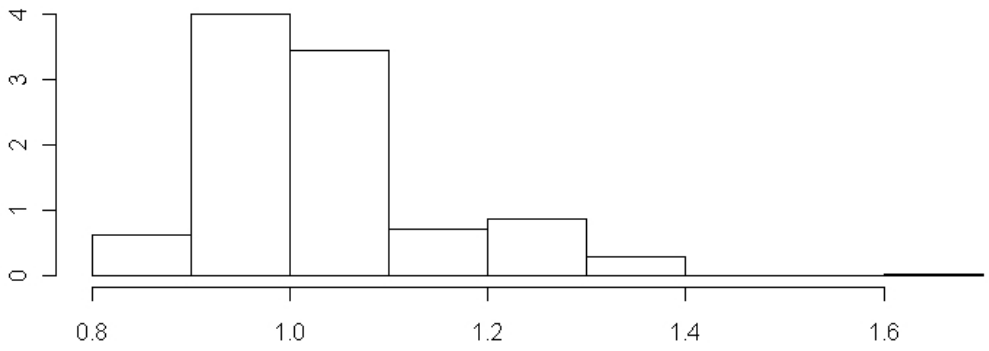
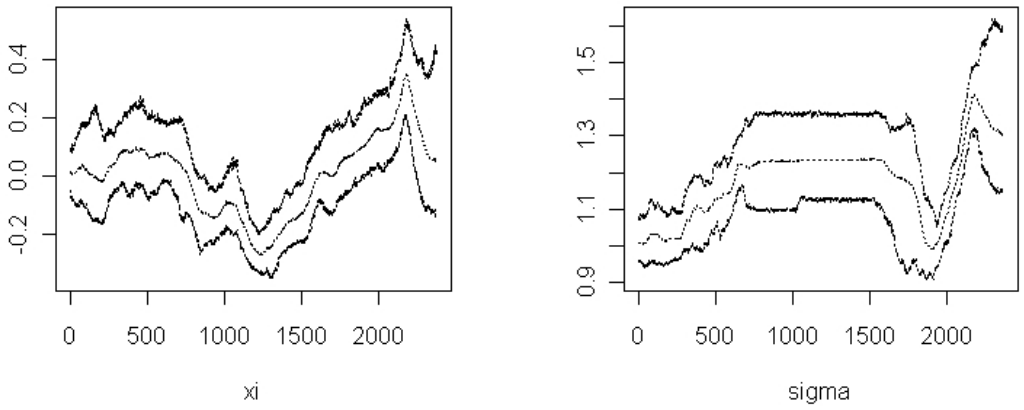


Figura 5.27: Intervalos de credibilidade de 95% da cauda dos dados da BOVESPA.



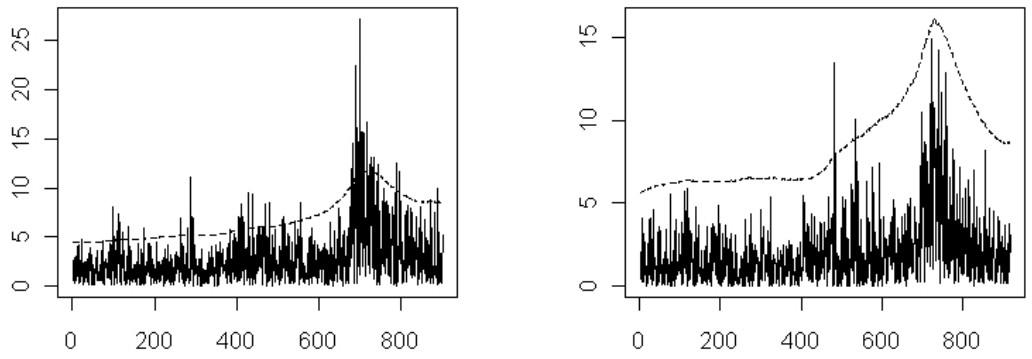
5.4.4 Cálculo de máximos e quantis altos

Após realizar a estimação dos parâmetros nas três aplicações estudadas, são encontradas outras medidas, baseadas nesta estimação. Uma pessoa que trabalha no mercado financeiro ou aplica parte de seu patrimônio em ações, tem interesse em saber qual o seu

risco de ter uma perda grande ou a probabilidade de ter um lucro elevado. Para isso, é interessante fornecer a probabilidade de se obter um retorno maior do que uma determinada quantidade, que são dados pela estimação dos quantis altos da distribuição. Além dos quantis, também pôde ser calculado o valor do máximo, quando o parâmetro ξ_t possui um valor negativo.

A Figura 5.28 mostra a média a posteriori do quantil 95% da Vale e do quantil 99% da Petrobras evoluindo no tempo. Pela figura, percebe-se que o quantil evolui no tempo junto com as observações, ou seja, no período de tempo onde os retornos são maiores, os quantis também aumentam. Nos dados da Vale este aumento não é muito grande, sendo que boa parte dos dados está acima do quantil no período da crise, mas na Petrobras o quantil 99% parece aumentar na mesma proporção das observações no período da crise.

