

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E MATEMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Uma Abordagem Clássica e Bayesiana para o Modelo  
de Tweedie

**Flávia Ferreira**

**Orientador: Prof. Dr. Carlos Alberto Ribeiro Diniz**

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística.

**São Carlos**

**Abril de 2005**

**Ficha catalográfica elaborada pelo DePT da  
Biblioteca Comunitária da UFSCar**

F383ac

Ferreira, Flávia.

Uma abordagem clássica e bayesiana para o modelo de Tweedie / Flávia Ferreira. -- São Carlos : UFSCar, 2008.  
81 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2005.

1. Estatística - análise. 2. Inferência bayesiana.  
3. Distribuição de Poisson. 4. Modelo de dispersão exponencial. 5. Seguros - métodos estatísticos. 6. Inferência clássica. I. Título.

CDD: 519.5 (20ª)

**Dedico este trabalho aos meus filhos  
Victor e Giulia.**

# Agradecimentos

Agradeço,

*a Deus e aos anjos que me protegem e me guiam em direção ao bem,*

*a minha mãe Yeda Benedicta Stringhetti Ferreira e demais familiares, que acreditaram em minha capacidade, e em especial aos meus filhos Victor e Giulia que pacientemente me acompanharam nessa minha jornada,*

*a meu orientador Dr. Carlos A. Ribeiro Diniz, pela paciência, apoio e disponibilidade em me ajudar nos momentos oportunos,*

*a CAPES (Coordenação de Aperfeiçoamento Pessoal de Nível Superior) pela assistência financeira,*

*aos professores Dr. Francisco Louzada Neto e Dr<sup>a</sup>. Vera Lúcia Damasceno Tomazella pelas correções e sugestões dadas em meu exame de qualificação,*

*a todos os professores e funcionários do Departamento de Estatística da UFSCar, em especial, agradeço ao professor Dr. Benedito Galvão Benze por acreditar na minha capacidade para fazer o curso de Mestrado em Estatística nesta instituição de ensino e a professora Dr<sup>a</sup> Tereza Cristina Martins Dias que contribuiu para o desenvolvimento do meu trabalho,*

*aos amigos Simone, Lia e Milton pela grande amizade nas horas alegres e difíceis,*

*enfim, gostaria de agradecer a todos, que, direta ou indiretamente, me acompanharam e me ajudaram durante o desenvolvimento deste trabalho, o meu muito obrigada.*

# *Resumo*

Nesse trabalho apresentamos uma abordagem clássica e Bayesiana para o modelo de Tweedie Poisson Composto na forma e aplicação apresentada por Jorgensen & Souza (1992). O objetivo específico é a estimação do custo médio individual de sinistro. Na obtenção dos estimadores clássicos é utilizado o método iterativo de *Newton-Raphson*. Na obtenção de estimadores Bayesianos são utilizadas densidades a *priori* não informativas para os parâmetros de interesse. Sob o enfoque Bayesiano duas metodologias são estudadas: *i)* sem a presença de covariáveis *ii)* com a presença de covariáveis. Uma comparação dos desempenhos dos estimadores clássicos e Bayesianos é verificada através de dados simulados.

**Palavras-chave:** Modelo de Tweedie Poisson Composto, Modelo de Dispersão Exponencial, Inferência Bayesiana, Seguros.

# *Abstract*

The goal of this work is to present a classic and a bayesian approach to the Tweedie Compound Poisson model considering the form and application shown at Jorgensen & Souza (1992). A specific goal is to estimate the individual claim mean values. Newton-Raphson iterative method is used in the classic inference in order to obtain the estimatives for the parameters. Non-informative priors are used to obtain the bayesian estimatives for the parameters. In the bayesian approach two different methodologies are considered: i) without the presence of covariance and ii) with the presence of covariance. The performance of both methods, classic and bayesian inference, are compared via simulated data.

**keywords:** Tweedie Compound Poisson model, Exponential dispersion model, bayesian inference, insurance

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Teoria de risco e Modelos de Dispersão Exponencial</b>	<b>5</b>
2.1	Teoria de risco . . . . .	5
2.2	Modelos de Dispersão Exponencial . . . . .	6
2.2.1	Modelo de dispersão exponencial aditivo e reprodutivo . . . . .	9
2.2.2	Propriedade de Convolução das Famílias $ED^*(\theta, \lambda)$ . . . . .	10
2.2.3	Casos especiais . . . . .	11
<b>3</b>	<b>Modelo de Tweedie</b>	<b>13</b>
3.1	Família Tweedie . . . . .	14
3.1.1	Distribuição Normal, $p = 0$ . . . . .	14
3.1.2	Distribuição Gamma, $p = 2$ . . . . .	14
3.1.3	Distribuição Poisson, $p = 1$ . . . . .	14
3.1.4	Distribuição Poisson-Composta, $p = 1/2$ . . . . .	15
<b>4</b>	<b>Inferência Clássica para o Modelo e Tweedie</b>	<b>18</b>
4.1	Modelo Estatístico e a Função de Verossimilhança . . . . .	19
4.2	Estimativas de Máxima Verossimilhança . . . . .	20
4.3	Intervalos de Confiança . . . . .	21
4.4	Aplicação dos Estimadores em Conjuntos de Dados Simulados . . . . .	22
<b>5</b>	<b>Uma abordagem Bayesiana para o Modelo de Tweedie</b>	<b>25</b>
5.1	Metodologia . . . . .	25
5.2	Inferência Bayesiana sobre os parâmetros $m$ , $\alpha$ e $\theta$ . . . . .	26

5.2.1	Estimação dos Parâmetros . . . . .	28
5.2.2	Análise de convergência e auto-correlação dos parâmetros de interesse . . . . .	28
5.2.3	Resultados apresentados pela reparametrização $\varphi = \frac{\theta}{\alpha}$ . . . . .	30
5.3	Resultados da Aplicação com os dados simulados . . . . .	31
<b>6</b>	<b>Abordagem Bayesiana para o Modelo de Tweedie envolvendo Covariáveis</b>	<b>35</b>
6.1	Metodologia . . . . .	35
6.2	O Modelo na Presença de Uma Covariável . . . . .	37
6.2.1	Inferência Bayesiana sobre os parâmetros $\beta_0, \beta_1, \varphi$ e $m$ . . . . .	38
6.3	O Modelo na presença de Duas Covariáveis . . . . .	40
6.3.1	Inferência Bayesiana sobre os parâmetros $\beta_0, \beta_1, \beta_2, \varphi$ e $m$ . . . . .	41
6.4	Uma Aplicação em conjuntos de dados simulados . . . . .	44
6.4.1	Ilustração dos Dados Simulados . . . . .	46
6.4.2	Resultados referentes ao modelo na presença de uma covariável . . . . .	47
6.4.3	Resultados referentes ao modelo na presença de duas covariáveis . . . . .	53
<b>7</b>	<b>Comparação entre estimadores Clássicos e Bayesianos</b>	<b>58</b>
7.1	Desempenho dos estimadores clássicos e Bayesianos . . . . .	58
7.2	Resultados Obtidos . . . . .	59
7.2.1	Baixa Taxa de Ocorrência de Sinistros-Poisson (3) . . . . .	60
7.2.2	Alta Taxa de Ocorrência de Sinistros-Poisson (50) . . . . .	62
7.3	Conclusão . . . . .	63
<b>8</b>	<b>Conclusões e Propostas Futuras</b>	<b>64</b>
	<b>Referências Bibliográficas</b>	<b>68</b>
<b>A</b>	<b>Demonstrações</b>	<b>68</b>
<b>B</b>	<b>Programas</b>	<b>72</b>



# Capítulo 1

## Introdução

O seguro de veículos motorizados é uma indústria extremamente competitiva, onde a combinação de crescimento de mercado e rentabilidade é vista como imperativo para o sucesso.

Decisões corretas relativas ao custo de sinistros implicam em crescimento de mercado e rentabilidade. Sendo assim é preciso prever o custo médio e a frequência dos sinistros, examinando o efeito do preço na rentabilidade. É necessário também estar ciente do efeito do preço nos padrões de retenção ao consumidor bem como fornecer o potencial para o crescimento do mercado considerando a natureza altamente competitiva do negócio.

Para a companhia de seguro o interesse maior é estabelecer um nível de preço compatível com o custo previsto para o sinistro que ao mesmo tempo consiga reter o cliente existente e adquirir novos clientes. Conhecendo quais consumidores são prováveis de renovar seus seguros, bem como seus níveis de risco, aumenta a facilidade na determinação do prêmio.

Na indústria de seguros, assim como em várias indústrias competitivas, o cliente é livre para escolher sua seguradora. Sua decisão é baseada numa combinação complexa de preço, serviço, preferência pessoal, conveniência e vários outros resultados. Neste meio é muito importante perceber quais clientes estão saindo e porque, de forma que a iniciativa de retenção possa se tornar mais evidenciada.

O risco de uma pessoa normalmente é acessado pela companhia de seguro de acordo com o comportamento estatístico de outros segurados. Usando o modelo de risco estatís-

tico padrão, pode-se assumir que os sinistros ocorrem conforme um processo de Poisson, dependendo das características do segurado e fatores ambientais.

O grande distanciamento tecnológico da área de seguros do Brasil em relação a outras economias fez com que a precificação do seguro de automóveis permanecesse durante muitos anos baseada somente nos dados do veículo ( marca, tipo, ano etc ) e em sua região de uso. Somente a poucos anos, os dados do motorista passaram a ser considerados no momento de calcular o preço do seguro. A primeira grande seguradora brasileira a implantar no cálculo do preço do seguro de automóveis fatores que levavam em conta dados do motorista usou uma metodologia bastante simples, da qual as únicas variáveis eram idade e sexo.

Evidentemente a operação de cálculo de prêmio com a introdução do "perfil" ficou mais sofisticada, exigindo mais tecnologia das companhias de seguro.

A situação atual mostra que as companhias de seguro mais evoluídas tecnologicamente e mais bem aparelhadas em termos de estatística e informática, tiraram melhores proveitos dos benefícios da nova forma de aceitação e precificação de seguros de automóveis e tem no momento, uma carteira de automóveis de alta qualidade e resultado superior aos concorrentes.

A introdução do perfil mostrou que os jovens, principalmente do sexo masculino, apresentam para o seguro de automóveis um grau de risco muito superior a média. Sem dúvida nenhuma, o preço dos seguros de automóveis para os jovens cresceu enormemente, na grande maioria das companhias, com a introdução do perfil. Portanto o preço do produto segurado não pode ser estudado sem a consideração de fatores inter-relacionados, como, idade do segurado, sexo, distância percorrida pelo veículo, idade do veículo, CEP, entre outros.

Existe um grande número de estudos sobre seguro de veículos que fazem uso da modelagem estatística. Muitos desses estudos, foram publicados há mais de vinte anos (Almer, 1957; Barley e Simon, 1960; Jung, 1966). Um progresso maior foi feito durante a década de setenta com os estudos de Johnson e Hey (1971) e Bennett (1978), e mais recentemente estudos de Baxter, Coutts e Ross (1980) e Coutts (1984).

Os autores desses estudos acreditam que para assegurar um aceitável retorno de capital uma companhia de seguro precisa ser uma empreendedora acima da média, e defendem

a teoria de que através de uma abordagem estatística completa de seus dados, as companhias possam alcançar maior lucratividade. Essa abordagem requer uma substancial análise dos dados, na qual técnicas de modelagem estatística tornam possível fornecer a seguradora informações adicionais sobre todos os fatores relevantes para o estudo de sinistros no seguro.

Nas estimativas dos parâmetros do modelo, geralmente, usam-se métodos de máxima verossimilhança. Assim, estes modelos dependem de suposições sobre distribuições conhecidas e tendem a ser baseados em um sistema de classificação pré-definida para determinar o risco individual do prêmio.

Para uma dada estrutura de categoria de fatores inter-relacionados, vários pesquisadores têm investigado como os sinistros podem ser melhor modelados usando máxima verossimilhança e conceitos estatísticos.

Jorgensen & Souza (1992) consideraram os modelos de dispersão exponencial, mais precisamente o uso do modelo de Tweedie Poisson Composto na determinação de precificação de seguros. Usando a teoria de modelos de dispersão exponencial, Jorgensen (1986, 1987) determinou, através de fórmulas de convolução, que o custo total de um determinado grupo de segurados, dado o número de sinistros, segue um modelo de dispersão exponencial e tem distribuição Gama.

Neste trabalho, assumindo que o número de sinistros segue um processo de Poisson e o custo dos sinistros possui uma distribuição Gama, apresentaremos a abordagem clássica e Bayesiana para o modelo de Tweedie Poisson Composto na forma e aplicação apresentada por Jorgensen & Souza (1992), considerando a introdução de covariáveis e uma reparametrização que reduz significativamente a correlação existente entre os parâmetros. O objetivo específico é a estimação do custo médio individual de sinistros, visto que o valor pago pelos sinistros tem grande efeito na rentabilidade de uma indústria de seguros.

A dissertação é desenvolvida em oito capítulos e apresenta a seguinte composição. No capítulo 2 é discutido a teoria de risco e é apresentado o modelo de dispersão exponencial. No capítulo 3 é apresentado o modelo de Tweedie. No capítulo 4 são apresentadas uma abordagem clássica para o modelo de Tweedie, as estimativas dos parâmetros de interesse para prever o custo médio individual e uma aplicação com dados simulados. No capítulo 5 são apresentadas a abordagem Bayesiana para modelo de Tweedie, bem como os re-

sultados das estimativas de seus parâmetros aplicadas em dados simulados. No capítulo 6 é apresentada a abordagem Bayesiana para o modelo de Tweedie com a presença de covariáveis. No capítulo 7 é apresentada uma comparação entre estimadores clássicos e Bayesianos e, finalmente, no capítulo 8 são apresentados a conclusão e propostas futuras.

# Capítulo 2

## Teoria de risco e Modelos de Dispersão Exponencial

Neste capítulo serão apresentados a teoria de risco e os modelos de dispersão exponencial, com suas propriedades e alguns de seus casos especiais.

A classe dos modelos de dispersão exponencial inclui várias distribuições para dados não normais, inclusive para dados positivos com zero, comuns à teoria de riscos e intensivamente estudada por Tweedie (1947), que descobriu muitas das importantes propriedades destes modelos e alguns de seus casos especiais. No entanto, os estudos de Tweedie não foram aplicados por décadas devido a não existência de computadores capazes de realizar as suas análises estatísticas. Os estudos mais sistemáticos das propriedades matemáticas dos Modelos de Dispersão Exponencial só se iniciaram com Jorgensen (1986, 1987).

### 2.1 Teoria de risco

Considerando a indústria de seguros, a teoria de riscos é responsável pelo estudo, análise e quantificação de riscos, além de ajuste de modelos de precificação e ruínas para seguros. As informações pertinentes são provenientes das observações das variáveis aleatórias *número de sinistros de um determinado grupo de segurados*,  $N(w)$ , e *valor total de sinistros por grupo*,  $Z(w)$ , definidas como

$$Z(w) = \sum_{i=1}^{N(w)} Z_i,$$

onde  $Z(w)$  é zero se  $N(w)$  for zero. Os  $Z_i$ 's são variáveis aleatórias identicamente distribuídas, independentes entre si e de  $N(w)$  e representam os valores individuais dos sinistros.

A exposição  $w$  é definida como  $w = vt$ , onde  $v$  é o valor assegurado e  $t$  é o intervalo de tempo durante o qual o valor  $v$  é exposto ao risco. A razão de sinistro observado por unidade de exposição é definido por

$$Y(w) = \frac{Z(w)}{w}.$$

E a média da razão de sinistros é definida como

$$\mu = E[Y(w)].$$

Neste estudo, por simplicidade e sem, no entanto, perder a generalidade,  $v$  e  $t$  são assumidos como sendo iguais a 1 para todas as apólices da carteira e, portanto, a exposição  $w$  é igual a 1.

## 2.2 Modelos de Dispersão Exponencial

A classe dos modelos de dispersão exponencial é uma das principais classes dos modelos de dispersão, que, por sua vez, inclui distribuições para dados na linha real, dados positivos, dados positivos com zero, dados de contagem e dados binomiais.

Um modelo de dispersão exponencial,  $ED$ , é uma família de distribuições com função densidade de probabilidade da forma

$$p(y; \theta, \lambda) = a(\lambda, y) \exp[\lambda \{y\theta - k(\theta)\}], \quad y \in R, \quad (2.1)$$

não degenerada, com função adequada  $a$ , com  $k(\theta)$  uma função definida por

$$k(\theta) = \log \int e^{\theta y} \nu(dy), \quad (2.2)$$

onde  $\nu$  é uma medida finita em  $R$ . O domínio de  $\theta$  é dado por  $\Theta$ , não vazio, da função  $\lambda k(\theta)$ , considerando que o domínio de  $\lambda k(\theta)$  é independente de  $\lambda$ . O parâmetro  $\lambda$  varia num conjunto  $\Lambda$  que, sem perda de generalidade, consideramos contido em  $R_+$ .

Para um valor dado de  $\lambda$ , (2.1) reduz-se a uma família exponencial natural cuja função densidade é dada por

$$p(y; \theta) = a(y) \exp\{\theta y - k(\theta)\}, y \in R.$$

O domínio considerado para o parâmetro canônico  $\theta$  da família segue sendo o domínio  $\Theta$ , agora da função  $k(\theta)$ .

Por definição, a família exponencial natural é caracterizada pela função de variância  $V(\mu)$ , onde

$$\begin{aligned} \mu &= k'(\theta) = \tau(\theta), \\ \tau^{-1} &= (k')^{-1}, \\ \theta &= \tau^{-1}(\mu), \\ \text{Var}(y) &= k''(\theta), \\ V(\mu) &= k''(\tau^{-1}(\mu)) = \text{Var}(Y). \end{aligned}$$

A função geradora de momentos em um modelo  $ED$  é da forma

$$G(t; \theta, \lambda) = \exp[\lambda \{k(\theta + t/\lambda) - k(\theta)\}], t \in \lambda(\Theta - \theta).$$

Assim, da mesma forma que na família exponencial natural, temos que na família de dispersão exponencial

$$\begin{aligned} \mu &= k'(\theta) = \tau(\theta), \theta \in \text{int}(\Theta), \\ \text{Var}(Y) &= \sigma^2 V(\mu), \mu \in \Omega, \end{aligned}$$

com

$$\begin{aligned} V(\mu) &= K''(\tau^{-1}(\mu)), \\ \Omega &= \tau(\text{int}(\Theta)), \end{aligned}$$

e

$$\sigma^2 = 1/\lambda \text{ o parâmetro de dispersão.}$$

Usa-se para os modelos *ED* a notação  $ED(\mu, \sigma^2)$  e a densidade (2.1) é reescrita como

$$f_y(y; \mu, \sigma^2) = a(y; \sigma^2) \exp\left\{-\frac{1}{2\sigma^2}d(y; \mu)\right\},$$

onde  $a \geq 0$  é uma função em  $y$  e  $\sigma^2$ ,  $d$  é uma medida de distância quadrática,  $\mu$  é o parâmetro de posição e  $\sigma^2$  o parâmetro de dispersão, já definidos anteriormente.

No caso de uma variável aleatória,  $Y$ , com distribuição normal  $N(\mu, \sigma^2)$ , a função de densidade tem a forma

$$\begin{aligned} p(y; \mu, \sigma^2) &= (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \\ &= (2\pi\sigma^2)^{-1/2} \exp\left\{\frac{-y^2}{2\sigma^2}\right\} \exp\left\{\left(y\mu - \frac{1}{2}\mu^2\right)/\sigma^2\right\}. \end{aligned}$$

Definindo  $\lambda = \frac{1}{\sigma^2}$  e fazendo  $\mu = \theta$ , obtém-se

$$p(y; \theta, \lambda) = a(\lambda, y) \exp[\lambda\{y\theta - k(\theta)\}],$$

com

$$a(\lambda, y) = (2\pi\sigma^2)^{-1/2} \exp\left\{\frac{-y^2}{2\sigma^2}\right\},$$

e

$$k(\theta) = \frac{1}{2}\theta^2.$$

E, portanto, a  $N(\mu, \sigma^2)$ , é um modelo de dispersão exponencial com  $V(\mu) = k''(\theta) = 1$ . Observe que, salvo  $\lambda$ , a função de variância caracteriza o modelo  $ED(\mu, \sigma^2)$ .