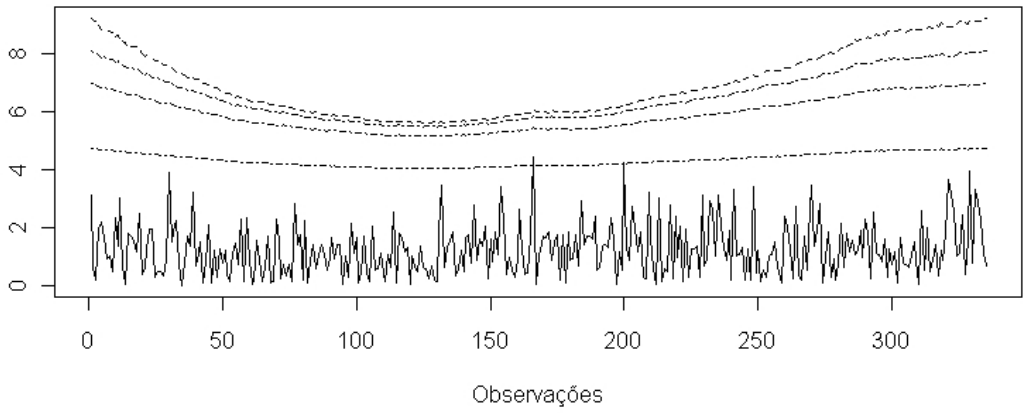
Figura 5.28: Série dos dados da Vale e Petrobras com quantis.
À direita, vale com quantil 95%. À direita, vale com quantil 99%. As linhas tracejadas representam a média a posteriori do quantil.



Em relação aos máximos, a Figura 5.29 mostra a estimação dos máximos para as observações da BOVESPA, nas observações onde a média a posteriori de ξ é menor do que $-0,10$, acompanhado do quantis elevados. Percebe-se pela figura que o máximo se situa

em um valor muito acima das observações, e a medida em que os quantis aumentam, a média a posteriori destes quantis se aproxima da média a posteriori do máximo. Tomando o limite do quantil indo a 100% a média deste limite coincide com a média do máximo.

Figura 5.29: Série dos dados da BOVESPA com máximos e quantis.
As linhas tracejadas representam os quantis 99%, 99,99%, 99,9999% e o máximo.



5.4.5 Conclusões das aplicações

Após modelar dados de retorno de ações pelo modelo $MGPDLM_k$ proposto neste capítulo, pode-se tirar as seguintes conclusões:

- Os parâmetros da cauda têm uma evolução dinâmica ao longo do tempo, com os parâmetros assumindo valores maiores em períodos onde há maiores níveis de retornos. Os resultados do DIC mostraram que nas três aplicações há um ganho no ajuste do que se fosse considerado todos os parâmetros fixos;

- o modelo estima bem o limiar, mesmo com uma informação relativamente vaga a respeito deste parâmetro, a posteriori teve uma distribuição com uma variância muito menor a da priori, com médias significativamente diferentes;

- a estimação dos parâmetros do modelo permite calcular de maneira eficiente medidas importantes de retornos financeiros, como quantis elevados e máximos.

Apêndice 5 - Condicionais completas e algoritmo

Baseado na equação (5.5), pode-se encontrar as seguintes funções condicionais completas:

- Condicional em $l\xi_t$

  Se $x_t < u$, então a distribuição condicional completa em $l\xi_t$ é dada por

$$\begin{aligned} \pi(l\xi_t \mid \Theta_{-l\xi_t}) &\propto \exp\left(-\frac{V_\xi}{2}(l\xi_t - \theta_{\xi,t})^2\right). \\ \text{Logo, } l\xi_t|\Theta_{-l\xi_t} &\sim N(\theta_{\xi,t}, 1/V_\xi) \end{aligned}$$

  onde $\Theta_{-l\xi_t}$ é o vetor com todos os parâmetros do modelo, exceto $l\xi_t$.

- Amostrando V_ξ

  De acordo com a equação (5.5), pode-se encontrar a seguinte distribuição condicional completa para V_ξ

$$\begin{aligned} \pi(V_\xi, \Theta_{-V_\xi}) &\propto V_\xi^{n/2+f_\xi-1} \exp\left(-V_\xi\left(\frac{\sum_{t=1}^n (l\xi_t - \theta_{\xi,t})^2}{2} + o_\xi\right)\right). \qquad (5.6) \\ \text{Logo, } V_\xi|\Theta_{-V_\xi} &\sim G\left(\frac{\frac{n}{2}+f_\xi}{\frac{\sum_{t=1}^n (l\xi_t-\theta_{\xi,t})^2}{2}+o_\xi}, \frac{n}{2}+f_\xi\right), \end{aligned}$$

  onde Θ_{-V_ξ} é o vetor com todos os parâmetros do modelo, exceto V_ξ.

- Amostrando W_ξ

  De acordo com a equação (5.5), pode-se encontrar a seguinte distribuição condicional completa para W_ξ

$$\begin{aligned}
\pi(W_\xi, \Theta_{-W_\xi}) &\propto V_\xi^{n/2+l_\xi-1} \exp\left(-W_\xi\left(\frac{\sum_{t=1}^{n}(\theta_{\xi,t}-\theta_{\xi,t-1})^2}{2}+m_\xi\right)\right).\\
\text{Logo, } W_\xi|\Theta_{-W_\xi} &\sim G\left(\frac{\frac{n}{2}+l_\xi}{\frac{\sum_{t=1}^{n}(\theta_{\xi,t}-\theta_{\xi,t-1})^2}{2}+m_\xi}, \frac{n}{2}+l_\xi\right),
\end{aligned}$$

onde Θ_{-W_ξ} é o vetor com todos os parâmetros, exceto W_ξ.

- Amostrando $\theta_{\xi,0}$

  De acordo com a equação (5.5), pode-se encontrar a seguinte distribuição condicional completa para $\theta_{\xi,0}$

$$\begin{aligned}
\pi(\theta_{\xi,0}|\Theta_{-\theta_{\xi,0}}) &\propto \exp\left(-\frac{W_\xi}{2}(\theta_{\xi,1}-\theta_{\xi,0})^2-\frac{1}{2C_{\xi,0}}(\theta_{\xi,0}-m_{\xi,0})^2\right)\\
&\propto \exp\left(-\frac{1}{2}\left[\theta_{\xi,0}^2\left(W_\xi+\frac{1}{C_{\xi,0}}\right)-2\theta_{\xi,0}\left(W_\xi\theta_{\xi,1}+\frac{m_{\xi,0}}{C_{\xi,0}}\right)\right]\right)\\
&= \exp\left(-\frac{1}{2}\left(W_\xi+\frac{1}{C_{\xi,0}}\right)\left[\theta_{\xi,0}^2-2\left(W_\xi\theta_{\xi,1}+\frac{m_{\xi,0}}{C_{\xi,0}}\right)/\left(W_\xi+\frac{1}{C_{\xi,0}}\right)\right]\right).\\
\text{Logo, } \theta_{\xi,0}|\Theta_{-\theta_{\xi,0}} &\sim N\left(\left(W_\xi\theta_1+\frac{m_{\xi,0}}{C_{\xi,0}}\right)/\left(W_\xi+\frac{1}{C_{\xi,0}}\right), 1/\left(W_\xi+\frac{1}{C_{\xi,0}}\right)\right),
\end{aligned}$$

  onde $\Theta_{-\theta_{\xi,0}}$ é o vetor com todos os parâmetros, exceto θ_0.

- Amostrando $\theta_{\xi,t}$, de acordo com a equação (5.5), tem-se

  Para $t = 1, \ldots, n-1$,

$$\begin{aligned}
\pi(\theta_{\xi,t}|\Theta_{-\theta_{\xi,t}}) &\propto \exp\left(-\frac{V_\xi}{2}(l\xi_t-\theta_{\xi,t})^2-\frac{W_\xi}{2}(\theta_{\xi,t+1}-\theta_{\xi,t})^2-\frac{W_\xi}{2}(\theta_{\xi,t}-\theta_{\xi,t-1})^2\right)\\
&\propto \exp\left(-\frac{1}{2}\left[\theta_{\xi,t}^2\left(V_\xi+2W_\xi\right)-2\theta_{\xi,n}\left(W(\theta_{\xi,t+1}+\theta_{\xi,t-1})+V_\xi l\xi_t\right)\right]\right)\\
&= \exp\left(-\frac{1}{2}\left(2W_\xi+V_\xi\right)\left[\theta_{\xi,t}^2-2\left(W_\xi(\theta_{\xi,t-1}+\theta_{\xi,t+1})+V_\xi l\xi_t\right)/\left(2W_\xi+V_\xi\right)\right]\right).\\
\text{Logo, } \theta_{\xi,t}|\Theta_{-\theta_{\xi,t}} &\sim N\left(\left(W_\xi(\theta_{\xi,t-1}+\theta_{\xi,t+1})+V_\xi l\xi_t\right)/\left(2W_\xi+V_\xi\right), 1/\left(2W_\xi+V_\xi\right)\right),
\end{aligned}$$

  onde $\Theta_{-\theta_{\xi,t}}$ é o vetor com todos os parâmetros, exceto $\theta_{\xi,t}$, $t = 1, \ldots, n-1$.

- Amostrando $\theta_{\xi,n}$

  De acordo com a equação (5.5), pode-se encontrar a seguinte distribuição condicional completa para $\theta_{\xi,n}$

$$\begin{aligned}
\pi(\theta_{\xi,n}|\Theta_{-\theta_{\xi,n}}) &\propto \exp\left(-\frac{V_\xi}{2}(l\xi_n-\theta_{\xi,n})^2-\frac{W_\xi}{2}(\theta_{\xi,n}-\theta_{\xi,n-1})^2\right)\\
&\propto \exp\left(-\frac{1}{2}\left[\theta_{\xi,n}^2\left(V_\xi+W_\xi\right)-2\theta_{\xi,n}\left(W_\xi\theta_{\xi,n-1}+V_\xi l\xi_n\right)\right]\right)\\
&= \exp\left(-\frac{1}{2}\left(W_\xi+V_\xi\right)\left[\theta_{\xi,n}^2-2\left(W_\xi\theta_{\xi,n-1}+V_\xi l\xi_n\right)/\left(W_\xi+V_\xi\right)\right]\right).\\
\text{Logo, } \theta_{\xi,n}|\Theta_{-\theta_{\xi,n}} &\sim N\left(\left(W_\xi\theta_{\xi,n-1}+V_\xi l\xi_n\right)/\left(W_\xi+V_\xi\right),1/\left(W_\xi+V_\xi\right)\right),
\end{aligned}$$

  onde $\Theta_{-\theta_{\xi,n}}$ é o vetor com todos os parâmetros, exceto $\theta_{\xi,n}$.