### 2.2.1 Modelo de dispersão exponencial aditivo e reprodutivo

Os modelos de dispersão exponencial podem ser definidos na forma aditiva ou reprodutiva. Assim, sendo  $Z$  uma variável aleatória com distribuição  $ED^*(\theta, \lambda)$  cuja função densidade de probabilidade é dada por

$$f_z^*(z; \theta, \lambda) = c^*(z; \lambda) \exp\{\theta z - \lambda K(\theta)\},$$

onde  $z \in R$ ,  $\theta$  é o parâmetro canônico,  $\lambda$  é o parâmetro de índice,  $c^*$  é uma função adequada e  $K(\theta)$  é a função cumulante definida em (2.2), então a família de distribuição de  $Z$  é chamada modelo de dispersão exponencial aditivo. Esta família tem esperança  $\lambda\mu$ , variância  $\lambda V(\mu)$  e função geradora de momentos  $G^*(t; \theta, \lambda) = \exp[\lambda\{k(\theta) + t - k(\theta)\}]$ .

Por outro lado, a correspondente família de distribuição da variável  $Y = Z/\lambda \sim ED(\mu, \sigma^2)$  é chamada de modelo de dispersão exponencial reprodutivo e sua densidade é dada por

$$f_y(y; \theta, \lambda) = c(y; \lambda) \exp[\lambda\{\theta y - K(\theta)\}],$$

onde  $y \in R$ ,  $\mu = \tau(\theta) = K'(\theta)$  é o parâmetro associado ao valor médio e  $\sigma^2 = 1/\lambda$  é o parâmetro de dispersão.

Portanto, se  $Y = Z/\lambda$ ,  $\mu = \tau(\theta)$  e  $\sigma^2 = 1/\lambda$ , as formas aditivas e reprodutivas podem ser resumidas pela equivalência de distribuição,

$$Y \sim ED(\mu, \sigma^2) \Leftrightarrow Z \sim ED^*(\theta, \lambda).$$

A transformação  $Z = \lambda Y = Y/\sigma^2$  é conhecida como transformação da dualidade.

A média e a variância no caso reprodutivo são

$$E(Y) = \mu \text{ e } Var(Y) = \sigma^2 V(\mu),$$

onde  $V(\mu) = \tau'\{\tau^{-1}(\mu)\}$  é a função de variância. No caso aditivo,

$$E(Z) = \xi \text{ e } Var(Z) = \lambda V(\xi/\lambda),$$

onde  $\xi = \lambda\tau(\theta) = \lambda\mu$ .

### 2.2.2 Propriedade de Convolução das Famílias $ED^*(\theta, \lambda)$

Os resultados a seguir foram propostos e demonstrados em Canal, G. Y. (1992).

**Resultado:** Considerando um modelo aditivo  $ED^*(\theta, \lambda)$  e assumindo que  $Z_1, \dots, Z_n$  são variáveis aleatórias independentes e  $Z_i \sim ED^*(\theta, \lambda_i)$ , para  $i = 1, \dots, n$ , então  $Z_+ = Z_1 + \dots + Z_n \sim ED^*(\theta, \lambda_1 + \dots + \lambda_n)$ , que corresponde à forma aditiva da fórmula de convolução, onde

$$ED^*(\theta, \lambda_1 + \dots + \lambda_K) = ED^*(\theta, \lambda_1) * \dots * ED^*(\theta, \lambda_K),$$

e  $*$  denota convolução. Ver prova no Apêndice A.

Como consequência da propriedade acima, temos o seguinte resultado. Se  $Y_i, \dots, Y_n$  são variáveis aleatórias independentes e

$$Y_i \sim ED(\mu, \sigma^2/w_i), i = 1, \dots, n,$$

com pesos  $w_i$ , tais que  $w_i/\sigma^2 \in \Lambda$ , então a forma reprodutiva da fórmula de convolução é dada por

$$\frac{1}{w_+} \sum_{i=1}^n w_i Y_i \sim ED(\mu, \sigma^2/w_+),$$

onde  $w_+ = w_1 + \dots + w_n$ .

Outro resultado que vem diretamente da propriedade de convolução acima é que, pelo Teorema Central do Limite, um modelo de dispersão exponencial é aproximadamente normal para  $\lambda$  grande. Assim para  $Z \sim ED^*(\theta, \lambda)$  temos, que, para qualquer  $\theta$  fixo,

$$\frac{Z - \lambda\mu}{\lambda^{-1/2}} \rightarrow^d N(0, V(\mu)) \text{ quando } \lambda \rightarrow \infty$$

onde  $d$  denota convergência em distribuição.

### 2.2.3 Casos especiais

Alguns modelos de dispersão exponencial são bastante conhecidos e estudados e são, mais frequentemente, denominados distribuição Normal, Gama, Poisson, Binomial e Binomial Negativa. A fórmula de convolução para a distribuição Gama, que é o caso de modelo de dispersão exponencial de maior interesse neste estudo, é apresentada na subseção a seguir.

#### Distribuição Gama

Uma importante distribuição para dados positivos é a distribuição Gama,  $Ga(\lambda, \Psi)$ , com função densidade para  $y > 0$  dada por

$$f_y(y; \lambda, \Psi) = \frac{\Psi^\lambda}{\Gamma(\lambda)} y^{\lambda-1} e^{-\Psi y}, \quad (2.3)$$

onde os parâmetros  $\lambda$  e  $\Psi$  são positivos.

Definindo  $\mu = \lambda/\Psi$  (média),  $\sigma^2 = 1/\lambda$  (coeficiente de variação quadrática) e parâmetro canônico  $\theta = -(1/\mu)$ , a densidade (2.3) pode ser reescrita como

$$f_y(y; \mu, \sigma^2) = \frac{\lambda^\lambda}{\Gamma(\lambda)} y^{\lambda-1} \exp \left\{ -\lambda \left( \frac{y}{\mu} + \log \mu \right) \right\}.$$

Assim, a distribuição Gama,  $Ga(u, \sigma^2)$ , é um modelo de dispersão exponencial reprodutivo com

$$\begin{aligned} K(\theta) &= \log \mu = \log\{-1/\theta\} = \log 1 - \log(-\theta) = -\log(-\theta), \\ c(y; \lambda) &= (\lambda^\lambda y^{\lambda-1})/\Gamma(\lambda), \\ \mu &= \tau(\theta) = K'(\theta) = -1/\theta, \\ V(\mu) &= \tau'(\tau^{-1}(\mu)) = \mu^2, \mu > 0 \text{ e} \\ \text{Var}(Y) &= \sigma^2 \mu^2. \end{aligned}$$

Aplicando a transformação de dualidade  $Z = \lambda Y$ , temos

$$\begin{aligned} Z &\sim Ga(\theta, \lambda) = Ga(-\lambda/\theta, 1/\lambda); \\ E(Z) &= E(\lambda Y) = \lambda\mu = \lambda(-1/\theta) = -\lambda/\theta; \\ Var(Z) &= \lambda^2 var(Y) = \lambda^2 \sigma^2 \mu^2 = \lambda^2 (1/\lambda) (-1/\theta)^2 = \lambda/\theta^2. \end{aligned}$$

Sejam  $Y_1, \dots, Y_n$  independentes e  $Y_i \sim Ga(\mu, \sigma^2/w_i)$ , a forma reprodutiva da fórmula de convolução é

$$\frac{1}{w_+} \sum_{i=1}^n w_i Y_i \sim Ga(u, \sigma^2/w_+),$$

onde  $w_+ = w_1 + \dots + w_n$ .

Sejam  $Z_1, \dots, Z_n$  independentes e  $Z_i \sim Ga(\theta, \lambda_i)$ , a forma aditiva da fórmula de convolução é

$$Z_1 + \dots + Z_n \sim Ga(\theta, \lambda_1 + \dots + \lambda_n).$$

# Capítulo 3

## Modelo de Tweedie

O modelo de Tweedie é uma classe do modelo de dispersão exponencial, com características particulares. Este modelo foi utilizado por Jorgensen & Sousa (1992) para resolver problemas de precificação de seguros. Em seu trabalho, os autores assumem a média da razão de sinistros como uma função de variáveis independentes.

A classe Tweedie de modelo de dispersão exponencial é definida como tendo função de variância unitária  $V_p(\mu) = \mu^p$  onde  $p \in (-\infty, 0) \cup (1, \infty)$ . Um resultado que mostra que para cada  $p \in (-\infty, 0) \cup (1, \infty)$  existe uma distribuição de probabilidade é encontrada em Jorgensen (1992). Se  $Y$  pertence à família Tweedie, escrevemos  $Y \sim ED^{(p)}(\mu, \sigma^2)$ . A família Tweedie é invariante para as transformações de escala, isto é,

$$\text{Se } Y \sim ED^{(p)}(\mu, \sigma^2) \text{ então, } cY \sim ED^{(p)}(c\mu, c^{2-p}\sigma^2),$$

já que

$$\begin{aligned} E(cY) &= c\mu, \text{ e} \\ \text{Var}(cY) &= c^2 \text{Var}(Y) = c^2 \sigma^2 \mu^p = \sigma^2 c^{2-p} (c\mu)^p, \end{aligned}$$

para  $c > 0$ . Esta propriedade é característica do modelo de Tweedie.

### 3.1 Família Tweedie

Alguns exemplos do Modelo de Tweedie são encontrados como distribuição Normal ( $p = 0$ ), distribuição Poisson ( $p = 1$ ), distribuição *Gama* ( $p = 2$ ) e distribuição Poisson Composta ( $1 < p < 2$ ), tais distribuições serão demonstradas nas subseções seguintes.

#### 3.1.1 Distribuição Normal, $p = 0$

Já sabemos da seção 2.2, que

$$\begin{aligned} k(\theta) &= \theta^2/2 \implies \tau(\theta) = \frac{\partial k(\theta)}{\partial \theta} = \theta, \\ \tau'(\theta) &= \frac{\partial \tau(\theta)}{\partial \theta} = 1 \implies V(\mu) = 1 = \mu^0, \\ \text{como } V_p(\mu) &= \mu^p = \mu^0, \\ \text{temos que } p &= 0. \end{aligned}$$

#### 3.1.2 Distribuição Gamma, $p = 2$

Na subseção 2.2.3 vimos que :

$$\begin{aligned} \mu &= -(1/\theta) \implies k(\theta) = -\log(-\theta), \\ \tau(\theta) &= \frac{\partial k(\theta)}{\partial \theta} = \frac{-1}{\theta}, \\ \tau'(\theta) &= \frac{\partial \tau(\theta)}{\partial \theta} = \frac{1}{\theta^2} = \frac{1}{1/\mu^2} = \mu^2 \implies V(\mu) = 1 = \mu^2, \\ \text{como } V_p(\mu) &= \mu^p = \mu^2, \text{ temos que } p = 2. \end{aligned}$$

#### 3.1.3 Distribuição Poisson, $p = 1$

Uma variável aleatória  $Z$  segue uma distribuição Poisson de média  $m$ ,  $Po(m)$ , se sua função densidade de probabilidade tem a forma

$$p(z; m) = \frac{m^z}{z!} e^{-m} = \frac{1}{z!} \exp\{z \log m - m\}, \quad z \in \mathbb{N}.$$

Se  $\theta = \log m$ , então

$$p(z; \theta) = \frac{1}{z!} \exp \{z\theta - e^\theta\},$$

que é uma família exponencial natural com

$$\begin{aligned} k(\theta) &= e^\theta, \quad \Theta = R, \\ \mu &= m, \\ V(\mu) &= \mu, \quad \Omega = R_+. \end{aligned}$$

Para incluir a Poisson no mundo das famílias exponenciais de convolução  $ED^*(\theta, \lambda)$ , fazemos

$$m = \lambda e^\theta,$$

para obter

$$p^*(z; \theta, \lambda) = \frac{\lambda^z}{z!} \exp \{z\theta - \lambda e^\theta\}.$$

A propriedade de convolução aplicada a essa família mostra que "soma de Poisson é Poisson"

A partir disso temos que

$$\begin{aligned} k(\theta) &= e^\theta, \\ \tau(\theta) &= \frac{\partial k(\theta)}{\partial \theta} = e^\theta = \mu, \\ \tau'(\theta) &= \frac{\partial \tau(\theta)}{\partial \theta} = e^\theta, \quad V(\mu) = \mu, \\ \text{sendo } V_p(\mu) &= \mu^p = \mu, \text{ temos que } p = 1. \end{aligned}$$

### 3.1.4 Distribuição Poisson-Composta, $p = 1/2$

Seja  $N, Z_1, Z_2, \dots, Z_n$  uma sequência de variáveis aleatórias independentes, tal que  $N$  segue Poisson com parâmetro  $m$ ,  $Po(m)$ , e as  $Z_i$ 's são identicamente distribuídos. Definindo

$$Z = \sum_{i=1}^N Z_i, \quad (3.1)$$

onde  $Z = 0$  para  $N = 0$ , a distribuição de  $Z$  é chamada de distribuição de Poisson Composta.

Quando as variáveis aleatórias  $Z_1, Z_2, \dots, Z_n$  seguem uma distribuição  $Ga(-\theta, -\alpha)$ , a distribuição resultante de  $Z$  é a forma aditiva de convolução do modelo de dispersão exponencial,  $ED^*(\theta, \lambda)$ , e, portanto,  $Z/\lambda$  é a forma reprodutiva de convolução do modelo de dispersão exponencial,  $ED(\mu, \sigma^2)$ , como afirma o seguinte resultado, proposto e demonstrado em Canal, G. Y. (1992), cuja prova é apresentada no Apêndice A.

**Resultado:** A variável aleatória  $Z$  dada em (3.1) com  $Z_i \sim Ga(-\theta, -\alpha)$  é uma distribuição  $ED^*(\theta, \lambda)$  com função densidade dada por

$$\begin{aligned} f_{Z,N}(z, n; \theta, \lambda, \alpha) &= \frac{(-\theta)^{-n\alpha} m^n z^{-n\alpha-1} \exp(\theta z - m)}{\Gamma(-n\alpha) n!} \\ &= \frac{\{\lambda k_p(-1/z)\}^n}{\Gamma(-n\alpha) n! z} \exp\{\theta z - \lambda k_p(\theta)\} \\ P(N=0) = P(Z=0) &= \exp\{-\lambda k_p(\theta)\} \end{aligned}$$

onde

$$\begin{aligned} K_p(\theta) &= \frac{\alpha - 1}{\alpha} \left( \frac{\theta}{\alpha - 1} \right)^\alpha, \\ V(\mu) &= \mu^p \quad \text{com } p \in (1, 2). \end{aligned}$$

Aplicando a transformação de dualidade  $Z = \lambda Y, Y \sim ED(\mu, \sigma^2)$ , obtemos a seguinte função densidade de probabilidade

$$f(y; \theta, \lambda, \alpha) = \sum_{n=1}^{\infty} \frac{\{\lambda k_p(-(\lambda y)^{-1})\}^n}{\Gamma(-n\alpha) n! y} \exp[\lambda\{\theta y - k_p(\theta)\}], \quad y \geq 0.$$

Uma aplicação do Modelo de Tweedie Poisson Composto é apresentada em Jorgensen & Souza (1992). O estudo tinha como objetivo principal obter prêmios de seguros con-



siderando a taxa média de valores de sinistros,  $\mu$ , como uma função de covariáveis independentes Neste trabalho, todo o estudo foi desenvolvido via modelos lineares generalizados.

# Capítulo 4

## Inferência Clássica para o Modelo e Tweedie

Em nosso trabalho apresentaremos uma aplicação para o Modelo de Tweedie Poisson Composto, tendo como objetivo principal obter o valor médio individual de sinistros por segurado considerando as variáveis "número total de sinistros por grupo",  $N$ , e "custo individual de sinistros",  $Z_i$ . Nesta parte do trabalho consideramos  $N$  seguindo uma Poisson com parâmetro  $m$ ,  $Po(m)$ , e  $Z_i$  seguindo uma distribuição Gama com parâmetros  $-\theta$  e  $-\alpha$ ,  $Ga(-\theta, -\alpha)$ .

Assim a distribuição condicional,  $Z|N$ , segue, pela propriedade de convolução, uma distribuição Gama com parâmetros  $-\theta$  e  $-n\alpha$ ,  $\sim Ga(\alpha n, \theta)$ . Desta forma a densidade conjunta de  $Z$  e  $N$ ,  $f_{Z,N}(z, n; \theta, \lambda, \alpha) = f_{Z|N}(z|n; \theta, \lambda, \alpha) f_N(n; \theta, \lambda, \alpha)$ , é dada por

$$f_{Z,N}(z, n; \theta, \lambda, \alpha) = \frac{(-\theta)^{-n\alpha}}{\Gamma(-n\alpha)} z^{n\alpha-1} e^{\theta z} \frac{m^n}{n!} e^{-m}, \quad (4.1)$$

para  $z > 0, n \geq 1$ .

Neste capítulo serão determinados os estimadores de máxima verossimilhança dos parâmetros envolvidos,  $\alpha$ ,  $\theta$  e  $m$ . O objetivo final consiste na estimação, através da inferência clássica, da razão  $\alpha/\theta$ , que corresponde ao custo médio de sinistros por segurado.

## 4.1 Modelo Estatístico e a Função de Verossimilhança

Denotemos por

$G$ : número de grupos de segurados;

$N_j$ : número total de sinistros por grupo  $j$ ,  $j = 1, 2, \dots, G$ ;

$Z_i$ : custo de sinistros por segurado,  $i = 1, \dots, N_j$ ;

$T_j = \sum_{i=1}^{N_j} Z_{ij}$ : custo total de sinistros do  $j$ -ésimo grupo,  $j = 1, 2, \dots, G$ ;

$T_j|N_j$ : custo total de um determinado grupo de segurados dado o número de sinistros.

Assumindo  $\alpha > 0$ ,  $\theta > 0$ ,  $n \geq 1$  e  $z > 0$ , temos que a função de densidade conjunta em(4.1) pode ser reescrita como

$$f_{Z,N}(z, n; m, \alpha, \theta) = \frac{\theta^{\alpha n}}{\Gamma(\alpha n)} z^{\alpha n - 1} e^{-\theta z} \frac{m^n e^{-m}}{n!}. \quad (4.2)$$

Considerando que  $j = 1, \dots, G$ , onde  $G$  é o número de grupos de segurados e  $N_j$  o número total de sinistros dentro do grupo  $j$ , e definindo  $T_j = \sum_{i=1}^{N_j} Z_{ij}$ , para  $i = 1, \dots, N_j$ , como sendo o custo total de sinistros do  $j$ -ésimo grupo, teremos  $T_1, T_2, \dots, T_G$  como uma sequência de variáveis aleatórias independentes e identicamente distribuídos e a distribuição condicional do custo total de um determinado grupo de segurados dado o número de sinistros, supondo  $\alpha_1 = \alpha_2 \dots = \alpha_G = \alpha$ , é uma Gama com parâmetros  $\alpha n_j$  e  $\theta$ ,  $Ga(\alpha n_j, \theta)$ , ou seja

$$T_j|N_j = n_j \sim Ga(\alpha n_j, \theta). \quad (4.3)$$

Considerando que  $T_j|N_j \sim Ga(\alpha n_j, \theta)$  e que estejam disponíveis os pares de dados  $(n_1, t_1), (n_2, t_2), \dots, (n_G, t_G)$ , que correspondem ao número total de sinistro e custo total destes sinistros por grupo, supondo  $\mathbf{n} = (n_1, n_2, \dots, n_G)$  e  $\mathbf{t} = (t_1, t_2, \dots, t_G)$ , a função de verossimilhança é dada por

$$L(\alpha, \theta, m | \mathbf{t}, \mathbf{n}) = \prod_{j=1}^G \frac{\theta^{\alpha n_j}}{\Gamma(\alpha n_j)} t_j^{\alpha n_j - 1} e^{-\theta t_j} \frac{m^{n_j} e^{-m}}{n_j!},$$

ou seja,

$$L(\alpha, \theta, m \mid \mathbf{t}, \mathbf{n}) = \frac{\theta^{\alpha \sum_{j=1}^G n_j}}{\prod_{j=1}^G \Gamma(\alpha n_j)} \cdot \prod_{j=1}^G t_j^{\alpha n_j - 1} e^{-\theta \sum_{j=1}^G t_j} \frac{m^{\sum_{j=1}^G n_j} e^{-Gm}}{\prod_{j=1}^G n_j!}. \quad (4.4)$$

A função de verossimilhança resume toda a informação relevante contida nos dados sobre os parâmetros do modelo.