Para os parâmetros $l\sigma_t$, V_σ, W_σ e $\theta_{\sigma,t}$, t=0,...,n, o cálculo das distribuições condicionais completas é análogo aos parâmetros que envolvem ξ.

Para os outros parâmetros, não é possível encontrar uma forma conhecida para as distribuições condicionais completas das suas distribuições a posteriori. O algoritmo MCMC de estimação das distribuições a posteriori dos parâmetros deste modelo é dado na seguinte forma:

Algoritmo 4

Dados os valores dos parâmetros até a iteração s, novos valores da cadeia são gerados da seguinte maneira:

- Amostrando $\{l\xi_t\}_{t=1}^n$

  Para um valor t entre 1 e n, se $x_t < u^{(s)}$, tem-se a forma conhecida da distribuição condicional completa, e o parâmetro é atualizado por: $l\xi_t^{(s+1)} \sim N(\theta_{\xi,t}^{(s)}, 1/V_\xi^{(s)})$. Agora, se $x_t \geq u^{(s)}$ sua condicional completa não tem forma conhecida, sendo necessário amostrar $l\xi_t$ pelo algoritmo de Metropolis. Um candidato para $l\xi_t^*$ é a Normal truncada $N(l\xi_t^{(s)}, K_{\xi,t})I(\xi_U, \infty)$, onde $\xi_U = \log(-\sigma_t^{(s)}/(x_t - u^{(s)}) + 1)$,

$\sigma_t^{(s)} = \exp(l\sigma_t^{(s)})$. Assim, $l\xi_t^{(s+1)} = l\xi_t^*$ com probabilidade $\alpha_{l\xi}$, onde

$$\alpha_{l\xi} = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x})\Phi((l\xi_t^{(s)} - \xi_U)/\sqrt{K_{\xi,t}}))}{\pi(\tilde{\Theta}|\mathbf{x})\Phi((l\xi_t^* - \xi_U)/\sqrt{K_{\xi,t}}))}\right\},$$

$\Theta^* = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^{(s)}, l\sigma^{(s)}, l\xi_{<t}^{(s+1)}, l\xi_t^*, l\xi_{>t}^{(s)})$ e $\tilde{\Theta} = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^{(s)}, l\sigma^{(s)}, l\xi_{<t}^{(s+1)}, l\xi_{\geq t}^{(s)})$.

- Amostrando $\{l\sigma_t\}_{t=1}^n$

  Para um valor t entre 1 e n, se $x_t < u^{(s)}$, tem-se a forma conhecida da distribuição condicional completa, e o parâmetro é atualizado por: $l\sigma_t^{(s+1)} \sim N(\theta_{\sigma,t}^{(s)}, 1/V_\sigma^{(s)})$. Agora, se $x_t \geq u^{(s)}$ sua condicional completa não tem forma conhecida e é necessário amostrar $l\sigma_t$ pelo algoritmo de Metropolis.

  Se $\xi_t^{(s+1)} = \exp(l\xi_t^{(s+1)}) - 1 > 0$, amostra-se $l\sigma_t^*$ de uma $N(l\sigma_t^{(s)}, K_{\sigma,t})$.

  Assim, $l\sigma_t^{(s+1)} = l\sigma_t^*$ com probabilidade $\alpha_{l\sigma}$, onde

  $$\alpha_{l\sigma} = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x})}{\pi(\tilde{\Theta}|\mathbf{x})}\right\}.$$

  Se $\xi_t^{(s+1)} < 0$, um candidato para $l\sigma_t^*$é a Normal truncada $N(l\sigma_t^{(s)}, K_{\sigma,t})I(\sigma_U, \infty)$, onde $\sigma_U = \log(-\xi_t^{(s)}) + \log(x_t - u^{(s)})$. Assim, $l\sigma_t^{(s+1)} = l\sigma_t^*$ com probabilidade $\alpha_{l\sigma}$, com

  $$\alpha_{l\sigma} = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x})\Phi((l\sigma_t^{(s)} - \sigma_U)/\sqrt{K_{\sigma,t}}))}{\pi(\tilde{\Theta}|\mathbf{x})\Phi((l\sigma_t^* - \sigma_U)/\sqrt{K_{\sigma,t}}))}\right\}, \text{ onde}$$

  $\Theta^* = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^{(s)}, l\sigma_{<t}^{(s+1)}, l\sigma_t^*, l\sigma_{>t}^{(s)}, l\xi^{(s+1)})$ e $\tilde{\Theta} = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^{(s)}, l\sigma_{<t}^{(s+1)}, l\sigma_{\geq t}^{(s)}, l\xi^{(s+1)})$.

- Amostrando u

  O valor do limiar u^* é amostrado de uma distribuição $N(u^{(s)}, V_u)I(u_L^{(s)}, \infty)$, onde

  $$u_L^{(s)} = \max\left\{\min(x_1, \ldots, x_t), \max_{\{t:\xi_t^{(s+1)}<0, x_t>u^{(s)}\}}(x_t + \sigma^{(s+1)}/(\xi_t^{(s+1)}(1+\xi_t^{(s+1)})))\right\},$$

V_u é a variância da distribuição proposta para o limiar. Aceita-se $u^{(s+1)} = u^*$ com probabilidade α_u, onde

$$\alpha_u = \min\left\{1, \frac{\pi(\Theta^*|\mathbf{x})\Phi((u^{(s)} - u_L^{(s)})/\sqrt{V_u})}{\pi(\tilde{\Theta}|\mathbf{x})\Phi((u^* - u_L^{(s)})/\sqrt{V_u})}\right\},$$

$\Theta^* = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^*, l\sigma^{(s+1)}, l\xi^{(s+1)})$ e $\tilde{\Theta} = (\mu^{(s)}, \eta^{(s)}, \mathbf{p}^{(s)}, u^{(s)}, l\sigma^{(s+1)}, l\xi^{(s+1)})$.

- Amostrando $\mu, \eta, \mathbf{p}$

  Para amostrar os vetores μ, η e $\mathbf{p}$, o procedimento é igual ao Algoritmo 2.

- Amostrando V_ξ, W_ξ, $\theta_{\xi,t}$, V_σ, W_σ e $\theta_{\sigma,t}$, t=0,...,n.

  Pela forma da distribuição a posteriori em 5.5, pode-se encontrar distribuições condicionais completas conhecidas para o seguintes parâmetros

$$\begin{aligned}
V_\xi^{(s+1)} &\sim G\left(\frac{f_\xi + \frac{n}{2}}{o_\xi + \sum_{t=1}^n (l\xi_t^{(s+1)} - \theta_{\xi,t}^{(s)})^2}, f_\xi + \frac{n}{2}\right),\\
W_\xi^{(s+1)} &\sim G\left(\frac{l_\xi + \frac{n}{2}}{m_\xi + \sum_{t=1}^n (\theta_{\xi,t}^{(s)} - \theta_{\xi,t-1}^{(s)})^2}, l_\xi + \frac{n}{2}\right),\\
\theta_{\xi,0}^{(s+1)} &\sim N\left(\frac{W_\xi^{(k+1)}\theta_{\xi,1}^{(k)} + m_{\xi,0}/C_{\xi,0}}{W_\xi^{(k+1)} + 1/C_{\xi,0}}, \frac{1}{W_\xi^{(k+1)} + 1/C_{\xi,0}}\right),\\
\theta_{\xi,t}^{(s+1)} &\sim N\left(\frac{V_\xi^{(s+1)} l\xi_t^{(s+1)} + W_\xi^{(s+1)}(\theta_{\xi,t+1}^{(s)} - \theta_{\xi,t-1}^{(s+1)})}{V_\xi^{(s+1)} + 2W_\xi^{(s+1)}}, \frac{1}{V_\xi^{(s+1)} + 2W_\xi^{(s+1)}}\right),\\
t &= 1, \quad \ldots \quad , n-1,\\
\theta_{\xi,n}^{(s+1)} &\sim N\left(\frac{V_\xi^{(s+1)} l\xi_n^{(s+1)} + W_\xi^{(s+1)}\theta_{\xi,t-1}^{(s+1)}}{V_\xi^{(s+1)} + W_\xi^{(s+1)}}, \frac{1}{V_\xi^{(s+1)} + W_\xi^{(s+1)}}\right),\\
V_\sigma^{(s+1)} &\sim G\left(\frac{f_\sigma + \frac{n}{2}}{o_\sigma + \sum_{t=1}^n (l\sigma_t^{(s+1)} - \theta_{\sigma,t}^{(s)})^2}, f_\sigma + \frac{n}{2}\right),\\
W_\sigma^{(s+1)} &\sim G\left(\frac{l_\sigma + \frac{n}{2}}{m_\sigma + \sum_{t=1}^n (\theta_{\sigma,t}^{(s)} - \theta_{\sigma,t-1}^{(s)})^2}, l_\sigma + \frac{n}{2}\right),\\
\theta_{\sigma,0}^{(s+1)} &\sim N\left(\frac{W_\sigma^{(k+1)}\theta_{\sigma,1}^{(k)} + m_{\sigma,0}/C_{\sigma,0}}{W_\sigma^{(k+1)} + 1/C_{\sigma,0}}, \frac{1}{W_\sigma^{(k+1)} + 1/C_{\sigma,0}}\right),\\
\theta_{\sigma,t}^{(s+1)} &\sim N\left(\frac{V_\sigma^{(s+1)} l\sigma_t^{(s+1)} + W_\sigma^{(s+1)}(\theta_{\sigma,t+1}^{(s)} - \theta_{\sigma,t-1}^{(s+1)})}{V_\sigma^{(s+1)} + 2W_\sigma^{(s+1)}}, \frac{1}{V_\sigma^{(s+1)} + 2W_\sigma^{(s+1)}}\right),
\end{aligned}$$

$$
\begin{aligned}
& t = 1, \quad \ldots \quad , n-1, \\
\theta_{\sigma,n}^{(s+1)} \quad & \sim \quad N\left(\frac{V_{\sigma}^{(s+1)} l\sigma_{n}^{(s+1)} + W_{\sigma}^{(s+1)}\theta_{\sigma,t-1}^{(s+1)}}{V_{\sigma}^{(s+1)} + W_{\sigma}^{(s+1)}}, \frac{1}{V_{\sigma}^{(s+1)} + W_{\sigma}^{(s+1)}}\right).
\end{aligned}
$$

Capítulo 6

Conclusões e continuações

Baseado no trabalho realizado, pode-se tirar as seguintes conclusões:

- embora a estimativa não-paramétrica baseada numa aproximação por mistura tenha resultado em boas predições para dados extremos, ela não é tão eficiente na estimação da cauda e dos valores em torno do limiar como o modelo que utiliza a distribuição GPD, que na teoria, pelo Teorema de Pickands (1975), é a indicada para a cauda;

- em alguns casos, como nos apresentados nas aplicações, é necessário predizer a distribuição por uma mistura com mais de uma componente. Embora eficiente em alguns casos, o modelo de Behrens et al. (2004) não seria apropriado em alguns casos onde os dados são positivos, mas não possuem distribuição Gama. Mostrou-se que a abordagem não-paramétrica por mistura engloba uma ampla classe de distribuições, e que não é necessário um grande número de componentes na mistura para se fazer uma boa predição;

- o modelo de mistura com GPD se mostra mais eficiente na estimação do limiar quando é maior o tamanho da amostra. Quando este tamanho é menor, é necessário impor uma distribuição a priori próxima do verdadeiro valor do limiar, ou próxima de um valor que supõe-se ser o limiar, geralmente um quantil alto dos dados. Com n

pequeno, há poucas observações para realizar a estimação dos parâmetros da cauda, o que torna difícil a estimação destes parâmetros;

- o critério de comparação de modelos BIC pareceu ser o mais eficiente de acordo com as simulações, e mostrou nas aplicações, que o melhor modelo é exatamente o que utiliza mistura com mais de uma componente com GPD na cauda. O critério DIC, embora com várias restrições na literatura, pareceu ser um bom critério de comparação de acordo com as simulações. Na abordagem que utiliza modelos dinâmicos, o DIC é a medida mais adequada para verificar o ajuste do modelo;

- o modelo que considera estrutura de modelos de regressão na cauda se mostrou muito eficiente, onde as covariáveis, escolhidas de maneira adequada, forneceram uma informação que ajudou a explicar o comportamento da variável, podendo-se estimar de maneira mais precisa valores dos parâmetros em particulares configurações de covariável, ou seja, pode-se encontrar distribuições para os parâmetros da cauda em uma determinada cidade em um específico mês do ano. Além disso, a estimação trouxe informações importantes como valores de máximos e quantis elevados, quando foram estudadas temperaturas máximas, e mínimos e quantis baixos, quando o objetivo foi analisar temperaturas mínimas;

- além de dados ambientais, este trabalho mostrou em seu último modelo a análise de retornos financeiros. Para estes dados, foi feita uma estrutura de modelos dinâmicos para os parâmetros da cauda da distribuição. Este método se mostrou eficiente, com o comportamento dos parâmetros variando com o tempo, acompanhando nas observações os períodos de crise na economia.

Este trabalho pode ter continuidade em várias direções, entre elas:

- Utilizar uma estrutura espaço-temporal para os parâmetros da cauda. Em muitas aplicações, o comportamento de uma variável pode estar relacionado com o que

ocorre em regiões vizinhas. Detalhes de modelos espaço-temporais podem ser vistos, por exemplo em Cressie (1993). Considere uma região espacialmente contínua $D \subset \mathcal{R}^2$, no qual somente para um conjunto n de posições fixas $(s_1, \ldots, s_n)$, são conhecidas as medidas de interesse $\mathbf{Y} = (Y(s_1), \ldots, Y(s_n))'$. O objetivo da estatística espacial é fornecer para qualquer $s_0 \subset D$ a melhor estimativa $Y(s_0)$ a partir de $\mathbf{Y}$. Uma abordagem usual decompõe $\mathbf{Y}$ nas seguintes componentes:

$$\begin{aligned} \mathbf{Y} &= (\text{média}) + (\text{componente espacial}) + (\text{erro de medida}) \\ &= \mu + \mathbf{Z} + \epsilon, \end{aligned}$$

onde μ é a média que pode ser por exemplo uma forma linear $\mu = X\beta$. A componente $\mathbf{Z}$ é um vetor $(n \times 1)$ de realização de um campo aleatório com vetor de médias $\mathbf{0}$. Em muitas aplicações, $\mathbf{Z}$ é assumido ser um processo Gaussiano, ou seja, possui distribuição multivariada com média $\mathbf{0}$ e matriz de covariância Ω. No contexto da teoria espacial clássica, é utilizada a simplificação $\Omega = \sigma^2\Sigma$, onde Σ é a matriz de correlação e σ^2 é uma variância constante, igual para todos os $Z(s)$.

Outra simplificação ocorre do fato da matriz de correlação depender apenas de uma função das distâncias s_i e s_j, ou seja,

$$\rho_{i,j} = \rho(Z(s_i), Z(s_j)) = \rho(\mid s_i - s_j \mid) = \rho(\mid \gamma \mid),$$

para todo $i = 1, \ldots, n$ e $j = 1, \ldots, n$. Existem vários tipos de função de correlação. Um exemplo é a função de correlação exponencial, dada por

$$\rho = \exp -\phi(\mid \gamma \mid),$$

onde ϕ é uma função que mede como a correlação decai com a distância.