## 4.2 Estimativas de Máxima Verossimilhança

Considerando a função de verossimilhança definida em (4.4), temos que o logaritmo dessa função é dado por

$$\begin{aligned} \ln [L(\alpha, \theta, m \mid \mathbf{t}, \mathbf{n})] &= \alpha \sum_{j=1}^G n_j \log \theta - \sum_{j=1}^G \log \Gamma(\alpha n_j) + \sum_{j=1}^G (\alpha n_j - 1) \times \\ &\quad \log t_j - \theta \sum_{j=1}^G t_j + \sum_{j=1}^G n_j \log m - mG + \sum_{j=1}^G \log n_j!. \end{aligned}$$

Assim, os estimadores de  $\alpha$ ,  $\theta$  e  $m$  são tais que satisfazem o seguinte sistema de equações

$$\frac{\partial \ln [L(\alpha, \theta, m \mid \mathbf{t}, \mathbf{n})]}{\partial \alpha} = \sum_{j=1}^G n_j \log \theta - \sum_{j=1}^G \frac{1}{\Gamma(\alpha n_j)} + \sum_{j=1}^G n_j \log t_j = 0; \quad (4.5)$$

$$\frac{\partial \ln [L(\alpha, \theta, m \mid \mathbf{t}, \mathbf{n})]}{\partial \theta} = \frac{\alpha \sum_{j=1}^G n_j}{\theta} - \sum_{j=1}^G t_j = 0 \quad (4.6)$$

$$\frac{\partial \ln [L(\alpha, \theta, m \mid \mathbf{t}, \mathbf{n})]}{\partial m} = \frac{\sum_{j=1}^G n_j}{m} - G = 0 \quad (4.7)$$

Os estimadores de máxima verossimilhança de  $\theta$  e  $m$  são facilmente obtidos das equações (4.6) e (4.7) e são dados respectivamente por

$$\hat{\theta} = \frac{\hat{\alpha} \sum_{j=1}^G n_j}{\sum_{j=1}^G t_j};$$

$$\hat{m} = \frac{\sum_{j=1}^G n_j}{G}.$$

Porém, o estimador de máxima verosimilhança de  $\alpha$  não tem forma analítica conhecida, sua expressão é dada por

$$\sum_{j=1}^G \frac{1}{\Gamma(\alpha n_j)} = \sum_{j=1}^G n_j \log \theta + \sum_{j=1}^G n_j \log t_j.$$

Assim, o estimador de  $\alpha$  é obtido somente através de um procedimento iterativo. Consequentemente,  $\hat{\theta}$  também será obtido através de um procedimento iterativo.

Estes estimadores serão determinados através do método iterativo de *Newton-Raphson* (Seber e Wild, 1989). Este método é utilizado para determinar raízes de funções baseado em um processo iterativo dado pela seguinte fórmula :

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

A sequência gerada desta maneira converge para uma raiz de  $f(x)$ , onde  $x_1$  é um chute inicial para a raiz.

Geometricamente o método de *Newton-Raphson* funciona traçando tangentes à curva nas aproximações de  $x$  e tomando a próxima aproximação dada pela intersecção da reta tangente com os eixos dos  $x$ 's.

### 4.3 Intervalos de Confiança

Um procedimento comum em Inferência Estatística é a utilização da aproximação normal para a obtenção de intervalos de confiança para parâmetros, ou funções de parâmetros, de interesse. Entretanto, estes procedimentos não apresentam, em alguns casos, precisão razoável quando o número de amostras é pequeno.

Para o estimador de máxima verossimilhança  $\hat{m} = \frac{\sum_{j=1}^G N_j}{G}$ , definindo  $S = \sum_{j=1}^G N_j$ , pela propriedade de convolução, temos que  $S$  segue distribuição Poisson( $Gm$ ), Desta forma,

$$P\left(\hat{m} = \frac{c}{G}\right) = P(S = c) = \frac{\exp(-Gm) (Gm)^c}{c!}$$

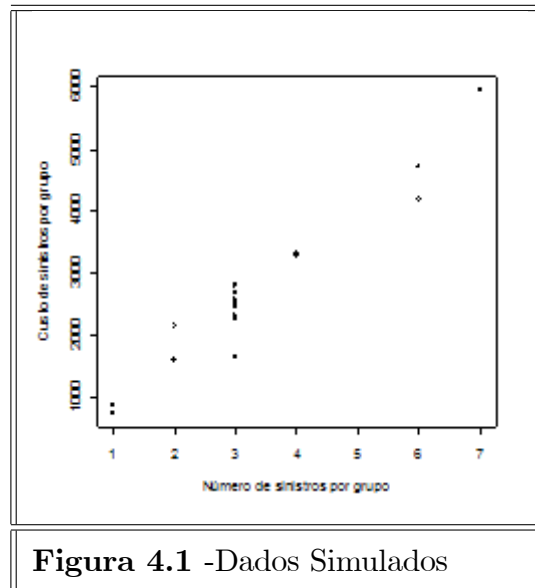
para  $c = 0, 1, 2, \dots$ , que especifica a distribuição exata de  $\hat{m}$ . Assim, o intervalo de confiança para o parâmetro  $m$  pode ser definido como

$$\hat{m} \pm 1,96\sqrt{\frac{\hat{m}}{G}}$$

Como existem dificuldades analíticas para encontrarmos intervalos de confiança exatos para os parâmetros de interesse  $\alpha$  e  $\theta$  usaremos um processo de simulação de Monte Carlo. Com esta técnica podemos obter as estimativas intervalares empíricas para os parâmetros de interesse. A principal idéia é simular conjuntos de dados provenientes da mesma distribuição, repetir esse processo um grande número de vezes e, em cada conjunto de dados calcular os valores de interesse. A partir de uma grande número de repetições do processo é possível determinar as distribuições empíricas dos estimadores dos parâmetros de interesse e, portanto, determinar as estimativas intervalares empíricas. Uma outra forma de determinarmos tais estimativas intervalares empíricas é através de técnicas bootstrap.

## 4.4 Aplicação dos Estimadores em Conjuntos de Dados Simulados

Para a aplicação dos estimadores definidos na Seção 4.1, dois conjuntos de dados compostos por 20 grupos de segurados foram simulados no software R. O primeiro conjunto envolve o número de sinistros de 20 grupos de segurados, gerados de uma distribuição Poisson(3) e o segundo conjunto envolve o custo total de um determinado grupo de segurados, dado o número de sinistros. Este segundo conjunto foi gerado de uma distribuição Gama( $16.n_j, 0.02$ ), onde  $n_j$  é o número de sinistros por grupo,  $n_j \geq 1$ , e, neste exemplo, o custo individual de sinistros, em média, é próximo de 800.



A Figura 4.1 ilustra os dados simulados. Os custos dos sinistros simulados variam entre o mínimo de R\$720,81 e o máximo de R\$5.951,91, correspondendo aos números de sinistros por grupo iguais a 1 e 7, respectivamente.

As estimativas de  $\alpha$ ,  $\theta$  e  $m$ , assim como uma comparação entre os valores originais dos parâmetros e suas estimativas, são apresentadas na Tabela 1.

Os I.C's apresentados na Tabela 1 para os parâmetros  $\alpha$  e  $\theta$  são empíricos, pois foram obtidos via simulação.

**Tabela 4.1** - Médias, erros padrões das estimativas dos parâmetros e I.C's

Parâmetros	Real	Média	ErroPadrão	I.C.
$\alpha$	16,0000	21,6025	6.8113	(14, 1160; 21, 2720)
$\theta$	0,0200	0,0265	0.00839	(0,01762; 0,02638)
$m$	3,0000	3.3000	1.5593	(0.2436; 6.3563)

Analisando a Tabela 4.1 verificamos que as estimativas de  $\theta$  e  $m$  estão mais próximas dos valores reais dos parâmetros, ao contrário de  $\alpha$ , cujo valor é menos preciso. Por outro lado, como podemos perceber pelo resultado apresentado na Tabela 4.2, a estimativa do custo medio individual, que depende das estimativas de  $\alpha$  e  $\theta$ , esta bem próximo de seu valor real.

**Tabela 4.2** - Custo Médio Individual de Sinistros

Fórmula	Real	Estimado
$\frac{\alpha}{\theta}$	800,0000	814,2065



# Capítulo 5

## Uma abordagem Bayesiana para o Modelo de Tweedie

Neste capítulo, será apresentada uma abordagem Bayesiana aplicada ao modelo de Tweedie. Similar ao Capítulo 4, as variáveis consideradas serão o número total de sinistros por grupo,  $N$ , e o custo individual de sinistros,  $Z_i$ . As distribuições serão as mesmas como definidas anteriormente, ou seja,  $N$  seguindo uma Poisson com parâmetro  $m$ ,  $Po(m)$ ,  $Z_i$  seguindo uma distribuição Gama com parâmetros  $-\theta$  e  $-\alpha$ ,  $Ga(-\theta, -\alpha)$ , a densidade conjunta de  $Z$  e  $N$ , como definida em (4.2) e a distribuição condicional,  $T_j|N_j$ , como definida em (4.3). O objetivo final do desenvolvimento da metodologia é a estimação, através da inferência Bayesiana, do custo médio de sinistros por segurado.

### 5.1 Metodologia

A análise Bayesiana consiste de métodos inferências alternativos que permitem incorporar ao modelo informações prévias a respeito dos parâmetros de interesse, a incorporação dessas informações podem melhorar a precisão dos resultados. A inferência Bayesiana considera os parâmetros do modelo como variáveis aleatórias e, portanto, distribuições a *priori* devem ser atribuídas aos parâmetros. Estas distribuições podem ser consideradas informativas, nos casos onde temos alguma informação disponível sobre os parâmetros do modelo, ou não informativa, nos casos onde não temos nenhuma informação a respeito dos parâmetros. Própria quando a distribuição atribuída a *priori* é uma distribuição de

probabilidade e imprópria quando a distribuição atribuída não é uma distribuição de probabilidade.

A escolha de uma distribuição a *priori* apropriada é fundamental para garantir a qualidade da inferência Bayesiana. Após atribuir as distribuições a *priori* para os parâmetros do modelo, combinamos estas distribuições com a função de verossimilhança e encontramos, através da utilização do teorema de Bayes, a distribuição a *posteriori* para os parâmetros. Na análise Bayesiana toda a inferência de interesse é baseada nesta distribuição a *posteriori*.

É comum a inferência Bayesiana apresentar integrais intratáveis analiticamente, neste caso, utiliza-se métodos de aproximação, tais como Gibbs Sampling, Metropolis-Hastings, Gibbs com Metropolis, Método de Aceitação e Rejeição Adaptativo, entre outros. Esses métodos geram amostras da distribuição de interesse fornecendo, assim, suas características de interesse.

A convergência das amostras geradas a partir dos métodos computacionais citados anteriormente tem grande importância. Existem vários métodos para o diagnóstico de convergência e um dos mais utilizados é o método de Gelman&Rubin (1992). Este método utiliza técnicas de análise de variância, comparando a variância dentro das cadeias com a variância entre as cadeias.

## 5.2 Inferência Bayesiana sobre os parâmetros $m$ , $\alpha$ e $\theta$

Considerando a verossimilhança, como definida no Capítulo 4, dada por

$$L(\alpha, \theta, m \mid \mathbf{t}, \mathbf{n}) = \frac{\theta^{\alpha \sum_{j=1}^G n_j}}{\prod_{j=1}^G \Gamma(\alpha n_j)} \prod_{j=1}^G t_j^{\alpha n_j - 1} e^{-\theta \sum_{j=1}^G t_j} \frac{m^{\sum_{j=1}^G n_j} e^{-Gm}}{\prod_{j=1}^G n_j!}, \quad (5.1)$$

e as distribuições a *priori* para  $m$ ,  $\alpha$ ,  $\theta$  dadas por

$$m \sim Ga(a; b)$$

$$\alpha \sim Ga(c; d)$$

$$\theta \sim Ga(r; s),$$

pelo Teorema de Bayes, supondo independência entre  $m$ ,  $\alpha$  e  $\theta$ , temos que a distribuição a *posteriori* conjunta para  $(m, \alpha, \theta)$  é dada por

$$\pi(\alpha, \theta, m | \mathbf{t}, \mathbf{n}) = L(T, N) \pi(\alpha) \pi(\theta) \pi(m).$$

Ou seja,

$$\begin{aligned} \pi(\alpha, \theta, m | \mathbf{t}, \mathbf{n}) &= \frac{\theta^{\alpha \sum_{j=1}^G n_j}}{\prod_{j=1}^G \Gamma(\alpha n_j)} \prod_{j=1}^G t_j^{\alpha n_j - 1} e^{-\theta \sum_{j=1}^G t_j} \frac{m^{\sum_{j=1}^G n_j} e^{-Gm}}{\prod_{j=1}^G n_j!} \frac{b^a}{\Gamma(a)} m^{a-1} e^{-bm} \times \\ &\quad \frac{d^c}{\Gamma(c)} \alpha^{c-1} e^{-d\alpha} \frac{s^r}{\Gamma(r)} \theta^{r-1} e^{-s\theta}. \end{aligned}$$

Assim a condicional de  $\alpha$  é dada por

$$i) \pi(\alpha | \theta, m, \mathbf{t}, \mathbf{n}) \propto \frac{\theta^{\alpha \sum_{j=1}^G n_j}}{\prod_{j=1}^G \Gamma(\alpha n_j)} \prod_{j=1}^G t_j^{\alpha n_j - 1} \alpha^{c-1} e^{-d\alpha},$$

onde  $\alpha^{c-1} e^{-d\alpha}$  corresponde ao núcleo de uma  $Ga(c; d)$ . A condicional de  $\theta$  é dada por

$$ii) \pi(\theta | \alpha, m, \mathbf{t}, \mathbf{n}) \propto \theta^{\alpha \sum_{j=1}^G n_j + r - 1} e^{-\theta \sum_{j=1}^G t_j - \theta s},$$

que corresponde a uma distribuição  $Ga(\alpha \sum_{j=1}^G n_j + r; \sum_{j=1}^G t_j + s)$ . E a condicional de  $m$  é dada por

$$iii) \pi(m | \alpha, \theta, \mathbf{t}, \mathbf{n}) \propto m^{\sum_{j=1}^G n_j + a - 1} e^{-m(b+G)},$$

que corresponde a uma distribuição  $Ga\left(\sum_{j=1}^G n_j + a; b + G\right)$ .

Neste trabalho consideraremos apenas piores não informativas para  $\alpha$ ,  $\theta$  e  $m$ . Para isto, foram impostos aos hiperparâmetros  $a$ ,  $b$ ,  $c$ ,  $d$ ,  $r$  e  $s$  os valores 0,001.

### 5.2.1 Estimação dos Parâmetros

Utilizando o algoritmo computacional Gibbs com Metropolis foram geradas amostras para  $\alpha$ ,  $\theta$  e  $m$ . O algoritmo Gibbs sampling é um método de simulação via cadeias de Markov que é utilizado na obtenção de amostras de uma distribuição de interesse baseada nas distribuições condicionais a *posteriori* dos parâmetros. Esse algoritmo foi utilizado para os parâmetros  $m$  e  $\theta$  uma vez que as distribuições condicionais apresentavam a forma de uma densidade conhecida.

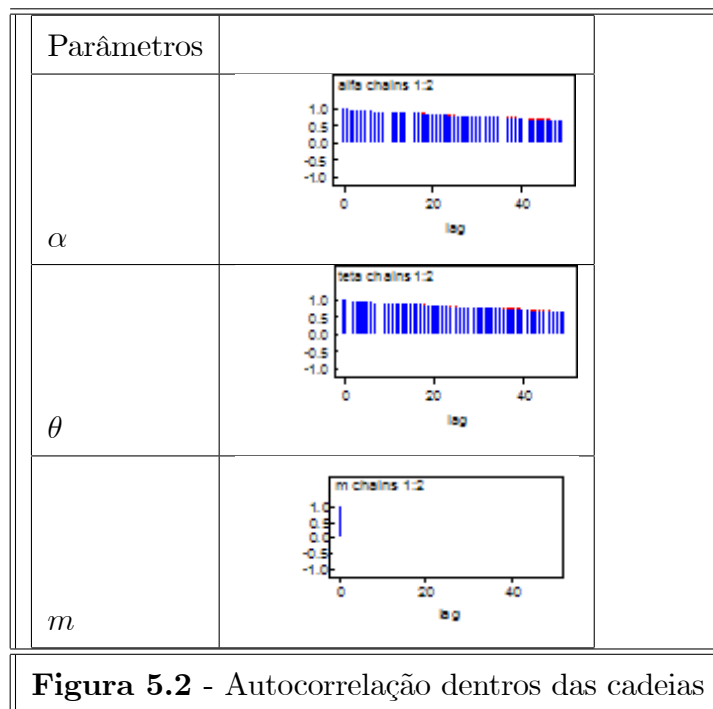
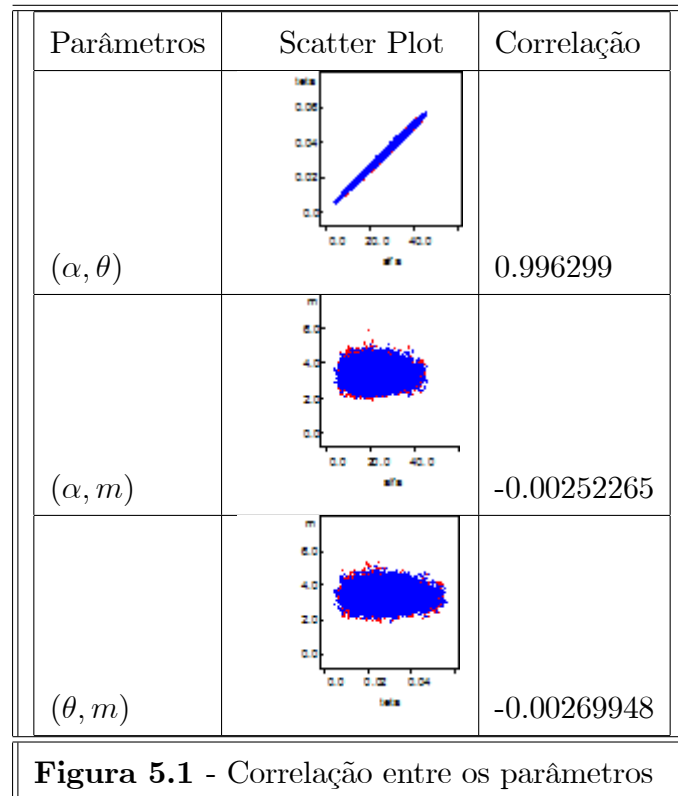
Na obtenção de amostras da distribuição do parâmetro  $\alpha$  foi utilizado o algoritmo de Metropolis-Hasting, pois a distribuição condicional não apresentava a forma de uma densidade conhecida.

Um problema grave ao fazer inferência utilizando simulação via cadeias de Markov é que ao término das iterações podem haver áreas do espaço paramétricos que não foram visitadas pela cadeia finita. Isto deve-se ao fato das simulações moverem-se muito lentamente, ou estarem presas à regiões restritas da distribuição de interesse.

Gelman & Rubin (1992) verificam que, quando analisadas separadamente, algumas cadeias parecem ter convergido, mas replicações independentes, com pontos de partida dispersos, podem mostrar que a convergência ainda não ocorreu. Isto pode ser causado por uma alta correlação entre os parâmetros e nestes casos uma reparametrização pode auxiliar a reduzir a autocorrelação.

### 5.2.2 Análise de convergência e auto-correlação dos parâmetros de interesse

Nesta seção é analisada a existência de convergência e alta correlação entre os parâmetros. Inicialmente são apresentadas algumas figuras, pelas quais pode-se avaliar a correlação entre os parâmetros, a autocorrelação dentro das cadeias e o comportamento das cadeias.



Observando as Figuras 5.1 e 5.2, nota-se que os parâmetros  $\alpha$  e  $\theta$  apresentam uma forte correlação e também uma autocorrelação acentuada. Para minimizar a presença de correlação e tentar diminuir a autocorrelação será utilizada a seguinte reparametrização para o parâmetro  $\theta$ ,

$$\varphi = \frac{\theta}{\alpha},$$

ou seja  $\theta = \alpha\varphi$ .

Considerando  $\theta = \alpha\varphi$  e a função de verossimilhança em (5.1), obtém-se a função de verossimilhança para os parâmetros  $\alpha$ ,  $\varphi$  e  $m$ , dada por

$$L(\alpha, \varphi, m \mid \mathbf{t}, \mathbf{n}) = \frac{(\alpha\varphi)^{\alpha \sum_{j=1}^G n_j}}{\prod_{j=1}^G \Gamma(\alpha n_j)} \prod_{j=1}^G t_j^{\alpha n_j - 1} e^{-(\alpha\varphi) \sum_{j=1}^G t_j} \frac{m^{\sum_{j=1}^G n_j} e^{-Gm}}{\prod_{j=1}^G n_j!}. \quad (5.2)$$

Assim, com a reparametrização, a condicional à *posteriori* para  $\alpha$  é dada por

$$i) \pi(\alpha \mid \varphi, m, \mathbf{t}, \mathbf{n}) \propto \frac{(\alpha\varphi)^{\left(\alpha \sum_{j=1}^G n_j + c + r - 1\right) - 1}}{\prod_{j=1}^G \Gamma(\alpha n_j)} \prod_{j=1}^G t_j^{\alpha n_j - 1} e^{-(\alpha\varphi) \sum_{j=1}^G t_j - s(\alpha\varphi)} e^{-d\alpha},$$

onde  $\alpha^{\left(\alpha \sum_{j=1}^G n_j + c + r - 1\right) - 1} e^{-\alpha \left(\varphi \sum_{j=1}^G t_j + \varphi s + d\right)}$  corresponde ao núcleo de uma Gama. A condicional à *posteriori* para  $\varphi$  é dada por

$$ii) \pi(\varphi \mid \alpha, m, \mathbf{t}, \mathbf{n}) \propto \varphi^{\alpha \sum_{j=1}^G n_j + r - 1} e^{-\varphi \left(\alpha \sum_{j=1}^G t_j + \alpha s\right)},$$

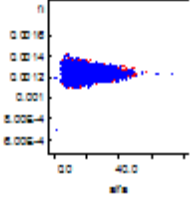
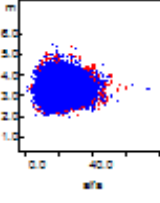
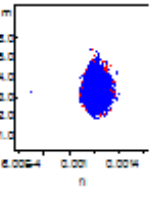
que corresponde a uma distribuição  $Ga\left(\alpha \sum_{j=1}^G n_j + r; \alpha \sum_{j=1}^G t_j + \alpha s\right)$ . E a condicional à *posteriori* para  $m$  é dada por

$$iii) \pi(m \mid \alpha, \theta, \mathbf{t}, \mathbf{n}) \propto m^{\sum_{j=1}^G n_j + a - 1} e^{-m(b+G)},$$

que corresponde a uma distribuição  $Ga\left(\sum_{j=1}^G n_j + a; b + G\right)$ .

### 5.2.3 Resultados apresentados pela reparametrização $\varphi = \frac{\theta}{\alpha}$

Nesta seção é feita a análise do efeito causado pela reparametrização na correlação existente entre os parâmetros.

Parâmetrizaçãõ	Scatter Plot	Correlaçãõ
$(\alpha, \varphi)$		0.00274242
$(\alpha, m)$		-9.83337E-4
$(\varphi, m)$		0.0155836

**Figura 5.3** - Correlaçãõ entre os parâmetros

Observando os resultados apresentados pela Figura 5.3 conclui-se que a reparametrizaçãõ proporcionou uma reduçãõ acentuada da correlaçãõ entre os parâmetros, o que reduz significativamente o número total de iterações necessárias ao alcance da convergência e os valores sugeridos como salto para a obtençãõ de baixa autocorrelaçãõ dentro das cadeias.

### 5.3 Resultados da Aplicaçãõ com os dados simulados

O conjunto de dados simulados para ilustrar a metodologia desenvolvida envolve duas variáveis. A primeira é o número de sinistros dos grupos de segurados. Esta variável foi gerada 20 vezes de uma distribuiçãõ Poisson com parâmetro (3). A segunda é o custo total de sinistros dos grupos de segurados condicionado ao número de sinistros do grupo. Esta variável foi gerada de uma distribuiçãõ  $Ga(16n_j, 0.02)$ .

Observe que, ao considerar a distribuiçãõ  $Ga(16n_j, 0.02)$  para  $T_j|N_j$ , a distribuiçãõ para  $Z_{ij}$ , o  $i$ -ésimo custo individual de sinistros dentro do grupo  $j$ , segue, pelas propriedades de convoluçãõ, distribuiçãõ  $Ga(16, 0.02)$ . Tem-se, entãõ, que em média este custo individual seja próximo de 800.

Considerando amostras de  $n_j$  e  $t_j$ ,  $j = 1, 2, \dots, 20$ , foram geradas 2 cadeias com 55.000 iterações, descartando as primeiras 5.000 como burn-in, e considerando um salto de 5, obtendo uma amostra final de 10.000 valores para  $\alpha$ ,  $\varphi$  e  $m$ .

Para a geração das amostras da distribuição a *posteriori* via amostrador de Gibbs e Metropolis-Hastings foi utilizado o software Winbugs (Spiegelhalter *et al*, 1996).

**Tabela 5.1** - Valores dos parâmetros estimados a *posteriori*

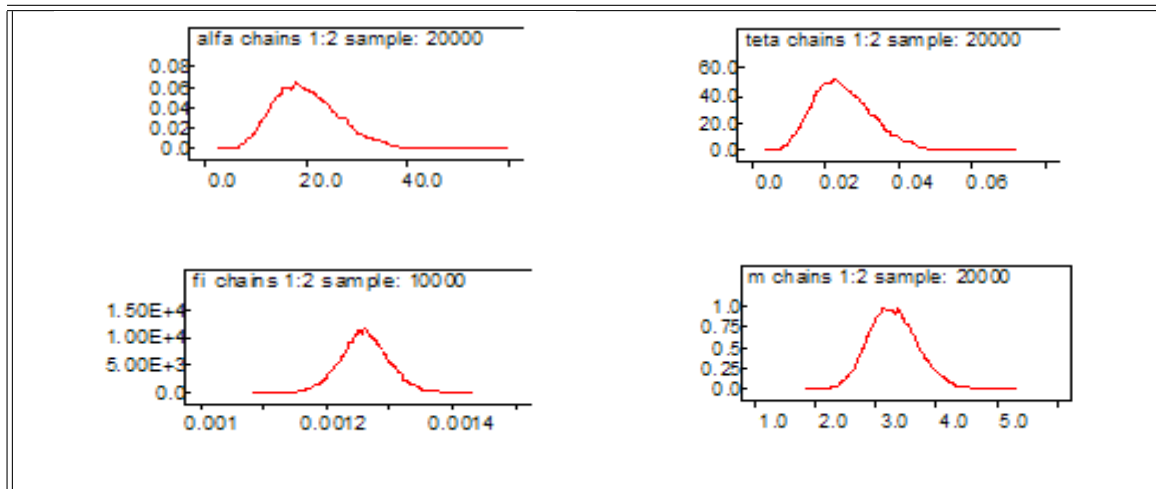
Parâmetros	Real	Média	DP	Região de Credibilidade
$\alpha$	16, 0000	20, 4500	6, 6400	(9, 6260; 35, 2600)
$\varphi$	0, 0012	0, 0012	0, 00003	(0, 0011; 0, 0012)
$\theta$	0, 0200	0, 0251	0, 0081	(0, 0117; 0, 0433)
$m$	3, 0000	3, 2990	0, 4072	(2, 5610; 4, 1530)

**Tabela 5.2** - Custo médio individual de sinistros

Parâmetros	Real	Média	DP	Região de Credibilidade
custo médio	800, 0000	814, 8000	23, 1100	(770, 50; 862, 20)

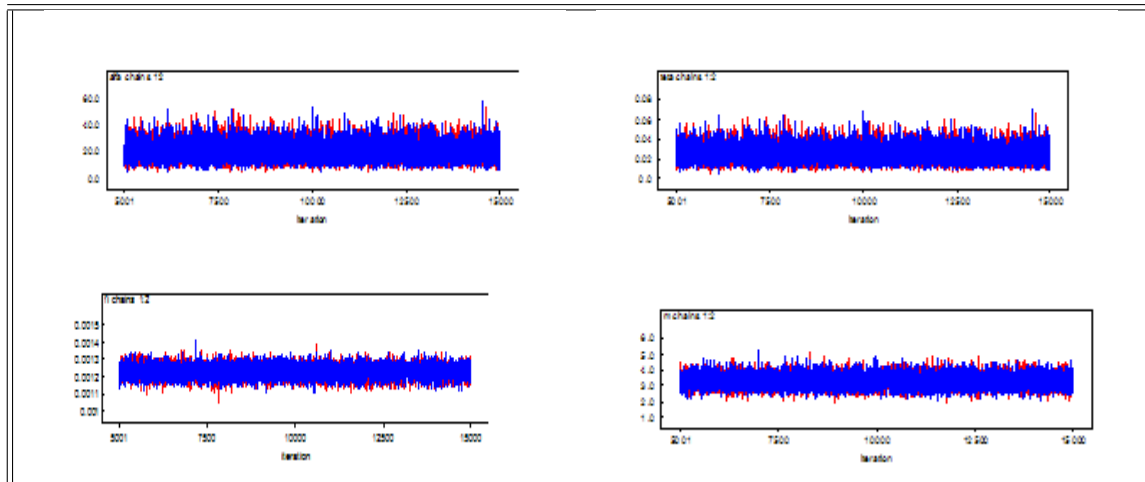
Os resultados a *posteriori* (Tabela 5.1) mostraram que as estimativas dos parâmetros são satisfatórias, pois estão próximas dos valores reais, assim como o valor do custo médio de sinistros estimado, observado na (Tabela 5.2).





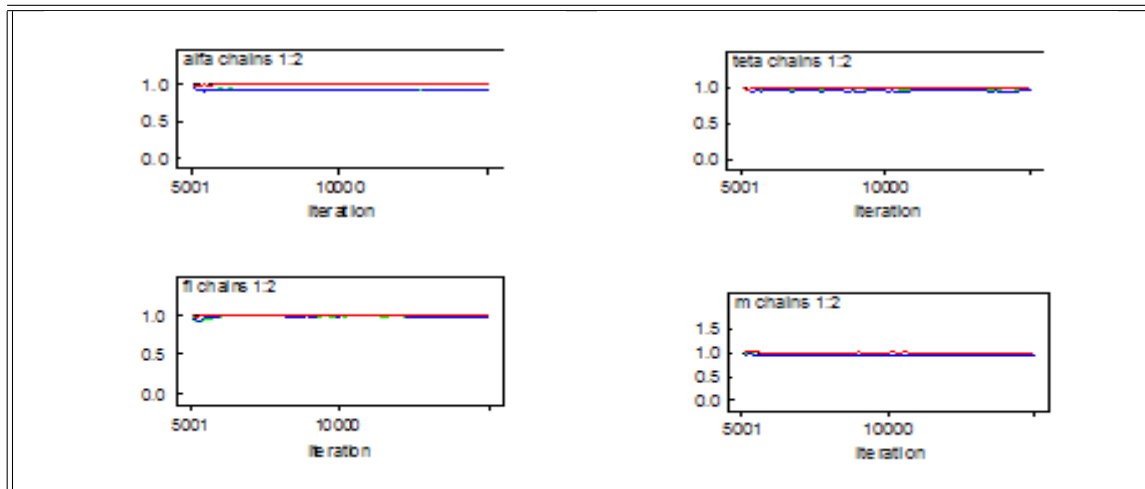
**Figura 5.4** - Densidade a Posteriori

A Figura 5.4 apresenta as densidades *a posteriori* dos parâmetros.



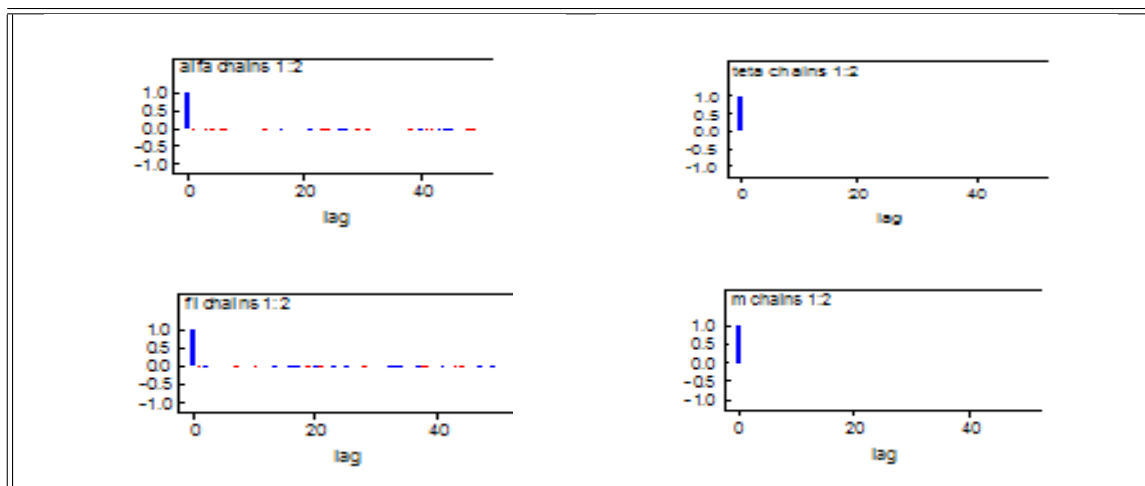
**Figura 5.5** - Traços a Posteriori

Através dos resultados gráficos da Figura 5.5 verifica-se a convergência para as cadeias devido a existência de uniformidade nos traços.



**Figura 5.6** - Testes Gráficos de Gelman&Rubin

Os valores obtidos na Figura 5.6, pelo diagnóstico de Gelman&Rubin (1992), são bem próximos de 1, o que indica a convergência das cadeias. Através do teste gráfico esta convergência também é verificada.



**Figura 5.7-** Auto Correlação

Observa-se da Figura 5.7 que os parâmetros apresentam baixa auto-correlação. Os resultados gráficos acima são obtidos considerando um salto igual a 5.

# Capítulo 6

## Abordagem Bayesiana para o Modelo de Tweedie envolvendo Covariáveis

Considerando que a determinação do prêmio de seguro depende de vários fatores inter-relacionados, como, por exemplo, idade, sexo, distância percorrida, entre outros, a estimação dos parâmetros será realizada reescrevendo, inicialmente, o parâmetro  $\alpha$  da distribuição Gama como função de covariáveis. O nosso objetivo, com isso, é obter estimativas mais precisas para os parâmetros, e conseqüentemente para o custo médio individual de sinistros.

### 6.1 Metodologia

Nessa seção é desenvolvida uma generalização do modelo de Tweedie na presença de covariáveis.

Considerando a distribuição condicional do custo total de um determinado grupo de segurados dado o número de sinistros, definida por

$$T_j | N_j = n_j \sim Ga(\alpha n_j, \theta),$$

onde  $j = 1, \dots, G$  determina o grupo de segurados e  $N_j$  o número total de sinistros dentro do

grupo  $j$ , ao atribuímos a  $\alpha$  uma função de covariáveis que incorpora as características dos segurados, teremos que adicionar a  $\alpha$  os índices que correspondem ao indivíduo e ao grupo, ou seja,  $\alpha_{ij}$ , onde  $i = 1, 2, \dots, n_j$ , determina o  $i$ -ésimo indivíduo com  $\alpha_{1j} \neq \alpha_{2j} \dots \neq \alpha_{n_j j}$

Desta forma, a distribuição condicional do custo total de um determinado grupo de segurados dado o número de sinistros do grupo é, por convolução, dada por

$$T_j | N_j = n_j \sim G(\alpha_{1j} + \alpha_{2j} + \dots + \alpha_{n_j j}, \theta) \sim Ga\left(\sum_{i=1}^{n_j} \alpha_{ij}, \theta\right). \quad (6.1)$$

Considerando que  $\alpha_{1j} \neq \alpha_{2j} \dots \neq \alpha_{n_j j}$ , a reparametrização apresentada no capítulo anterior,  $\varphi = \frac{\theta}{\alpha}$ , pode ser reescrita da seguinte maneira

$$\varphi = \frac{\theta \sum_{j=1}^G n_j}{\sum_{j=1}^G \sum_{i=1}^{n_j} \alpha_{ij}}$$

e assim, a densidade conjunta de  $T_j$  e  $N_j$ , para  $n \geq 1$  e  $z > 0$ , pode ser reescrita como

$$f_{T_j, N_j} = f_{T_j | N_j} \cdot f_{N_j},$$

onde

$$f_{(T_j, N_j)}(t_j, n_j; \alpha, \varphi, m) = \frac{\left(\sum_{j=1}^G \sum_{i=1}^{n_j} \alpha_{ij} \varphi / \sum_{j=1}^G n_j\right)^{\sum_{i=1}^{n_j} \alpha_{ij}}}{\Gamma\left(\sum_{i=1}^{n_j} \alpha_{ij}\right)} t_j^{\sum_{i=1}^{n_j} \alpha_{ij} - 1} e^{-\left(\sum_{j=1}^G \sum_{i=1}^{n_j} \alpha_{ij} \varphi / \sum_{j=1}^G n_j\right) t_j} \frac{m^{n_j} e^{-m}}{n_j!}.$$

Considerando que  $T_j \sim Ga\left(\sum_{i=1}^{n_j} \alpha_{ij}, \theta\right)$  e que estejam disponíveis os pares de dados  $(n_1, t_1), (n_2, t_2), \dots, (n_G, t_G)$ , que correspondem ao número total de sinistro por grupo e o custo total destes sinistros por grupo, supondo  $\mathbf{n} = (n_1, n_2, \dots, n_G)$  e  $\mathbf{t} = (t_1, t_2, \dots, t_G)$ , a função de verossimilhança em (5.2) é reescrita como

$$L(\alpha, \varphi, m \mid \mathbf{t}, \mathbf{n}) = \prod_{j=1}^G \frac{\left( \sum_{i=1}^{n_j} \alpha_{ij} \varphi / \sum_{j=1}^G n_j \right)^{\sum_{i=1}^{n_j} \alpha_{ij}}}{\Gamma \left( \sum_{i=1}^{n_j} \alpha_{ij} \right)} t_j^{\sum_{i=1}^{n_j} \alpha_{ij} - 1} e^{-\left( \sum_{i=1}^{n_j} \alpha_{ij} \varphi / \sum_{j=1}^G n_j \right) t_j} \frac{m^{n_j} e^{-m}}{n_j!},$$

ou seja,

$$L(\alpha, \varphi, m \mid \mathbf{t}, \mathbf{n}) = \frac{\left( \sum_{j=1}^G \sum_{i=1}^{n_j} \alpha_{ij} \varphi / \sum_{j=1}^G n_j \right)^{\sum_{j=1}^G \sum_{i=1}^{n_j} \alpha_{ij}}}{\prod_{j=1}^G \Gamma \left( \sum_{i=1}^{n_j} \alpha_{ij} \right)} \prod_{j=1}^G t_j^{\sum_{i=1}^{n_j} \alpha_{ij} - 1} \times \quad (6.2)$$

$$e^{-\left( \sum_{j=1}^G \sum_{i=1}^{n_j} \alpha_{ij} \varphi / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j} \frac{m^{\sum_{j=1}^G n_j} e^{-mG}}{\prod_{j=1}^G n_j!}$$

## 6.2 O Modelo na Presença de Uma Covariável

Considerando o parâmetro  $\alpha$  como uma função de covariável,  $\alpha = e^{\beta_0 + \beta_1 x_1}$ , onde a covariável  $x_1$  poderia corresponder, por exemplo, a idade ou a distância percorrida pelos segurados semanalmente, a função de verossimilhança em (6.2) é reescrita na forma

$$L(\beta_0, \beta_1, \varphi, m \mid \mathbf{t}, \mathbf{n}, \mathbf{x}_1) = \frac{\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1i j}} \varphi / \sum_{j=1}^G n_j \right)^{\sum_{j=1}^G \left( \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1i j}} \right)}}{\prod_{j=1}^G \Gamma \left( \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1i j}} \right)} \times$$

$$\prod_{j=1}^G t_j^{\sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1i j}} - 1} \times$$

$$e^{-\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1i j}} \varphi / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j} \frac{m^{\sum_{j=1}^G n_j} e^{-mG}}{\prod_{j=1}^G n_j!} \quad (6.3)$$

onde  $\mathbf{t} = (t_1, t_2, \dots, t_G)$ ,  $\mathbf{n} = (n_1, n_2, \dots, n_G)$  e  $\mathbf{x}_1 = (x_{111}, x_{121}, \dots, x_{1n_G G})$

### 6.2.1 Inferência Bayesiana sobre os parâmetros $\beta_0$ , $\beta_1$ , $\varphi$ e $m$

Para estimar os parâmetros de interesse serão utilizadas as seguintes distribuições *a priori* para  $\beta_0$ ,  $\beta_1$ ,  $\varphi$  e  $m$ .

$$\beta_0 \sim N(e, f)$$

$$\beta_1 \sim N(g, h)$$

$$\varphi \sim Ga(c, d)$$

$$m \sim Ga(a, b)$$

Pelo teorema de Bayes e considerando independência entre os parâmetros  $\beta_0$ ,  $\beta_1$ ,  $\varphi$  e  $m$ , temos que a distribuição *a posteriori* dos parâmetros é dada por

$$\pi(\beta_0, \beta_1, \varphi, m | \mathbf{t}, \mathbf{n}, \mathbf{x}_1) \propto L(T_j, N_j) \pi(\beta_0) \pi(\beta_1) \pi(\varphi) \pi(m),$$

ou, seja,

$$\begin{aligned}
\pi(\beta_0, \beta_1, \varphi, m | \mathbf{t}, \mathbf{n}, \mathbf{x}_1) &= \frac{\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij}} \varphi / \sum_{j=1}^G n_j \right)^{\sum_{j=1}^G \left( \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij}}) \right)}}{\prod_{j=1}^G \Gamma \left( \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij}}) \right)} \times \\
&\prod_{j=1}^G t_j^{\sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij}} - 1} e^{-\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij}} \varphi / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j} \times \\
&\frac{m^{\sum_{j=1}^G n_j} e^{-mG}}{\prod_{j=1}^G n_j!} \frac{b^a}{\Gamma(a)} m^{a-1} e^{-bm} \frac{d^c}{\Gamma(c)} \varphi^{c-1} e^{-d\varphi} \times \\
&\frac{1}{\sqrt{2\pi f}} \exp\left(\frac{-1}{2f} (\beta_0 - e)^2\right) \frac{1}{\sqrt{2\pi h}} \exp\left(\frac{-1}{2h} (\beta_1 - g)^2\right)
\end{aligned}$$

Assim, as condicionais de  $\beta_0$ ,  $\beta_1$ ,  $\varphi$  e  $m$ , são dadas por

$$\begin{aligned}
i) \pi(\beta_0 | \beta_1, \varphi, m, \mathbf{t}, \mathbf{n}, \mathbf{x}_1) &\propto \frac{\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij}} \varphi / \sum_{j=1}^G n_j \right)^{\sum_{j=1}^G \left( \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij}} \right)}}{\prod_{j=1}^G \Gamma \left( \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij}}) \right)} \times \\
&\prod_{j=1}^G t_j^{\sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij}}) - 1} e^{-\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij}} \varphi / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j} \times \\
&\exp\left(\frac{-1}{2f} (\beta_0 - e)^2\right);
\end{aligned}$$

onde  $\exp\left(\frac{-1}{2f} (\beta_0 - e)^2\right)$  corresponde ao núcleo de uma  $N(e, f)$ .

$$\begin{aligned}
ii) \pi(\beta_1 | \beta_0, \varphi, m, \mathbf{t}, \mathbf{n}, \mathbf{x}_1) &\propto \frac{\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij}} \varphi / \sum_{j=1}^G n_j \right)^{\sum_{j=1}^G \left( \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij}}) \right)}}{\prod_{j=1}^G \Gamma \left( \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij}}) \right)} \times \\
&\prod_{j=1}^G t_j^{\sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij}}) - 1} e^{-\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij}} \varphi / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j} \times \\
&\exp \left( \frac{-1}{2h} (\beta_1 - g)^2 \right);
\end{aligned}$$

sendo que  $\exp \left( \frac{-1}{2h} (\beta_1 - g)^2 \right)$  corresponde ao núcleo de uma  $N(g, h)$ .

$$iii) \pi(\varphi | \beta_0, \beta_1, m, \mathbf{t}, \mathbf{n}, \mathbf{x}_1) \propto \varphi^{\sum_{j=1}^G \left( \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij}}) \right) - 1} e^{-\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij}} \varphi / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j} \varphi^{c-1} e^{-d\varphi},$$

que corresponde a distribuição de uma gama

$$Ga \left( \sum_{j=1}^G \left( \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij}}) \right) + c, \left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij}} / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j + d \right).$$

$$iv) \pi(m | \beta_0, \beta_1, \varphi, \mathbf{t}, \mathbf{n}, \mathbf{x}_1) \propto m^{\sum_{j=1}^G n_j} e^{-mG} m^{a-1} e^{-bm},$$

que corresponde a distribuição de uma  $Ga \left( \sum_{j=1}^G n_j + a, G + b \right)$ .

### 6.3 O Modelo na presença de Duas Covariáveis

Atribuindo ao parâmetro  $\alpha$  a função de covariáveis  $\alpha = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}$ , supondo independência entre os parâmetros  $\beta'$ s, a função de verossimilhança (6.3) é reescrita na forma



$$\begin{aligned}
L(\beta_0, \beta_1, \beta_2, \varphi, m \mid \mathbf{t}, \mathbf{n}, \mathbf{x}_1, \mathbf{x}_2) &= \frac{\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} \varphi / \sum_{j=1}^G n_j \right)^{\sum_{j=1}^G \left( \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} \right)}}{\Gamma \left( \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} \right)} \times \\
&\prod_{j=1}^G t_j^{\sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} - 1} \times \\
&e^{-\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} \varphi / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j} \times \\
&\frac{\sum_{j=1}^G n_j e^{-mG}}{m \prod_{j=1}^G n_j!}
\end{aligned}$$

### 6.3.1 Inferência Bayesiana sobre os parâmetros $\beta_0, \beta_1, \beta_2, \varphi$ e

$m$

Para estimar os parâmetros de interesse, serão utilizadas as seguintes distribuições *a priori* para  $\beta_0, \beta_1, \beta_2, \varphi$  e  $m$

$$\beta_0 \sim N(e, f)$$

$$\beta_1 \sim N(g, h)$$

$$\beta_2 \sim N(r, s)$$

$$\varphi \sim Ga(c, d)$$

$$m \sim Ga(a, b)$$

Pelo teorema de Bayes e considerando independência entre os parâmetros  $\beta_0, \beta_1, \beta_2, \varphi$  e  $m$ , temos que a distribuição *a posteriori* dos parâmetros é dada por

$$\pi(\beta_0, \beta_1, \beta_2, \varphi, m \mid \mathbf{t}, \mathbf{n}, \mathbf{x}_1, \mathbf{x}_2) = L_{(T_j, N_j)} \pi(\beta_0) \pi(\beta_1) \pi(\beta_2) \pi(\varphi) \pi(m)$$

ou seja,

$$\begin{aligned} \pi(\beta_0, \beta_1, \beta_2, \varphi, m \mid \mathbf{t}, \mathbf{n}, \mathbf{x}_1, \mathbf{x}_2) &= \frac{\left( \prod_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} \varphi / \sum_{j=1}^G n_j \right)^{\sum_{j=1}^G \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}})}}{\prod_{j=1}^G \Gamma \left( \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}}) \right)} \times \\ &\prod_{j=1}^G t_j^{\sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} - 1} \times \\ &e^{-\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} \varphi / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j m^{\sum_{j=1}^G n_j} e^{-mG}} \times \\ &\frac{1}{\prod_{j=1}^G n_j!} \times \\ &\frac{b^a}{\Gamma(a)} m^{a-1} e^{-bm} \frac{d^c}{\Gamma(c)} \varphi^{c-1} e^{-d\varphi} \times \\ &\frac{1}{\sqrt{2\pi f}} \exp\left(\frac{-1}{2f} (\beta_0 - e)^2\right) \frac{1}{\sqrt{2\pi h}} \exp\left(\frac{-1}{2h} (\beta_1 - g)^2\right) \times \\ &\cdot \frac{1}{\sqrt{2\pi s}} \exp\left(\frac{-1}{2s} (\beta_2 - r)^2\right) \end{aligned}$$

onde  $\mathbf{t} = (t_1, t_2, \dots, t_G)$ ,  $\mathbf{n} = (n_1, n_2, \dots, n_G)$ ,  $\mathbf{x}_1 = (x_{111}, x_{121}, \dots, x_{1n_G G})$  e

$\mathbf{x}_2 = (x_{211}, x_{221}, \dots, x_{2n_G G})$

Assim as condicionais de  $\beta_0, \beta_1, \beta_2, \theta$  e  $m$  são dadas por

$$\begin{aligned}
i) \pi(\beta_0|\beta_1, \beta_2, \varphi, m, \mathbf{t}, \mathbf{n}, \mathbf{x}_1, \mathbf{x}_2) &\propto \frac{\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} \varphi / \sum_{j=1}^G n_j \right)^{\sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}}}}{\prod_{j=1}^G \Gamma \left( \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}}) \right)} \times \\
&\prod_{j=1}^G t_j^{\sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}}) - 1} \times \\
&e^{-\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} \varphi / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j} \times \\
&\exp \left( \frac{-1}{2f} (\beta_0 - e)^2 \right),
\end{aligned}$$

onde  $\exp \left( \frac{-1}{2f} (\beta_0 - e)^2 \right)$  corresponde ao núcleo de uma  $N(e, f)$ .

$$\begin{aligned}
ii) \pi(\beta_1|\beta_0, \varphi, m, \mathbf{t}, \mathbf{n}, \mathbf{x}_1, \mathbf{x}_2) &\propto \frac{\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} \varphi / \sum_{j=1}^G n_j \right)^{\sum_{j=1}^G \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}})}}{\prod_{j=1}^G \Gamma \left( \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}}) \right)} \times \\
&\prod_{j=1}^G t_j^{\sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}}) - 1} \times \\
&e^{-\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} \varphi / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j} \times \\
&\exp \left( \frac{-1}{2h} (\beta_1 - g)^2 \right),
\end{aligned}$$

sendo que  $\exp \left( \frac{-1}{2h} (\beta_1 - g)^2 \right)$  corresponde ao núcleo de uma  $N(g, h)$ .

$$\begin{aligned}
iii) \pi(\beta_2 | \beta_0, \beta_1, \varphi, m, \mathbf{t}, \mathbf{n}, \mathbf{x}_1, \mathbf{x}_2) &= \frac{\left( \prod_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} \varphi / \sum_{j=1}^G n_j \right)^{\sum_{j=1}^G (\sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}}))}}{\prod_{j=1}^G \Gamma \left( \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}}) \right)} \times \\
&\prod_{j=1}^G t_j^{\sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}}) - 1} \times \\
&e^{-\left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} \varphi / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j} \times \\
&\exp \left( \frac{-1}{2s} (\beta_2 - r)^2 \right),
\end{aligned}$$

onde  $\exp \left( \frac{-1}{2s} (\beta_2 - r)^2 \right)$  corresponde ao núcleo de uma  $N(r, s)$ .

$$\begin{aligned}
iv) \pi(\varphi | \beta_0, \beta_1, \beta_2, m, \mathbf{t}, \mathbf{n}, \mathbf{x}_1, \mathbf{x}_2) &\propto \frac{\sum_{j=1}^G \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}}) - (\varphi e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}}) \sum_{j=1}^G t_j}{\varphi^{c-1} e^{-d\varphi}} \times \\
&\varphi^{c-1} e^{-d\varphi},
\end{aligned}$$

que corresponde a distribuição de uma gama

$$Ga \left( \sum_{j=1}^G \sum_{i=1}^{n_j} (e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}}) + c, \left( \sum_{j=1}^G \sum_{i=1}^{n_j} e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}} / \sum_{j=1}^G n_j \right) \sum_{j=1}^G t_j + d \right)$$

$$v) \pi(m | \beta_0, \beta_1, \beta_2, \varphi, \mathbf{t}, \mathbf{n}, \mathbf{x}_1, \mathbf{x}_2) \propto m^{\sum_{j=1}^G n_j} e^{-mG} m^{a-1} e^{-bm},$$

que corresponde a distribuição  $Ga(\sum_{j=1}^G n_j + a, G + b)$ .

Serão consideradas apenas *prioris* não informativas para  $\beta_0, \beta_1, \beta_2, \theta$  e  $m$ . Para isto, foram impostos aos hiperparâmetros a,b,c,d, os valores 0,001, para os parâmetros  $e, g$  e  $r$  os valores 0 e para os parâmetros  $f, h$  e  $s$  os valores 1000.

## 6.4 Uma Aplicação em conjuntos de dados simulados

Utilizando o algoritmo computacional Gibbs com Metropolis, foram geradas amostras para  $\beta_0, \beta_1, \beta_2, \varphi$  e  $m$ . O algoritmo Gibbs foi utilizado para os parâmetros  $m$  e  $\theta$  uma

vez que as distribuições condicionais apresentam a forma de uma densidade conhecida.

Na obtenção de amostras da distribuição dos parâmetros  $\beta_0, \beta_1, \beta_2$  foi utilizado o algoritmo de Metropolis-Hasting, pois as distribuições condicionais não apresentavam a forma de uma densidade conhecida.

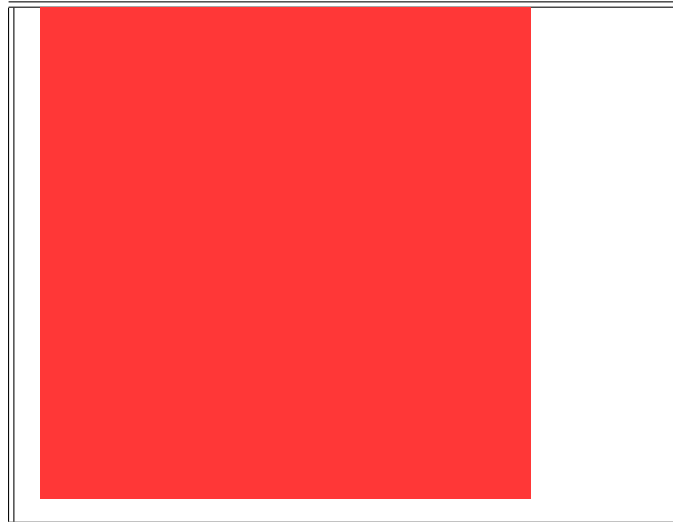
O conjunto de dados simulados para ilustrar a metodologia desenvolvida envolvia quatro variáveis. A primeira é o número de sinistros dos grupos de segurados. Esta variável foi gerada 20 vezes de uma distribuição Poisson com parâmetro (3). A segunda é o custo total de sinistros dos grupos de segurados condicionado ao número de sinistros do grupo. Esta variável foi gerada de uma distribuição  $Ga\left(\sum_{i=1}^{n_j} \alpha_{ij}, 0.02\right)$ .

A terceira variável é a idade de cada segurado sinistrado, gerada considerando idades entre 18 e 70 anos, esta variável foi gerada de uma distribuição  $U(18, 70)$ . E a quarta variável é a distância percorrida semanalmente pelos segurados sinistrados, gerada considerando valores entre 100 a 500Km. Esta variável foi gerada de uma distribuição  $U(100, 500)$ .

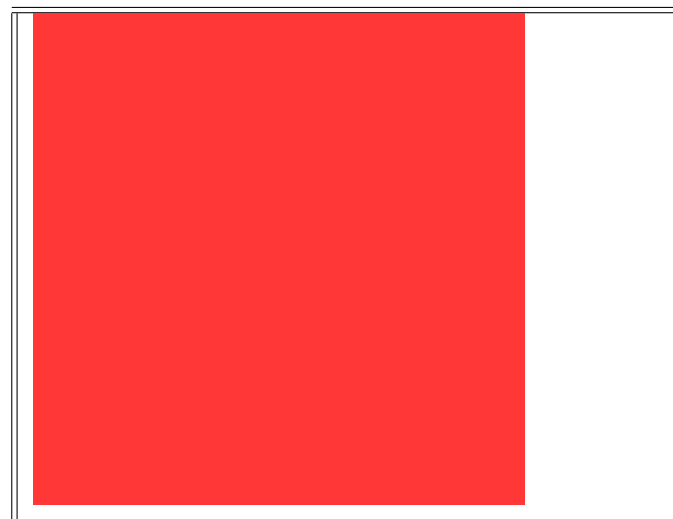
Com amostras de  $n_j$  e  $z_j$ ,  $j = 1, 2, \dots, 20$ , e  $x_1$  e  $x_2$  de tamanho 66 (o número total de sinistros de todos os grupos é igual a sessenta e seis sinistros), foram geradas cadeias com 105.000 iterações, descartando as primeiras 5.000 como burn-in, e considerando um salto de 10, obtendo assim, uma amostra final de 10.000 valores para  $\beta_0, \beta_1, \beta_2, \varphi$  e  $m$ .

Para a geração das amostras da distribuição a *posteriori* via amostrador de Gibbs e Metropolis-Hastings foi utilizado o software Winbugs (Spiegelhalter et al, 1996).

### 6.4.1 Ilustração dos Dados Simulados



**Figura 6.1** - Covariável Idade



**Figura 6.2** - Covariável Distância Percorrida

Observando a Figura 6.1, verificamos que os dados foram simulados considerando que indivíduos mais novos sinistram mais e conseqüentemente o valor dos sinistros são maiores. Na Figura 6.2 observamos que a simulação dos dados foi feita aleatoriamente, ou seja o custo de sinistros independe da distância percorrida pelo segurado.

### 6.4.2 Resultados referentes ao modelo na presença de uma covariável

Nesta subseção serão apresentados os resultados obtidos para o modelo na presença de uma covariável, para isto serão utilizadas as covariáveis idade e distância percorrida.

1- Covariável Idade

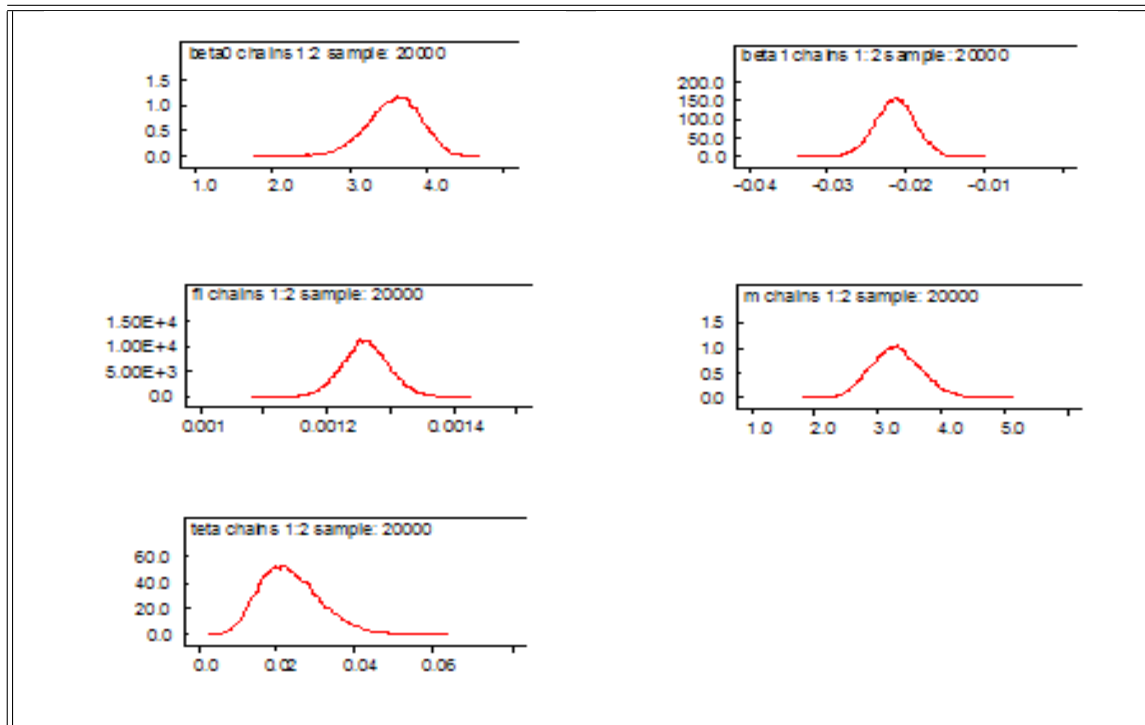
**Tabela 6.1** -Valores dos parâmetros estimados a *posteriori*

Parâmetros	Real	Média	DP	Região de Credibilidade
$\beta_0$	3,4000	3,5690	0,3576	(2,8100; 4,1950)
$\beta_1$	-0,0200	-0,0214	0,0026	(-0,0267; -0,0162)
$\varphi$	0,00125	0,00120	0,00003	(0,0012; 0,0013)
$m$	3,0000	3,3040	0,4037	(2,5610; 4,1400)
$\theta$	0,0200	0,0237	0,0079	(0,0107; 0,0413)
$\alpha$	15,8200	18,8500	6,2990	(8,5420; 32,7400)

**Tabela 6.2** - Custo Médio Individual de Sinistros

Fórmula	Real	Média	DP	Região de Credibilidade
$\frac{\sum_{j=1}^G \sum_{i=1}^{n_j} \alpha_{ij}}{\sum_{j=1}^G n_j \theta}$	795,8000	793,7000	23,8100	(747,6000; 842,7000)

As estimativas obtidas para os parâmetros a *posteriori* (Tabela 6.1) são satisfatórias, pois estão próximas dos valores reais, assim como a estimativa do custo médio (Tabela 6.2)



**Figura 6.3** - Densidade a Posteriori

Na Figura 6.3 observamos as densidades *a posteriori* dos parâmetros



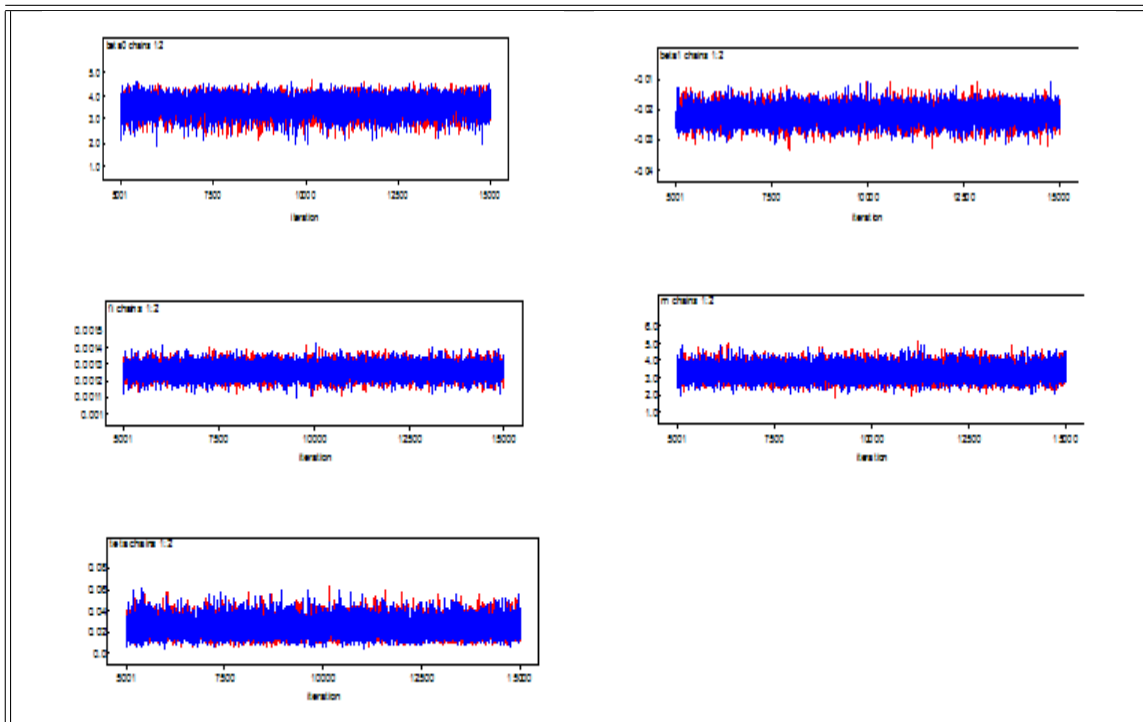


Figura 6.4 - Traços a Posteriori

A uniformidade existente nos gráficos apresentados na Figura 6.4 indica convergência para as cadeias.

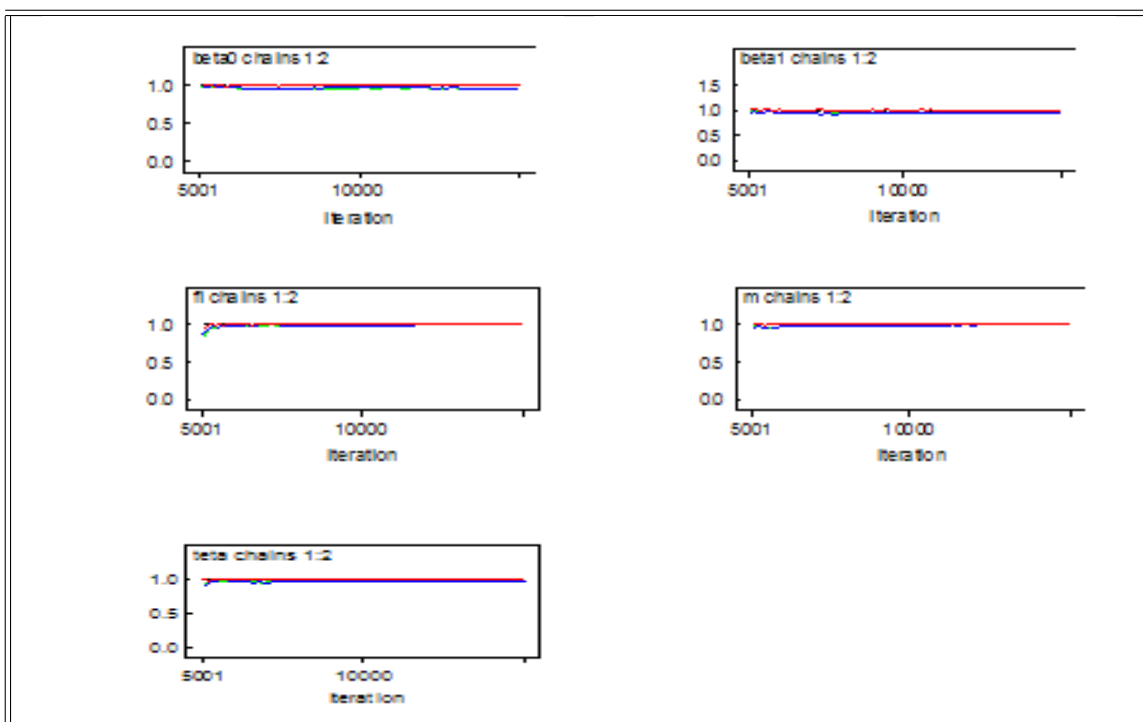
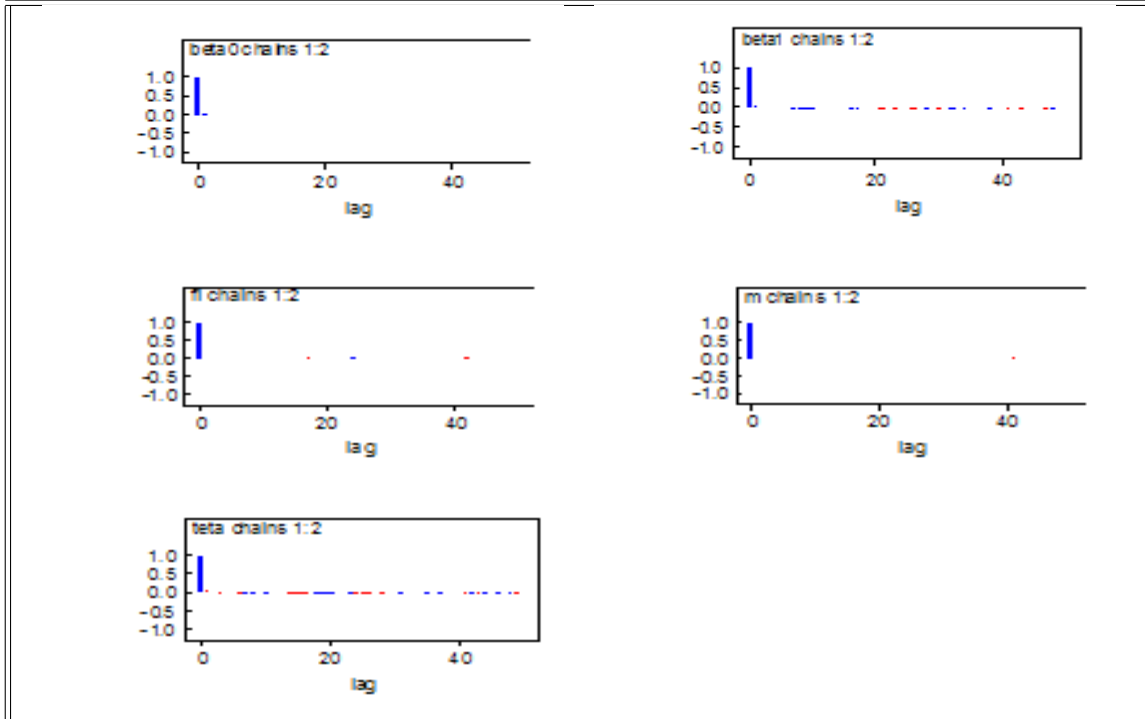


Figura 6.5 - Testes Gráficos de Gelman&Rubin

Os valores obtidos na Figura 6.5, pelo diagnóstico de Gelman&Rubin (1992), são muito próximos de 1, o que indica a convergência das cadeias. Através do teste gráfico da Figura 6.5 também verifica-se a convergência para todas as cadeias.



**Figura 6.6** - Auto Correlação

Observa-se da Figura 6.6 que os parâmetros apresentam baixa auto-correlação, considerando um salto igual a 10.

## 2- Covariável Distância Percorrida

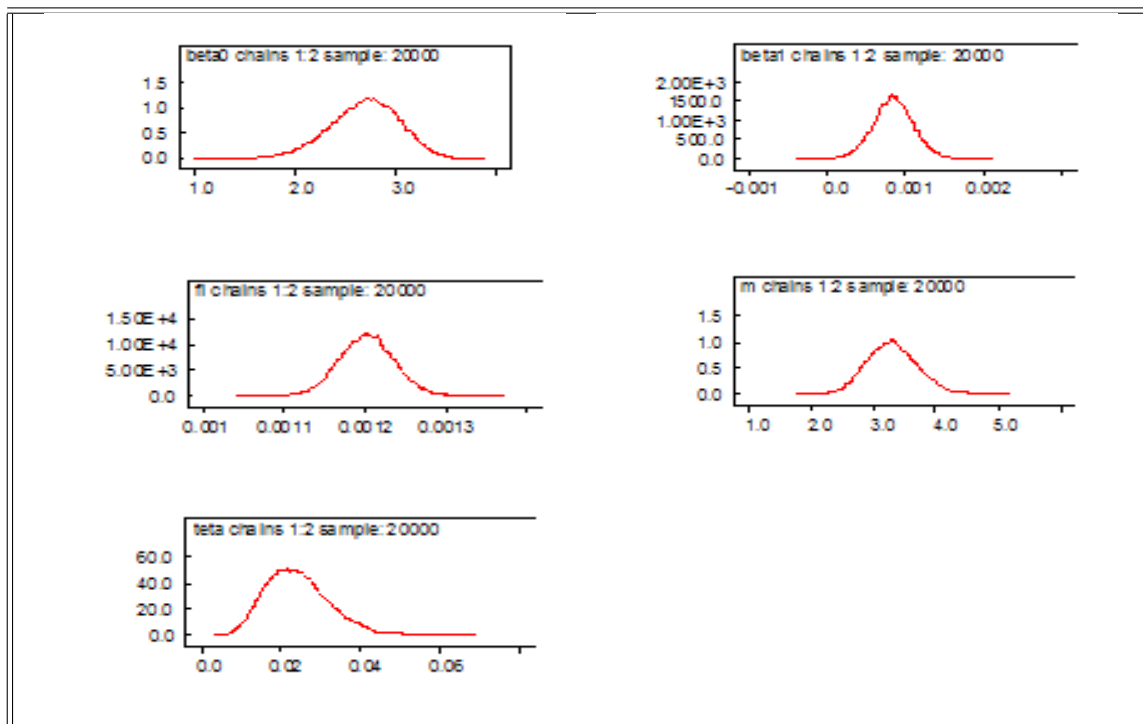
**Tabela 6.3** -Valores dos parâmetros estimados *a posteriori*

Parâmetros	real	média	DP	Região de Credibilidade
$\beta_0$	2,5000	2,6960	0,3479	(1,9570; 3,3170)
$\beta_1$	0,0010	0,0008	0,0002	(0,0001; 0,0013)
$\varphi$	0,00125	0,00120	0,00003	(0,0011; 0,0012)
$m$	3,0000	3,2980	0,4061	(2,5540; 4,1430)
$\theta$	0,0200	0,0242	0,0080	(0,0113; 0,0422)
$\alpha$	16,1669	20,2100	6,6890	(9,3090; 35,1800)

**Tabela 6.4** -Custo médio individual de sinistros

Fórmula	Real	Média	DP	Região de Credibilidade
$\frac{\sum_{j=1}^G \sum_{i=1}^{n_j} \alpha_{ij}}{\sum_{j=1}^G n_j \theta}$	831, 8911	832, 8000	24, 0900	(786, 6000; 881, 5000)

Através dos resultados acima, verificamos que as estimativas obtidas para os parâmetros (Tabela 6.3) são satisfatórias, estão próximas do valor real, e que o custo médio de sinistros estimado também está próximo de seu valor real (Tabela 6.4)

**Figura 6.7** - Densidade a Posteriori

As densidades dos parâmetros *a posteriori* são observadas na Figura 6.7

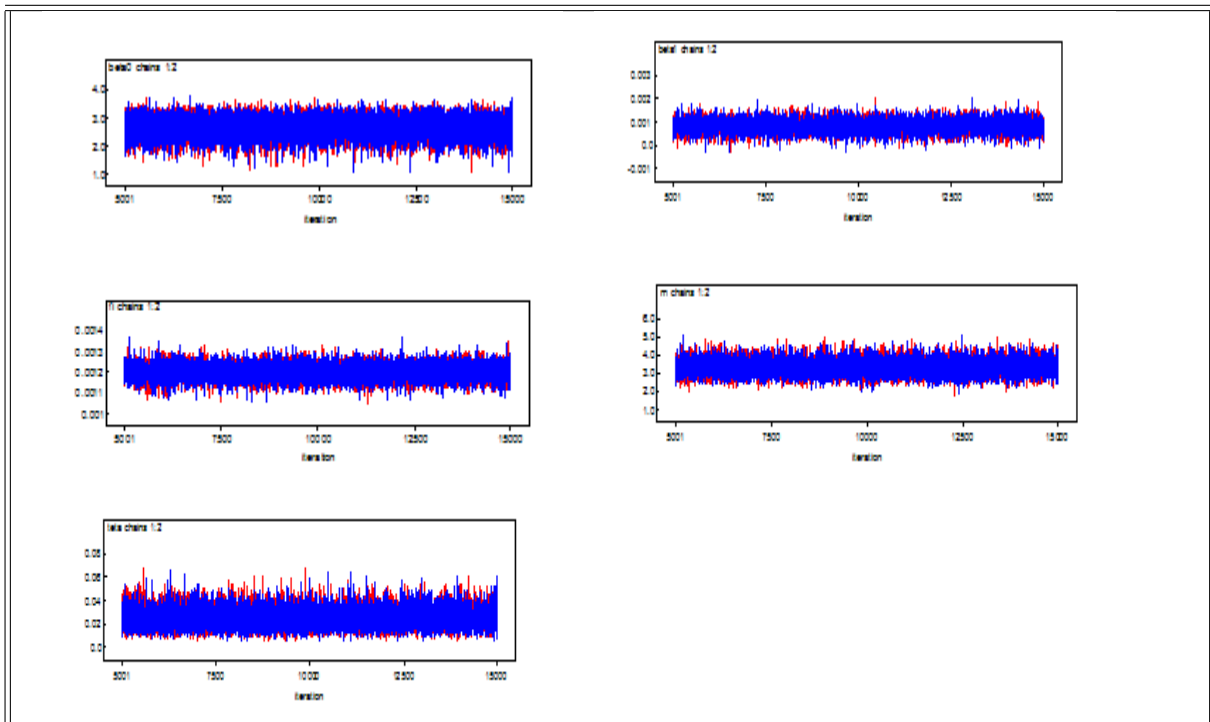


Figura 6.8 - Traços a Posteriori

Os gráficos apresentados na Figura 6.8 indicam convergência para as cadeias pois existe uniformidade no decorrer das iterações.

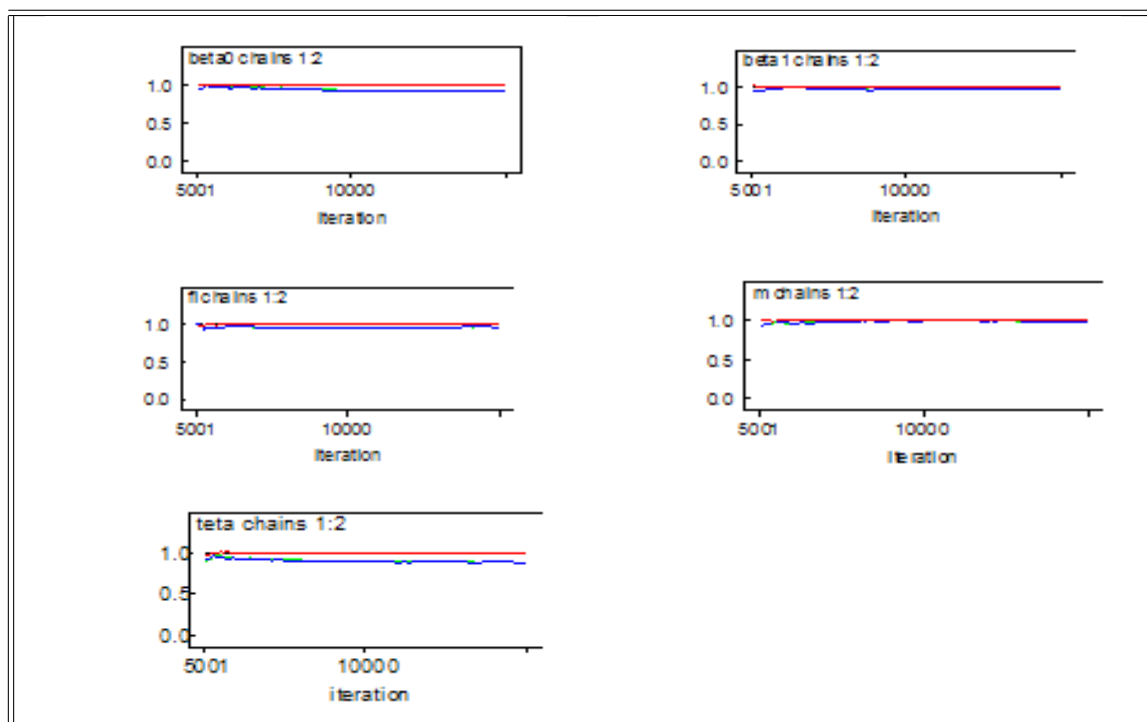
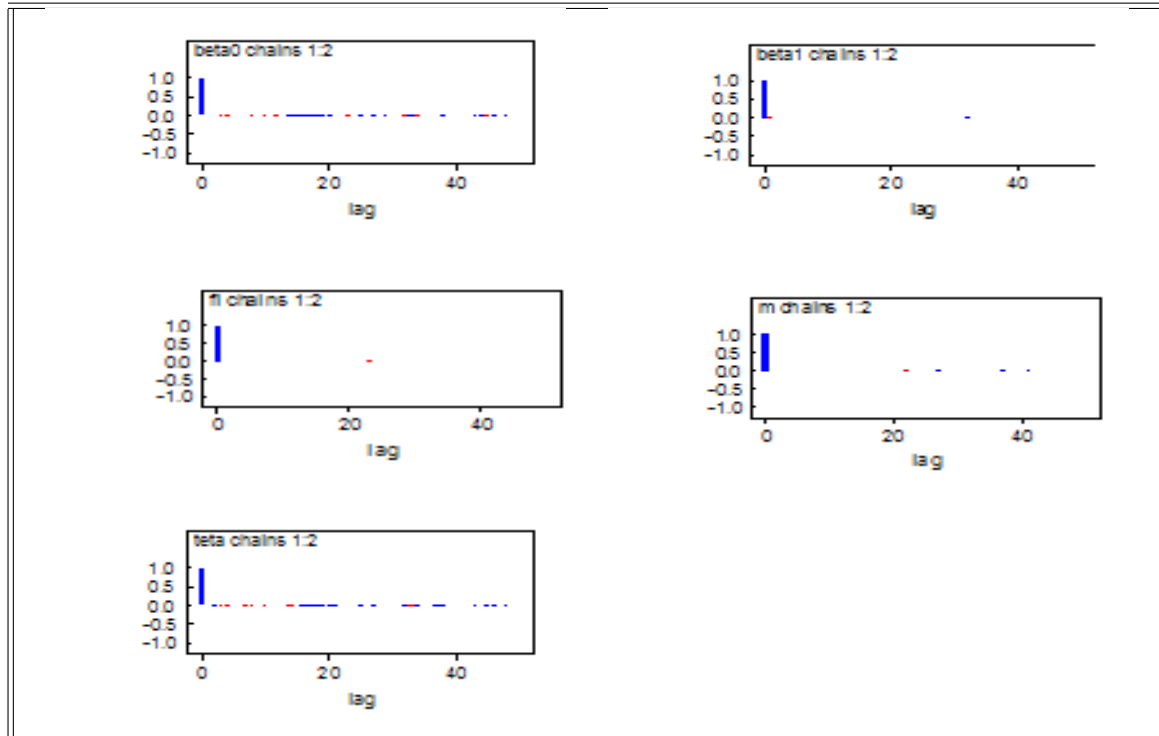


Figura 6.9 - Testes Gráficos de Gelman&Rubin

Através do teste gráfico, Figura 6.9, verificamos a convergência para todas as cadeias, esta convergência é confirmada pelos valores do diagnóstico de Gelman&Rubin (1992) que são muito próximos de 1.



**Figura 6.10** - Auto Correlação

Na Figura 6.10 observamos que os parâmetros apresentam baixa auto-correlação, considerando um salto igual a 10.

### 6.4.3 Resultados referentes ao modelo na presença de duas covariáveis

Nessa subsecção serão apresentados os resultados obtidos para o modelo quando são consideradas ao mesmo tempo as covariáveis idade e distância percorrida.

**Tabela 6.5** - Valores dos parâmetros estimados a *posteriori*

Parâmetros	Real	Média	DP	Região de Credibilidade
$\beta_0$	2,8000	3,0980	0,3639	(2,3200; 3,7460)
$\beta_1$	-0,0100	-0,0111	0,0023	(-0,0159; -0,0065)
$\beta_2$	0,0010	0,0007	0,0002	(0,0003; 0,0012)
$\varphi$	0,00125	0,00120	0,00003	(0,0011; 0,0013)
$m$	3,0000	3,2970	0,4058	(2,5540; 4,1330)
$\theta$	0,0200	0,0255	0,00879	(0,01137; 0,0454)
$\alpha$	15.95364	19,5800	7,0280	(9,1750; 36,560)

**Tabela 6.6** - Custo médio individual de sinistros

Fórmula	Real	Média	DP	Região de Credibilidade
$\frac{\sum_{j=1}^G \sum_{i=1}^{n_j} \alpha_{ij}}{\sum_{j=1}^G n_j \theta}$	804,9524	805,200	23,4800	(760,3000; 852,9000)

As estimativas obtidas para os parâmetros a *posteriori*, (Tabela 6.5), são satisfatórias pois os valores dos parâmetros estão muito próximos dos valores reais utilizados. O custo médio de sinistros estimado, (Tabela 6.6), também é muito satisfatório.

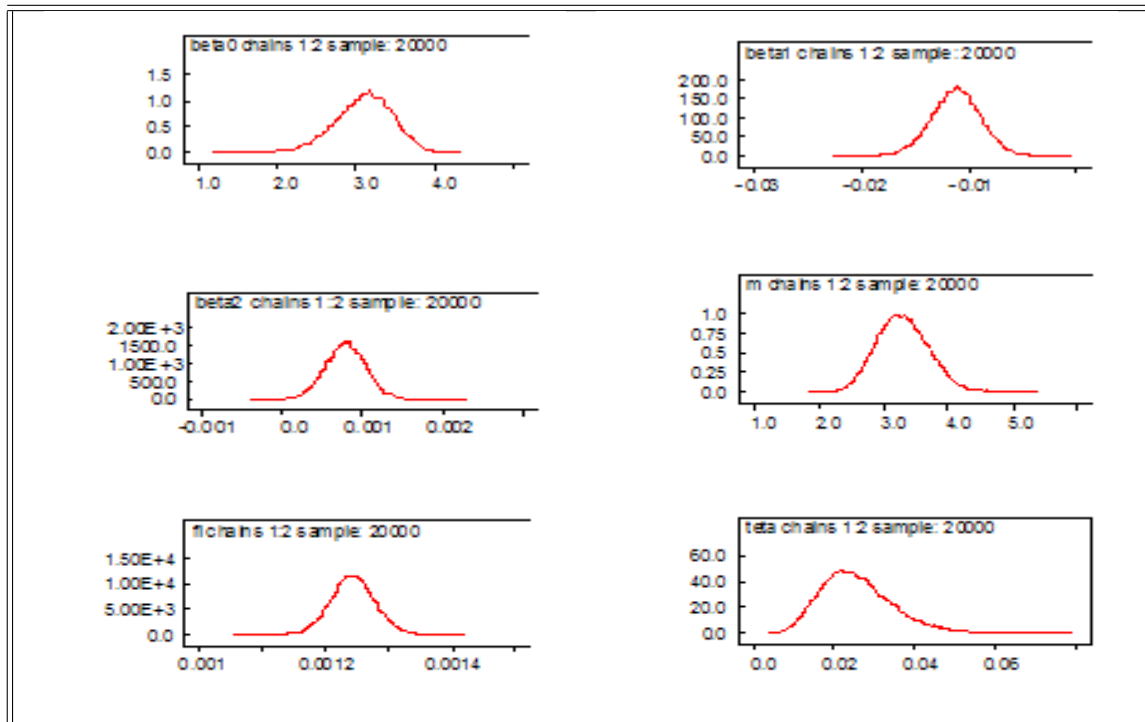


Figura 6.11- Densidade a Posteriori

A densidade dos parâmetros  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ ,  $m$ ,  $\varphi$  e  $\theta$  são dadas na Figura 6.11.

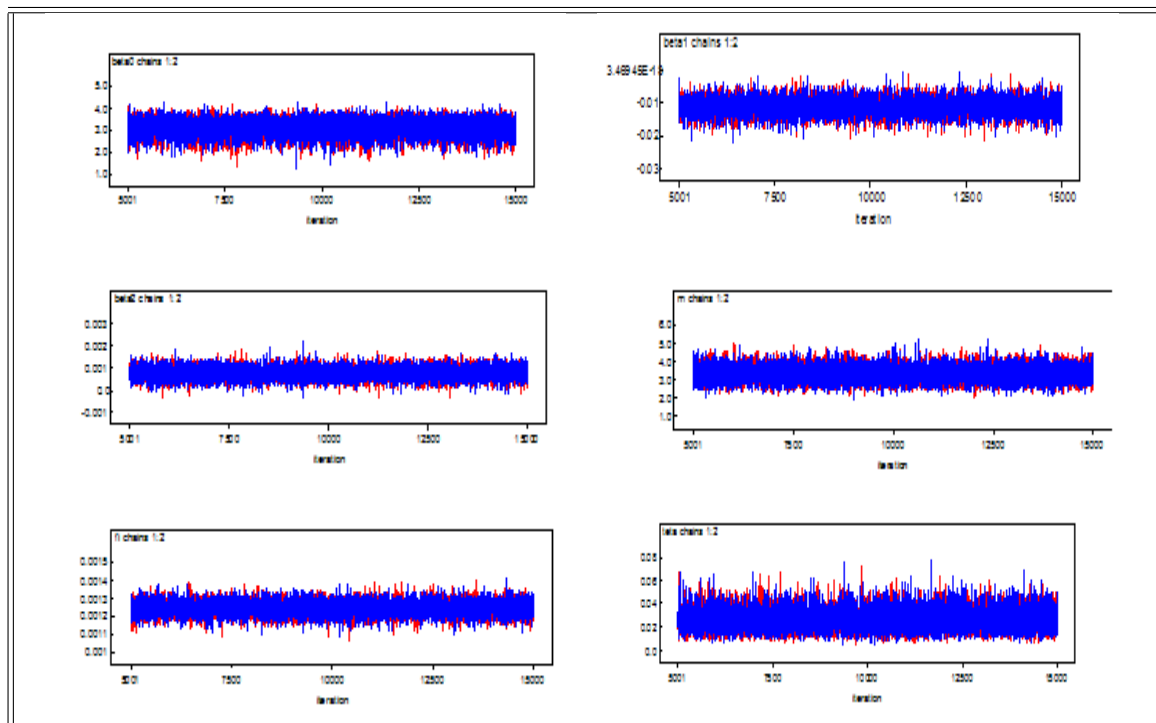
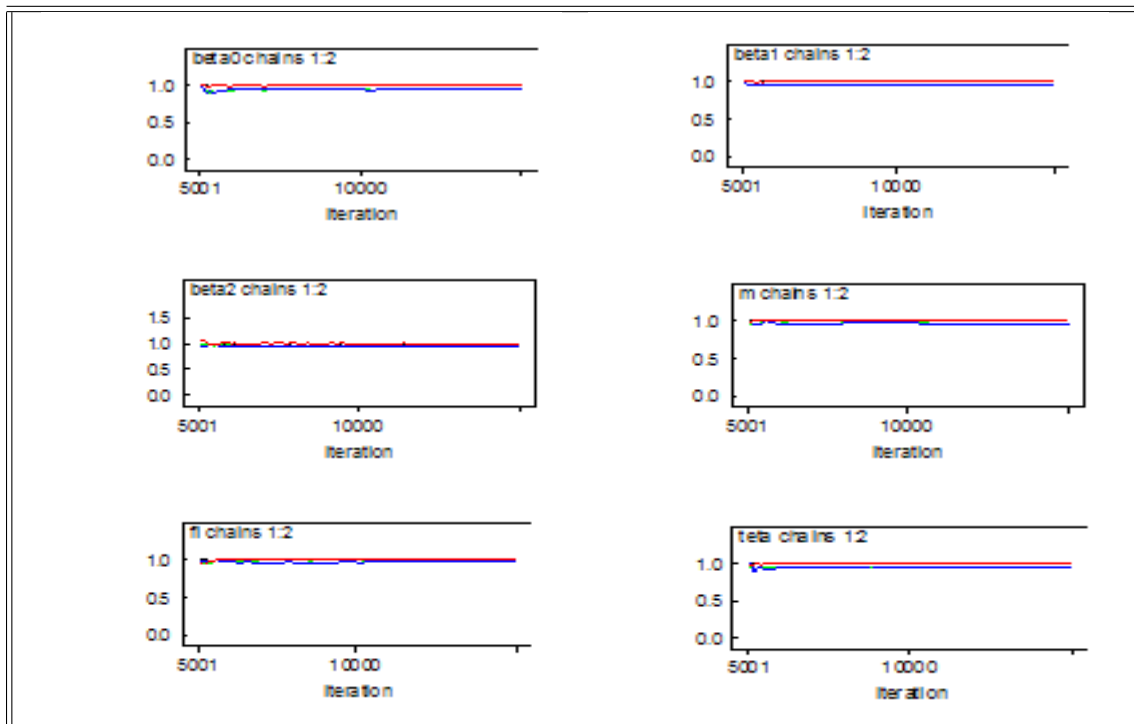


Figura 6.12 - Traços a Posteriori

Os gráficos apresentados na Figura 6.12 indicam convergência para as cadeias pois existe uniformidade nos traços ao longo das iterações.



**Figura 6.13** - Testes Gráficos de Gelman&Rubin

Através dos diagnósticos de Gelman&Rubin (1992) verificamos a convergência das cadeias, pois são muito próximos de 1, esta convergência também é verificada nos testes gráficos apresentados na Figura 6.13.



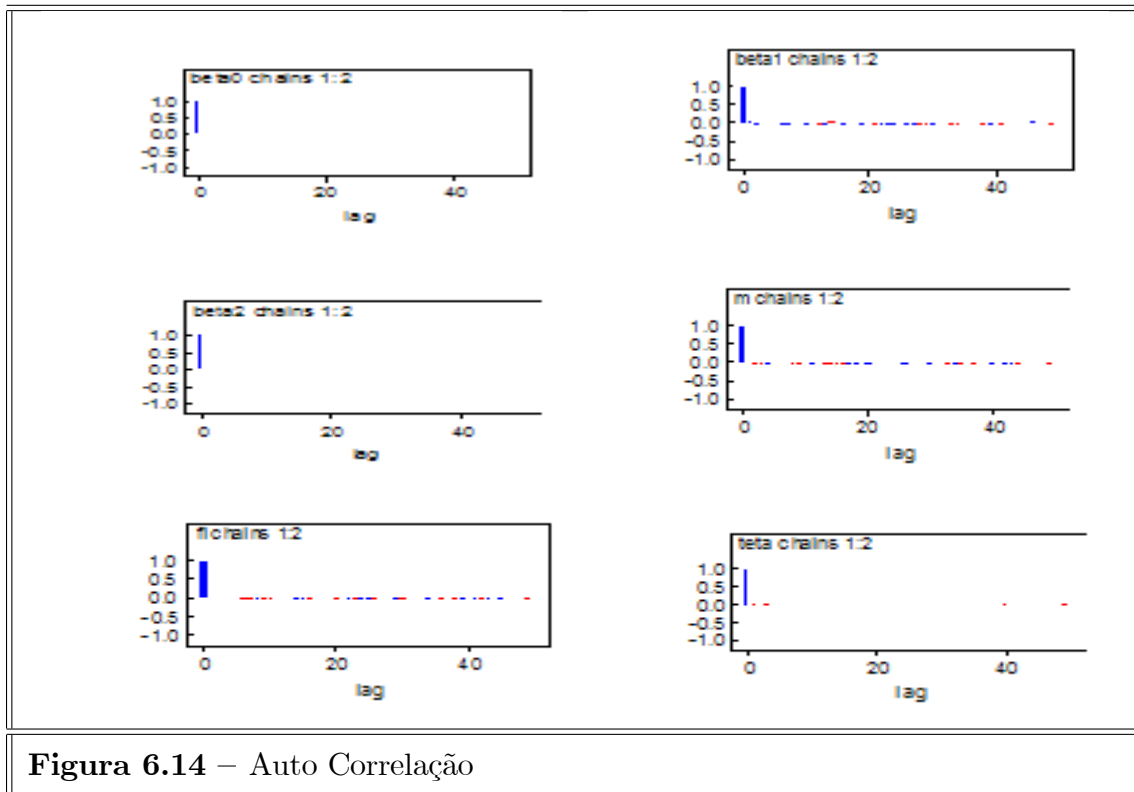


Figura 6.14 – Auto Correlação

Na Figura 6.14 verifica-se que os parâmetros apresentam baixa auto-correlação, considerando um salto igual a 10.

# Capítulo 7

## Comparação entre estimadores Clássicos e Bayesianos

Considerando o Modelo de Tweedie Poisson Composto (Jorgensen & Souza, 1992), é possível avaliar as precisões dos estimadores clássicos e Bayesianos sobre o custo médio individual de sinistros. Neste capítulo utilizaremos a reparametrização apresentada no Capítulo 5 para compararmos, via simulação, os estimadores clássicos e Bayesianos para o modelo de Tweedie sem considerar a presença de covariáveis.

### 7.1 Desempenho dos estimadores clássicos e Bayesianos

Para avaliar o desempenho dos estimadores clássicos e Bayesianos, considerando que a reparametrização apresentada no Capítulo 5 proporcionou uma redução acentuada de correlação entre os parâmetros, utilizaremos a função de verosimilhança definida em (5.2)

Através de técnicas de reamostragem verificamos o desempenho dos estimadores clássicos e Bayesianos supondo  $N_j \sim Po(m)$ , onde  $N_j$  representa o número de sinistros por grupo, e  $T_j|N_j = n_j \sim Ga(\alpha n_j, \theta)$ , onde  $T_j$  representa o custo total de sinistros por grupo.

Para todas as 300 replicações da simulação são definidos  $\alpha = 16$  e  $\theta = 0,02$ . Os valores de  $m$  são gerados a partir da distribuição  $Po(3)$  e  $Po(50)$ , o que correspondem a grupos com pequenas taxas de sinistros e grupos com altas taxas de sinistros, respectivamente. Três tamanhos amostrais são considerados:  $n = 10$ ,  $n = 20$  e  $n = 50$ . As métricas usadas para comparação são o erro quadrático médio (EQM) e o vício.

Os softwares R, SAS e Winbugs (Spiegelhater *et al*, 1996) foram utilizados na implementação das simulações e cálculo das estimativas clássicas e Bayesianas. As estimativas Bayesianas são dadas pelas médias das cadeias geradas a partir das distribuições a *posteriori* de cada parâmetro do modelo.

Para estimar os parâmetros  $\alpha$ ,  $\theta$  e  $m$  consideraremos como priori não informativa a distribuição  $G(1, 0.1)$ , que produz média 10 e variância 100. Lembramos que os valores reais dos parâmetros são  $\alpha = 16$ ,  $\theta = 0,02$  e  $m = 3$  ou 50, desta forma a priori  $G(1, 0.1)$  é extremamente não informativa para o conjunto de parâmetros

## 7.2 Resultados Obtidos

Nas Tabelas abaixo apresentamos os EQM 's, vícios e as médias das estimativas, além do custo médio individual de sinistros determinado pela razão das estimativas médias de  $\alpha$  e  $\theta$ . Os valores presentes na coluna Média -Bayesiana são obtidos considerando as seguintes fases: i) em cada execução MCMC calculamos as médias dos valores obtidos nas cadeias, para cada parâmetro; ii) após 300 execuções do processo MCMC, calculamos as médias destas médias. Os valores presentes na coluna Média-Clássica são obtidos considerando as médias das estimativas obtidas em 300 execuções Newton-Raphson para  $\alpha$  e  $\theta$ . e 300 estimativas de máxima verossimilhança para  $m$ .

Duas situações foram consideradas na análise: a primeira com pequena taxa  $m$  de ocorrência de sinistros em cada grupo e a segunda com uma alta taxa  $m$ .

### 7.2.1 Baixa Taxa de Ocorrência de Sinistros-Poisson (3)

**Tabela 7.1** - Estimativas de  $\alpha$ . (Real = 16).

n	10	20	50
Média-Clássica	21,6116	18,3385	16,5271
Média -Bayesiana	15,8012	16,1421	16,1758
EQM-Clássico	190,0936	54,5951	14,6380
EQM-Bayesiano	32,0082	24,1076	9,9872
Vício-Clássico	5,6210	2,3425	0,5279
Vício-Bayesiano	0,1990	0,1423	0,1761

Observando os resultados da Tabela 7.1, verifica-se que as estimativas médias estimadas via métodos Bayesianos para o parâmetro  $\alpha$  estão mais próximas do valor real em comparação com as estimativas clássicas, para todos os tamanhos amostrais e o EQM para o método Bayesiano é bem menor que o EQM para o método clássico em amostras pequenas e moderadas e os EQM's são similares para grandes amostras. Os vícios associados ao método Bayesiano são, em todo os casos, menores que os do método clássico.

**Tabela 7.2** - Estimativas de  $\theta$ . (Real = 0.02).

n	10	20	50
Média-Clássica	0,02719	0,02302	0,02065
Média -Bayesiana	0,01990	0,02020	0,02020
EQM-Clássico	0,00030	0,00008	0,00002
EQM-Bayesiano	0,00005	0,00003	0,00001
Vício-Clássico	0,00720	0,00303	0,00065
Vício-Bayesiano	0,00007	0,00026	0,00021

Pelos resultados apresentados na Tabela 7.2 observa-se que as estimativas clássicas e Bayesianas para  $\theta$  são similares, mas, por outro lado, o EQM para o método Bayesiano é menor, em amostras pequenas, que o EQM para o método clássico. Em amostras

moderadas e grandes os EQM's para os dois métodos são similares. Novamente, os vícios associados ao método Bayesiano são menores que os do método clássico em todo os casos.

**Tabela 7.3** - Estimativas de  $m$ . (Real = 3).

n	10	20	50
Média-Clássica	3,1986	3,1776	3,1862
Média -Bayesiana	3,2467	3,1857	3,1889
EQM-Clássico	0,3099	0,1655	0,0833
EQM-Bayesiano	0,3081	0,1651	0,0879
Vício-Clássico	0,1788	0,1865	0,1760
Vício-Bayesiano	0,2471	0,1860	0,1892

Pelos resultados apresentados na Tabela 7.3 e 7.4 observa-se que os dois métodos são similares em praticamente todas as métricas estudadas, com o vício clássico menor em pequenas amostras que o vício Bayesiano.

**Tabela 7.4** - Estimativas de  $\frac{\alpha}{\theta}$ . (Real = 800)

n	10	20	50
Média-Clássica	799,9536	798,2815	800,4803
Média -Bayesiana	793,0719	798,3056	800,1069
EQM-Clássico	1341,1950	576,1979	255,0164
EQM-Bayesiano	1344,8090	578,7271	254,8089
Vício-Clássico	4,0532	1,7214	0,4810
Vício-Bayesiano	5,8576	1,6972	0,1071

### 7.2.2 Alta Taxa de Ocorrência de Sinistros-Poisson (50)

**Tabela 7.5** - Estimativas de  $\alpha$ . (Real = 16)

n	10	20	50
Média-Clássica	21,3312	18,4978	16,8100
Média -Bayesiana	15,5610	16,1742	16,0248
EQM-Clássico	175,1840	42,3973	11,2618
EQM-Bayesiano	28,7918	18,6156	8,3479
Vício-Clássico	5,3401	2,5020	0,8113
Vício-Bayesiano	0,4396	0,1744	0,0248

Com altas taxas de sinistro observa-se, pelos resultados apresentados nas Tabelas 7.5 a 7.8, que o método Bayesiano é claramente melhor que o método clássico quando analisamos as estimativas e as métricas para os parâmetros em estudo.

**Tabela 7.6** - Estimativas de  $\theta$ . (Real = 0,02)

n	10	20	50
Média-Clássica	0,02666	0,02313	0,02101
Média -Bayesiana	0,01945	0,02023	0,02002
EQM-Clássico	0,00027	0,00006	0,00001
EQM-Bayesiano	0,00004	0,00002	0,00001
Vício-Clássico	0,00667	0,00314	0,00101
Vício-Bayesiano	0,00054	0,00023	0,00002

**Tabela 7.7** - Estimativas de  $m$ . (Real = 50)

n	10	20	50
Média-Clássica	50,1007	50,0318	50,0629
Média -Bayesiana	49,7045	49,8315	49,9815
EQM-Clássico	2231,2080	2221,6480	2223,2690
EQM-Bayesiano	2193,8110	2202,7660	2215,5890
Vício-Clássico	47,1796	47,1106	47,1417
Vício-Bayesiano	46,7828	46,9108	46,0603

**Tabela 7.8** - Estimativas de  $\frac{\alpha}{\theta}$ . (Real = 800)

n	10	20	50
Média-Clássica	799,6148	799,8209	800,0827
Média -Bayesiana	799,7277	799,7712	800,0730
EQM-Clássico	92,4892	38,3481	14,8356
EQM-Bayesiano	92,3232	38,39463	14,6688
Vício-Clássico	0,3857	0,1794	0,08281
Vício-Bayesiano	0,4868	0,2291	0,0732

### 7.3 Conclusão

Comparando as estimativas para os parâmetros  $\alpha$  e  $\theta$ , verifica-se que o método Bayesiano é claramente melhor que o método clássico levando em conta as métricas estudadas. O método bayesiano é ainda melhor quando ocorreu aumento da taxa de sinistro. Para o parâmetro  $m$  os dois métodos são similares em praticamente todas as métricas estudadas.

# Capítulo 8

## Conclusões e Propostas Futuras

Nessa dissertação desenvolvemos uma abordagem clássica e Bayesiana para o modelo de Tweedie Poisson Composto, apresentado por Jorgensen & Souza (1992), com foco principal na estimação do custo médio individual de sinistros. Com esta estimativa é possível a empresa determinar o valor do prêmio dentro do segmento de grupo estudado.

Ambas abordagens, clássica e Bayesiana, trouxeram estimativas satisfatórias para os parâmetros de interesse  $\alpha$ ,  $\theta$  e  $m$  e também para o custo médio individual de sinistros ( $\alpha/\theta$ ). O método Bayesiano apresenta resultados melhores considerando as métricas de comparação.

No enfoque Bayesiano foi utilizada uma reparametrização para os parâmetros  $\alpha$  e  $\theta$  com o objetivo de minimizar a presença de correlação existente entre os parâmetros e para tentar diminuir a autocorrelação. A reparametrização mostrou-se extremamente eficiente na solução destes dois problemas.

Outro ponto central da dissertação foi o desenvolvimento de inferência Bayesiana para o modelo de Tweedie na presença de covariáveis. Os resultados apresentados mostraram satisfatórios e mais precisos que os casos sem a presença de covariáveis.

Em todos os casos estudados foram utilizados *prioris* não informativas sobre os parâmetros  $\alpha$ ,  $\varphi$  e  $m$  para os modelos sem covariáveis e para os parâmetros  $\alpha$ ,  $\varphi$  e  $m$ ,  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  para os modelos com covariáveis. Os diagnósticos apresentaram, em todos os cenários, convergências e baixa auto-correlação.

Como continuação desse trabalho, propõe-se a utilização de *prioris* informativas, a utilização de outras distribuições para representar o custo individual de sinistros e, se



possível, a ilustração da metodologia desenvolvida em conjuntos de dados reais.

# Referências Bibliográficas

- [1] ALMER, B. Risk analysis in theory and practical statistical. *15 th International Congress of Actuaries*, 2, 314, 1957.
- [2] BAYLEY, R. A.; SIMON, L. Two studies in automobile insurance ratemaking. *ASTIN Bulletin*, 1, 192, 1960.
- [3] BAXTER, L. A.; COUTTS, S. M. ; ROSS, G. A. F. Applications of linear models in motor insurance. *21st International Congress of Actuarial*, 2, 11, 1979.
- [4] BENNET, M. C. Models in motor insurance, *Journal Student Society*, 22, 134, 1978.
- [5] CANAL, G. Y. Estimaco na famlia tweedie e uma aplicaco  seguros de automveis, *Informes de Matemtica*, IMPA, Srie D-047/92, 1992.
- [6] COUTTS, S. M. Motor insurance rating: an Actuarial Approach. *Journal of the Institute of Actuaries*, 111, 87, 1984.
- [7] FERREIRA, P. P. Modelos de precificaco e runas para seguros de curto prazo, *Fundaco Escola. Nacional de Seguros*, 2002.
- [8] GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences, *Statistical Sciences*, 7, 457-472, 1992.
- [9] JOHNSON, P. D.; HEY, G. B. Statistical studies in motor insurance. *Journal of the Institute of Actuaries*, 97, 199, 1971.
- [10] JORGENSEN, B. *The theory of dispersion models*. Londres: Chapman & Hall, p.273-352, 1997.

- [11] JORGENSEN, B.; SOUZA, M. C. P. Fitting tweedies's compound poisson. model to insurance claims data. *Informes de Matemática*, IMPA, Série B-072/92, 1992.
- [12] JUNG, J. On automobile insurance ratemaking. *ASTIN Bulletin*, 5, 41, 1968.
- [13] LUNN, D. J.; THOMAS, A.; BEST, N.; SPIEGELHALTER, D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10: 325-337, 2000.
- [14] SEBER, G. A. F.; WILD, C. J. *Nonlinear regression*. Wiley Series in Probability & Mathematical Statistics, 1989.
- [15] SAS Copyright by SAS Institute INC., Cary, NC, USA, 2002.
- [16] TWEEDIE, M, C, K. An index which distinguishes between some important exponential families in Statistics: application and new directions. *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*. J. K. Ghosh and J. Roy, Calcutta: Indian Statistical Institute, 579-604, 1984.

# Apêndice A

## Demonstrações

As proposições e corolários a seguir foram propostas e demonstradas em Canal, G. Y. (1992).

**Proposição A.1:** Propriedade de Convolução das Famílias  $ED^*(\theta, \lambda)$ . Sejam  $Z_1, \dots, Z_n$  variáveis aleatórias independentes com distribuições

$$Z_i \sim ED^*(\theta, \lambda_i),$$

para  $i = 1, \dots, n$ .

Se  $Z_+ = Z_1 + \dots + Z_n$  então  $Z_+ \sim ED^*(\theta, \lambda_1 + \dots + \lambda_n)$ .

**Demonstração:** Lembrando que a função geradora de momentos da soma de variáveis aleatórias independentes é o produto delas, obtém-se

$$G_{Z_+}(t; \theta) = \exp \left[ \sum_{i=1}^n \lambda_i \{k(\theta + t) - k(\theta)\} \right], t \in \Theta - \theta, t \in \Theta - \theta,$$

que é a função geradora de momentos de uma variável  $ED^*(\theta, \lambda_1 + \dots + \lambda_n)$ . Portanto,  $Z_+ \sim ED^*(\theta, \lambda_1 + \dots + \lambda_n)$ .

Como

$$Y \sim ED(\mu, \sigma^2) \Leftrightarrow \lambda Y \sim ED^*(\theta, \lambda), \sigma^2 = \frac{1}{\lambda},$$

temos o seguinte corolário:

**Corolário A.1:** Se  $Y_i, \dots, Y_n$  são variáveis aleatórias independentes e

$$Y_i \sim ED(\mu, \sigma^2/w_i), i = 1, \dots, n,$$

com pesos  $w_i$ , tais que  $w_i/\sigma^2 \in \Lambda$ , então a forma reprodutiva da fórmula de convolução é dada por

$$\frac{1}{w_+} \sum_{i=1}^n w_i Y_i \sim ED(\mu, \sigma^2/w_+),$$

onde  $w_+ = w_1 + \dots + w_n$ .

**Demonstração:** (diagramática)

$$\begin{aligned} Y_i &\sim ED(\mu, \sigma^2/w_i) \implies^{(2.2)} \sigma^{-2} w_i Y_i \sim ED^*(\theta, \sigma^{-2} w_i) \\ &\Downarrow \text{Pr oposição 2.2.1} \\ \frac{1}{w_+} \sum_{i=1}^n w_i Y_i &\sim ED(\mu, \sigma^2/w_+) \iff^{(2.2)} \sigma^{-2} \sum_{i=1}^n w_i Y_i \sim ED^*(\theta, \sigma^{-2} w_+). \end{aligned}$$

**Proposição A.2:**

A variável aleatória  $Z$  dada em (3.1) com  $Z_i \sim Ga(-\theta, -\alpha)$  é uma distribuição  $ED^*(\theta, \lambda)$  com função densidade dada por

$$\begin{aligned} f_Z(z; \theta, \lambda, \alpha) &= \sum_{n=1}^{\infty} \frac{\{\lambda k_p(-1/z)\}^n}{\Gamma(-n\alpha) n! z} \exp\{\theta z - \lambda k_p(\theta)\} \\ P(Z=0) &= \exp\{-\lambda k_p(\theta)\} \end{aligned}$$

onde

$$\begin{aligned} K_p(\theta) &= \frac{\alpha - 1}{\alpha} \left( \frac{\theta}{\alpha - 1} \right)^\alpha, \\ V(\mu) &= \mu^p \quad \text{com } p \in (1, 2). \end{aligned}$$

**Demonstração:**

Dado  $Z_i \sim Ga(-\theta, -\alpha)$  ou seja, sua função densidade é da forma

$$\frac{(-\theta)^{-\alpha}}{\Gamma(-\alpha)} z^{-\alpha-1} e^{\theta z},$$

Simplificando a notação, temos que

$$f_Z(z) = \sum_{n=1}^{\infty} f_{Z,N}(z, n)$$

$$f_{Z,N}(z, n) = f_{Z/N}(z) \cdot f_N(n).$$

Pela propriedade de convolução, proposição A.1, temos que

$$Z|N = n \sim Ga(-\theta, -n\alpha).$$

e portanto

$$f_{Z,N}(z, n; \theta, \lambda, \alpha) = \frac{(-\theta)^{-\alpha}}{\Gamma(-n\alpha)} z^{-\alpha-1} e^{\theta z} \frac{m^n}{n!} e^{-m},$$

para  $z > 0, n \geq 1$ .

Por conveniência se introduz o parâmetro  $\lambda$ , definido através de  $m$ , na seguinte maneira (lembre-se de 3.1.3)

$$m = \lambda K_p(\theta),$$

onde

$$K_p(\theta) = \frac{\alpha - 1}{\alpha} \left( \frac{\theta}{\alpha - 1} \right)^\alpha,$$

para obter

$$f_{Z,N}(z, n; \theta, \lambda, \alpha) = \frac{\{\lambda k_p(-1/z)\}^n}{\Gamma(-n\alpha) n! z} \exp\{\theta z - \lambda k_p(\theta)\}.$$

e

$$\begin{aligned}P(N = 0) &= P(Z = 0) = \exp\{-\lambda k_p(\theta)\}, \\K_p''(\theta) &= \left(\frac{\theta}{\alpha - 1}\right)^{\alpha-2}, \\V(\mu) &= k''(\tau^{-1}(\mu)) = \mu^{\frac{\alpha-2}{\alpha-1}} = \mu^p,\end{aligned}$$

onde  $p = (\alpha - 2) / (\alpha - 1)$ . Como  $\alpha < 0$  temos que  $p \in (1, 2)$ .

# Apêndice B

## Programas

Seguem as linhas de instruções em SAS, para os procedimentos de estimativas intervalares empíricas, utilizadas nas Inferências Clássicas.

### Código 1

Código do programa utilizado no *software* SAS afim de obtermos a estimativa dos parâmetros de interesse a partir da função de verossimilhança definida em (4.4).

```
options nonotes;
data z;
infile "c:\flavia\zc_dez.txt";
input z;
run;
data mn;
infile "c:\flavia\nc_dez.txt";
input mn;
run;
data dados;
set z;
set mn;
run;
%let iter=300; /* número de iterações*/
/* separando os dados*/
```



```

%macro dados;
%do i=1 %to &iter;
%put "Separando Matriz &i";
proc iml;
use dados var _all_;
read all into dados;
dados&i= shape(0,20,2);
dados&i= dados[20*&i-19:20*&i,];
colnames={z&i nn&i};
create dados&i from dados&i[c=colnames];
append from dados&i;
close dados&i;
quit;
%put "Estimando da Matriz &i";
proc nlp data = dados&i tech = tr phes cov = 2 vardef=n;
max l;
parms teta&i = 0.04,
q&i = 12;
a&i = sum(nn&i);
d&i = gamma(nn&i+1);
m&i = sum(nn&i)/20;
L = q&i*a&i*(log(teta&i)) - sum(log(gamma(q&i*nn&i)))+sum((q&i*nn&i-1)*log(z&i))-
teta&i*sum(z&i)+(sum(nn&i))*log(m&i)-m&i*20+sum(log(d&i));
run;
%end;
%mend dados;
%dados;

```

## Código 2

Código do programa no software SAS utilizado na implementação das simulações e cálculos das estimativas clássicas demonstradas no capítulo 7.

O código abaixo é referente quando utilizamos uma amostra de tamanho 20 e distribuição Poisson com parâmetro 50. Para os demais casos, os programas são análogos a este.

```
options nonotes;
data z;
infile "c:\poisson50_20janeiro\custo20.txt";
input z;
run;
data mn;
infile "c:\poisson50_20janeiro\sinistro20.txt";
input mn;
run;
data dados;
set z;
set mn;
run;
%let iter=300; /* número de iterações*/
/* separando os dados*/
%macro dados;
%do i=1 %to &iter;
%put "Separando Matriz &i";
proc iml;
use dados var _all_;
read all into dados;
dados&i= shape(0,20,2);
dados&i= dados[20*&i-19:20*&i,];
colnames={z&i mn&i};
create dados&i from dados&i[c=colnames];
append from dados&i;
close dados&i;
m&i= shape(0,300,1);
```

```

m&i = dados&i[,2][+]/20;
create m&i from m&i;
append from m&i;
close m&i;
quit;
%put "Estimando da Matriz &i";
proc nlp data = dados&i tech = tr phes cov = 2 vardef=n;
max l;
parms teta&i = 0.01,
q&i = 12;
a&i = sum(nn&i);
d&i = lgamma(nn&i+1);
m&i = sum(nn&i)/20;
L = q&i*a&i*(log(teta&i)) - exp(sum(log(lgamma(q&i*nn&i))))+sum((q&i*nn&i-1)*log(z&i))-
teta&i*sum(z&i)+(sum(nn&i))*log(m&i)-m&i*20-exp(sum(log(d&i)));
run;
%end;
%mend dados;
%dados;

```

Seguem as linhas de instruções em WinBUGS para o procedimentos MCMC usados nas análises Bayesianas apresentadas nesta dissertação.

### Código 3

Código do programa no *software* WinBUGS, a fim de obtermos amostras da distribuição condicional dos parâmetros de interesse, sob o modelo definido em (5.2).

```

model;
{
for( i in 1 : N ) {
p[i] <- exp(e[i])
uns[i] ~ dbern(p[i])

```



```
fi ~ dgamma(0.001,0.001)
beta0 ~ dnorm( 0,0.001)
beta1 ~ dnorm( 0,0.001)
for (i in 1:N1) {
  alfa1[i] <-exp(beta0*x0[1]+beta1*x1[i])
}
soma[1]<-sum(alfa1[])
n[1]<-N1
for (i in 1:N2) {
  alfa2[i] <-exp(beta0*x0[2]+beta1*x2[i])
}
soma[2]<-sum(alfa2[])
n[2]<-N2
for (i in 1:N3) {
  alfa3[i] <-exp(beta0*x0[3]+beta1*x3[i])
}
soma[3]<-sum(alfa3[])
n[3]<-N3
for (i in 1:N4) {
  alfa4[i] <-exp(beta0*x0[4]+beta1*x4[i])
}
soma[4]<-sum(alfa4[])
n[4]<-N4
for (i in 1:N5) {
  alfa5[i] <-exp(beta0*x0[5]+beta1*x5[i])
}
soma[5]<-sum(alfa5[])
n[5]<-N5
for (i in 1:N6) {
  alfa6[i] <-exp(beta0*x0[6]+beta1*x6[i])
}
```

```
soma[6]<-sum(alfa6[])
n[6]<-N6
for (i in 1:N7) {
  alfa7[i] <-exp(beta0*x0[7]+beta1*x7[i])
}
soma[7]<-sum(alfa7[])
n[7]<-N7
for (i in 1:N8) {
  alfa8[i] <-exp(beta0*x0[8]+beta1*x8[i])
}
soma[8]<-sum(alfa8[])
n[8]<-N8
for (i in 1:N9) {
  alfa9[i] <-exp(beta0*x0[9]+beta1*x9[i])
}
soma[9]<-sum(alfa9[])
n[9]<-N9
for (i in 1:N10) {
  alfa10[i] <-exp(beta0*x0[10]+beta1*x10[i])
}
soma[10]<-sum(alfa10[])
n[10]<-N10
for (i in 1:N11) {
  alfa11[i] <-exp(beta0*x0[11]+beta1*x11[i])
}
soma[11]<-sum(alfa11[])
n[11]<-N11
for (i in 1:N12) {
  alfa12[i] <-exp(beta0*x0[12]+beta1*x12[i])
}
soma[12]<-sum(alfa12[])
```

```
n[12]<-N12
for (i in 1:N13) {
  alfa13[i] <-exp(beta0*x0[13]+beta1*x13[i])
}
soma[13]<-sum(alfa13[])
n[13]<-N13
for (i in 1:N14) {
  alfa14[i] <-exp(beta0*x0[14]+beta1*x14[i])
}
soma[14]<-sum(alfa14[])
n[14]<-N14
for (i in 1:N15) {
  alfa15[i] <-exp(beta0*x0[15]+beta1*x15[i])
}
soma[15]<-sum(alfa15[])
n[15]<-N15
for (i in 1:N16) {
  alfa16[i] <-exp(beta0*x0[16]+beta1*x16[i])
}
soma[16]<-sum(alfa16[])
n[16]<-N16
for (i in 1:N17) {
  alfa17[i] <-exp(beta0*x0[17]+beta1*x17[i])
}
soma[17]<-sum(alfa17[])
n[17]<-N17
for (i in 1:N18) {
  alfa18[i] <-exp(beta0*x0[18]+beta1*x18[i])
}
soma[18]<-sum(alfa18[])
n[18]<-N18
```

```
for (i in 1:N19) {
  alfa19[i] <-exp(beta0*x0[19]+beta1*x19[i])
}
soma[19]<-sum(alfa19[])
n[19]<-N19
for (i in 1:N20) {
  alfa20[i] <-exp(beta0*x0[20]+beta1*x20[i])
}
soma[20]<-sum(alfa20[])
n[20]<-N20
const<-1
total<-sum(soma[])
alfamedio<-total/66
teta<-fi*alfamedio
customedio<-total/(teta*66)
}
list(N=20,z=c(3023.322,757.1312,3420.356,3150.857,3595.393,2476.457,
3055.705,1958.668,709.7797,975.6207,1179.688,4600.59,3125.756,2334.021,
3054.797,4299.165,5175.727,2561.269,1037.936,1829.836),
x1=c(18,19,18),
x2=c(22),
x3=c(24,21,20),
x4=c(23,22,19),
x5=c(22,24,18),
x6=c(27,26,35),
x7=c(22,19,21),
x8=c(25,31,33),
x9=c(39),
x10=c(48,56),
x11=c(36,45),
x12=c(18,24,23,20),
```



```
x13=c(52,55,47,49,51,53),
x14=c(39,46,37),
x15=c(22,19,24),
x16=c(34,25,32,31,35,30),
x17=c(30,29,28,31,35,34,26),
x18=c(44,40,42,36),
x19=c(69,59),
x20=c(58,60,62,69),
N1=3,
N2=1,N3=3,N4=3,N5=3,N6=3,N7=3,N8=3,N9=1,N10=2,N11=2,N12=4,
N13=6,N14=3,N15=3,N16=6,N17=7,N18=4,N19=2,N20=4,
uns=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),
x0=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1))
list(m=1,beta0=2,beta1= -0.03,fi=0.04)
list(m=5,beta0=6,beta1= -0.04,fi=0.01)
```