Em relação a análise de valores extremos, pode-se propor a adição da componente espacial para os parâmetros da cauda de maneira parecida ao do Capítulo 4, da

seguinte maneira

$$\begin{aligned} \underline{u} &= \mathbf{x}_u\beta_u + \mathbf{Z}_u \\ \underline{\xi} &= \exp\{\mathbf{x}_\xi\beta_\xi + \mathbf{Z}_\xi\} - 1 \\ \underline{\nu} &= \exp\{\mathbf{x}_\nu\beta_\nu + \mathbf{Z}_\nu\}, \end{aligned}$$

onde $\underline{u} = (u_1, \ldots, u_n)$, $\underline{\xi} = (\xi_1, \ldots, \xi_n)$ e $\underline{\nu} = (\nu_1, \ldots, \nu_n)$. Cada parâmetro possui um vetor de componente espacial, e os novos parâmetros seriam os parâmetros da função de correlação em $\mathbf{Z}_u$, $\mathbf{Z}_\xi$ e $\mathbf{Z}_\nu$.

- Fazer um modelo dinâmico mais estruturado, considerando também o limiar variando no tempo. O modelo linear dinâmico geral, visto, por exemplo em West e Harrison (1997), é escrito da seguinte forma

$$\begin{aligned} \mathbf{y}_t &= F_t'\theta_t + v_t \qquad v_t \sim N(0, 1/V) \\ \theta_t &= G_t\theta_{t-1} + w_t \quad w_t \sim N(0, 1/W). \end{aligned} \tag{6.1}$$

  Note que o modelo do Capítulo 5 é um caso particular do modelo acima, onde $\mathbf{y}_t$, é um dos parâmetros da cauda, por exemplo $l\nu_t = \log(\nu_t)$, $F_t' = 1$ e $G_t = 1$. Uma primeira extensão natural seria considerar também o modelo dinâmico para o limiar, pois o valor da variável considerada evento extremo pode se alterar ao longo do tempo, principalmente em dados financeiros onde níveis de retornos e valores de ações têm uma grande diferença ao longo das décadas.

  Além do modelo dinâmico de primeira ordem, pode ser sugerido um modelo dinâmico de ordem maior, que pode fornecer maior informação de como os parâmetros se movem com o tempo. Huerta e Sansó (2007) utilizam um modelo linear dinâmico de segunda ordem para o parâmetro de locação da distribuição GEV, onde $F_t' = (1\ 0)$, $\theta_t' = (\delta_t, \beta_t)$,

$$G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

$v_t \sim N(0, V)$ e $w_t' = (w_\delta, w_\beta) \sim N(\mathbf{0}, \mathbf{W})$. Assim, obtém-se o seguinte modelo para o parâmetro:

$$\begin{aligned} \mu_t &= \delta_t + v_t \\ \delta_t &= \delta_{t-1} + \beta_{t-1} + w_\delta \\ \beta_t &= \beta_{t-1} + w_\beta. \end{aligned}$$

Seguindo esta ideia, pode-se se fazer um modelo linear dinâmico de segunda ordem para cada um dos parâmetros da cauda.

- Unir as idéias dos Capítulos 4 e 5, realizando uma regressão dinâmica para os parâmetros da cauda. Em aplicações de valores extremos, além das variáveis de interesse, os valores das covariáveis também podem se alterar ao longo do tempo. Por isso, é importante propor um modelo que considera este tipo de comportamento para as covariáveis. A estrutura deste modelo também é um caso particular da equação (6.1), onde para um único parâmetro da cauda, considerando a presença de p covariáveis no modelo, $F_t' = (1, x_{1,t}, \ldots, x_{t,p})$, $G = I_n$, $\theta_t' = (\beta_{t,0}, \ldots, \beta_{t,p})$. Assim, no caso específico de regressão no parâmetro de escala da ditribuição GPD denotado por ν, o modelo para $l\nu_t = \log(\nu_t)$ é dado por

$$\begin{aligned} l\nu_t &= \beta_{t,0} + \beta_{t,1}x_{t,1} + \ldots + \beta_{t,p}x_{t,p} + v_t \\ \beta_{t,i} &= \beta_{t-1,i} + w_{i,t}, \quad i = 0, \ldots, p. \end{aligned}$$

Note que neste modelo os parâmetros β também variam ao longo do tempo. Um modelo análogo pode ser construído para os parâmetros $l\xi_t = \log(1 + \xi_t)$ e u_t.

Em relação ao limiar, é necessário ter cautela em fazer a estimação devido a dificuldade de estimação deste parâmetro, principalmente quando o tamanho da amostra não é muito grande, sendo necessário dar uma informação com uma precisão alta, como pode ser visto na estimação dos Capítulos 3, 4 e 5.

Referências

ASMUSSEN, S. (1987). *Applied Probability and Queues*, New York: Wiley.

BEHRENS, C. , GAMERMAN, D. e LOPES, H. F. (2004). Bayesian analysis of extreme events with threshold estimation, *Statistical Modelling*, **4**, 227-244.

BERMUDEZ, P. , TURKMAN, M. A. e TURKMAN, K. F. (2001). A predictive approach to tail probability estimation, *Extremes*, **4:4**, 295-314.

BOX, G. E. P. e TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis*, Addison Wesley, California.

CABRAS, S. , CASTELLANOS, M. A. e GAMERMAN, D. (2009). A default approach for regression on parameters of generalized Pareto distribution, *Relatórios Técnicos*, LES, UFRJ.

CASTELLANOS, M. A. e CABRAS, S. (2007). A default Bayesian procedure for the generalized Pareto distribution, *Journal of Statistical Planning and Inference*, **137**, 473-483.

CELEUX, G., HURN, M. e ROBERT, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions, *Journal of the American Statistician Association*, **95**, 957-970.

CHAVES-DEMOULIN, V. e DAVISON, A. C. (2005). Generalized additive modelling of sample extremes, *Applied Statistics*, **54**, 207-222.

COLES, S. G. e TAWN, J. A. (1996). A Bayesian analysis of extreme rainfall data, *Applied Statistics*, **45**, 463-478.

COLES, S. G. (2001). *Extreme Value Theory an Applications*, Edited by Galambos,

J.;Lechner, J.;Mimiu, E.[s.l.]: Kluver Academic Publishers.

CRESSIE, N. (1993). *Statistical for Spatial Data*, New York: Wiley

DAVISON, A. C. e SMITH, R. L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society: Series B*, **52**, 393-442.

DE VORE R., e LORENTZ, G. (1993). *Constructive Approximation*, New York: Springer Verlag

DIEBOLT, J. e ROBERT, C. (1994). Estimation of finite mixture distributions by Bayesian sampling, *Journal of the Royal Statistical Society Series B*, **56**, 363-375.

DOORNIK, J. A. (1996). *Ox: Object Oriented Matrix Programming*, 3.1 Console Version. Nuffield College, Oxford University, London.

EMBRECHTS, P., KÜPPELBERG, C. e MIKOSCH, T. (1997). *Modelling Extremal Events for Insurance and Finance*, New York: Springer.

FISHER, R. A. e TIPPETT, L. H. C. (1928). On the estimation of the frequency distributions of the largest and smallest sumber of a sample, *Proceedings of the Cambridge Philosophycal Society*, **24**, 180-190.

FRIGESSI, A., HAUG, O. e RUE, H. (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection, *Extremes*, **5**, 219-235.

FRÜHWIRTH-SCHNATTER, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models, *Journal of the American Statistical Association*, **96**, 194-209.

GAMERMAN, D. e LOPES, H. F. (2006) *Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2a. edição, Chapman & Hall.

GELFAND, A. E. e GHOSH, S. (1998). Model choice: a minimum posterior predictive loss approach, *Biometrika*, **85**, 1–11.

GREEN, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82**, 711-732.

HUERTA, G. e SANSÓ, B. (2007). Time-varying models for extreme values, *Environmental and Ecological Statistics*, **14(3)**, 285-299.

JASRA, A., HOLMES, C. C. e STEPHENS, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modelling, *Statistical Science*, **20(1)**, 50-67.

JENKINSON, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological events, *Quarterly Journal of the Royal Meteorology Society* , **81**, 158-171.

LOPES, H. F., NASCIMENTO, F.F. e GAMERMAN, D. (2009). Generalized Pareto models with time-varying tail behavior, *Relatórios Técnicos*, GSB, University of Chicago.

MENDES, B. V. M (2004). *Introdução a análise de eventos extremos*, Rio de Janeiro, E-papers.

NEAL, R. (1996). Sampling from multimodal distributions using tempered transitions, *Statistical Computing*, **4**, 353-366.

PARMESAN, C., ROOT, T.L., and WILLING, M.R. (2000). Impacts of extreme weather and climate on terrestrial biota *Bulletin of the American Meteorological Society*, **81**, 443-450.

PICKANDS, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics*, **3**, 131-199.

RICHARDSON, S. e GREEN, P. (1997). On Bayesian analysis of mixtures with an unknown number of components, *Journal of the Royal Statistical Society, Series B*, **59(4)**, 731-792.

ROBERTS, G. O. e ROSENTHAL, J. S. (2006). Examples of adaptive MCMC, *Relatórios Técnicos,*, Lancaster University.

ROEDER, K. e WASSERMAN, L. (1997), Practical Bayesian density estimation using mixtures of Normals, *Journal of the American Statistical Association*, **92**, 894-902.

SCHWARZ, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6(2)**, 461-464.

SILVERMAN, B. W. (1986). *Density Estimation*, London: Chapman and Hall.

SMITH, R. L., (1984). Threshold models for sample extremes. In J. Tiago de Oliveira,

Reidel, Dordrecht (editors), *Statistical Extremes and Applications*, 621-638.

SPIEGELHALTER, D. , BEST, N. G. , CARLIN, B. P. e VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society, Series B*, **64(4)**, 583-639.

STEPHENS, M. (1997). Bayesian methods for mixtures of normal distributions. D.Phil. dissertations, Dept. Statistics, Univ. Oxford.

TANCREDI, A., ANDERSON, C. e O'HAGAN, A. (2006). Accounting for threshold uncertainty in extreme value estimation, *Extremes*, **9**, 87-106.

TANNER, M. e WONG, W. (1987). The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, **82**, 528-550.

TITTERINGTON, D., SMITH, A. F. M and MAKOV, U. (1985). *Statistical Analysis of Finite Mixture Distributions*, New York,Wiley.

VON MISES, R. (1954). La distribution de la plus grande de n valeurs, *American Mathematical Society: Selected Papers*, **2**, 271-294.

WEST, M. e HARRISON, J. (1997). *Bayesian Forecasting and Dynamic Models* , 2a. Edição, New York, Springer-Verlag.

WIPER, M., RIOS INSUA, D. e RUGGERI, F. (2001). Mixtures of Gamma distributions with applications, *Journal of Computational and Graphical Statistics*, **10**, 440-454.

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo