

Sumário

Resumo	2
1. Introdução	3
2. Material e Métodos.....	7
2.1. Modelos Dinâmicos.....	12
2.2. Modelos Propostos	14
2.3. Critérios de Comparação	16
2.4. Inferência Bayesiana e MCMC.....	17
2.5. Exemplo de funcionamento do WinBUGS.....	19
2.6. Implementação	23
3. Resultados e Conclusões	26
Referências Bibliográficas.....	29
Anexos.....	31

Resumo

A Inferência Bayesiana possui uma série de aplicações nas mais diversas áreas, seja no mercado agrícola, na implementação de experimentos ou mesmo na área financeira e atuarial. Este estudo trata desta última, em que métodos computacionalmente intensivos são utilizados no desenvolvimento de modelos para constituição de provisões matemáticas.

Neste trabalho, vários modelos Bayesianos serão propostos e comparados em termos de sua capacidade preditiva. Como será visto, fazer inferência para os parâmetros desses modelos não é trivial e por isso, serão utilizados métodos de Monte Carlo via Cadeias de Markov (MCMC) para estimação e comparação de modelos. Será utilizado o software WinBUGS para simulação das cadeias de Markov e por fim, será feita a análise dos resultados, produzindo um comparativo entre as técnicas bayesianas e o método atualmente vigente, a fim de encontrar o melhor preditor para o cálculo da IBNR (Provisão para Eventos Ocorridos mas Não Avisados).

Palavras-chave: MCMC, modelos dinâmicos, WinBUGS, IBNR, PEONA.

Capítulo 1

Introdução

Este estudo tem por objetivo apresentar uma aplicação da Inferência Bayesiana no ramo das Ciências Atuariais. Todo o problema tratado nesta monografia refere-se a encontrar um método eficiente para estimação de uma Provisão Técnica de Eventos Ocorridos mas Não Avisados, utilizada pelas companhias seguradoras e operadoras de planos de saúde complementar.

Luiz Roberto Castiglione define “Provisões Técnicas” como sendo montantes que “têm por finalidade garantir o pagamento de ocorrências futuras, não permitindo que as parcelas destinadas a cobrir os riscos assumidos (prêmio puro) seja consumida pela companhia seguradora antes do (término do) prazo de vigência do negócio”.

Dentre as muitas provisões técnicas existentes, as operadoras de planos de saúde geralmente utilizam a IBNR – *Incurred But Not Reported*, traduzida como PEONA – *Provisão para Eventos Ocorridos mas Não Avisados*. A idéia da IBNR é constituir um montante para que a companhia possa arcar com seus compromissos futuros, especificamente sobre as datas de ocorrência e aviso do sinistro.

Para um melhor entendimento do processo que dará origem à constituição da IBNR / PEONA, pode ser analisado como exemplo o caso das operadoras de planos de saúde. A seguir, é apresentada a cronologia mais comumente usada entre os prestadores e as operadoras:

- **Liberação:** quando o beneficiário do plano é liberado, pela operadora, a efetuar determinado procedimento.
- **Ocorrência:** quando o usuário efetua o procedimento. Geralmente, a data de liberação e de ocorrência acabam coincidindo.
- **Aviso:** quando o prestador apresenta as guias referentes aos procedimentos efetuados para ser ressarcido.
- **Pagamento:** quando a operadora efetivamente efetua o pagamento referente a determinado procedimento.

O que geralmente ocorre é que o prestador, por diversos motivos, demora um certo tempo para comunicar a operadora da ocorrência do sinistro. Assim, temos um intervalo de tempo (meses de atraso) entre a data de ocorrência do procedimento e a data de aviso do sinistro referente àquele procedimento. O esquema geral de como os montantes a serem pagos ficam organizados no tempo pode ser exemplificado com o uso do Triângulo de Run-Off:

Triângulo Superior:

Mês de Origem	Período de Desenvolvimento (Meses de Atraso)				
	1	2	...	k-1	k
1	y_{11}	y_{12}	...	$y_{1,k-1}$	y_{1k}
2	y_{21}	y_{22}	...	$y_{2,k-1}$	
...	...				
k-1	$y_{k-1,1}$	$y_{k-1,2}$			
k	y_{k1}				

Triângulo Inferior:

Mês de Origem	Período de Desenvolvimento (Meses de Atraso)				
	1	2	...	k-1	k
1					
2			...		$y_{2,k}$
3				$y_{3,k-1}$	$y_{3,k}$
...			...		
k-1			...	$y_{k-1,k-1}$	$y_{k-1,k}$
k		$y_{k,2}$...	$y_{k,k-1}$	$y_{k,k}$

O triângulo superior apresenta os dados observados pela operadora, ou seja, os valores totais de sinistros a serem pagos, de acordo com o mês de origem e o número de meses de atraso referente àquele mês. Ao mesmo tempo, o triângulo inferior é o que deve ser estimado a fim de prever as despesas que a operadora terá de arcar no futuro próximo, levando-se em conta a distribuição dos montantes a serem pagos de acordo com os meses de atraso em relação ao mês de origem.

Na prática, observa-se que a grande maioria dos pagamentos efetuados concentra-se na parte esquerda da tabela, ou seja, de 1 a 4 meses de atraso com relação à ocorrência e o aviso, enquanto que os montantes destes valores pagos ocorrem com menor frequência na parte direita da tabela, o que nos leva a crer que, quanto maior é o tempo de atraso, menor é a quantidade de despesas a ser paga. Isso nos leva a pensar em certa dependência dinâmica entre os dados. O que se deseja estimar são as despesas que a operadora terá no futuro (mensalmente, anualmente, etc.), seja com 5 meses de atraso referente a um mês passado, ou com 1 mês de atraso, referente ao mês corrente, e assim por diante.

Várias técnicas para a provisão da IBNR têm sido discutidas. Com o uso do Triângulo de Run-Off, Straub (1980) sugere o método conhecido como “Chain Ladder” (Cadeia Escalonada), cuja metodologia também pode ser verificada em Boulter (2000). O método consiste em encontrar fatores médios entre as colunas e, com esses fatores, preencher o triângulo inferior. Esses fatores são encontrados simplesmente dividindo-se a soma de uma coluna pela soma de sua coluna anterior. Assim, num triângulo com 6 colunas, por exemplo, teríamos 5 fatores. Para preencher o primeiro campo em branco da segunda coluna, bastaria multiplicar o último campo da primeira coluna pelo fator obtido com a divisão da soma da segunda coluna pela primeira. E assim por diante. O método “Cape Code”, desenvolvido pelo próprio Straub, é uma extensão do “Chain Ladder”, mas o grande problema por trás desses métodos não é difícil de ser verificado: o método considera que os valores futuros sempre seguirão uma mesma proporção dos valores anteriores, o que nem sempre é verdade e pode ocasionar muitos problemas.

Mais recentemente, Oliveira Jr. (2004), baseado em um artigo de Ntzoufras e Dellaportas (2002), apresenta uma nova abordagem para esta estimação, através da Inferência Bayesiana. Com o uso do método MCMC (*Monte Carlo Markov Chain*), implementado pelo Algoritmo de Gibbs através do software WinBUGS, ele propõe um

modelo dinâmico para este cálculo, atribuindo uma distribuição Log-Normal para os montantes de sinistros e uma distribuição Multinomial para as quantidades.

Um outro estudo, fornecido pela Unimed Seguros (2004), também propõe uma abordagem Bayesiana para o caso em questão, mas atribuindo uma distribuição Gama para os montantes e uma distribuição de Poisson para as quantidades.

Fonseca *et al* (2006), por outro lado, utiliza a abordagem Bayesiana com um modelo de Poisson generalizado. Nascimento (2006), entretanto, sugere a estimação utilizando técnicas semi-paramétricas.

Estes são apenas alguns exemplos do que vem sendo discutido e das técnicas propostas para o problema em questão. Existem ainda muitas outras técnicas sugeridas, mas ainda não se chegou a uma conclusão eficaz sobre qual seria o melhor método para efetuar o cálculo da provisão. Longe de dar um ponto final nesta linha de pesquisa do ramo das ciências atuariais, este estudo visa propor, a partir desses estudos anteriores, uma nova alternativa Bayesiana para se modelar os montantes de sinistros a serem pagos e, conseqüentemente, chegar a um novo método para constituição da IBNR / PEONA.

Capítulo 2

Material e Métodos

Através de uma breve análise descritiva, pode-se verificar qual é o comportamento dos dados observados em quatro diferentes operadoras de planos de saúde complementar. As informações foram fornecidas pela *Actuarial Assessoria e Consultoria Atuarial Ltda*, uma empresa que presta consultoria a essas operadoras. Foram analisadas quatro empresas diferentes para identificar se os dados comportam-se seguindo algum tipo de padrão ou se eles variam de empresa para empresa. Desta forma, foram construídos os histogramas apresentados nas figuras 2.1 a 2.4.

Pode-se perceber, através dos gráficos apresentados, que parece haver uma “bimodalidade” tanto para os montantes de sinistros quanto para suas respectivas quantidades apresentadas. Convém lembrar que, conforme mencionado no Capítulo 1, os dados também parecem evoluir ao longo do tempo, o que nos leva a pensar em uma proposta de modelos dinâmicos para o problema em questão.

Antes de propor modelos que levem esta natureza bimodal e dinâmica dos dados em conta, um breve apanhado teórico de modelos dinâmicos se faz necessário, sendo dado na próxima seção.

Figura 2.1 – Histogramas para os Dados do Cliente “A”

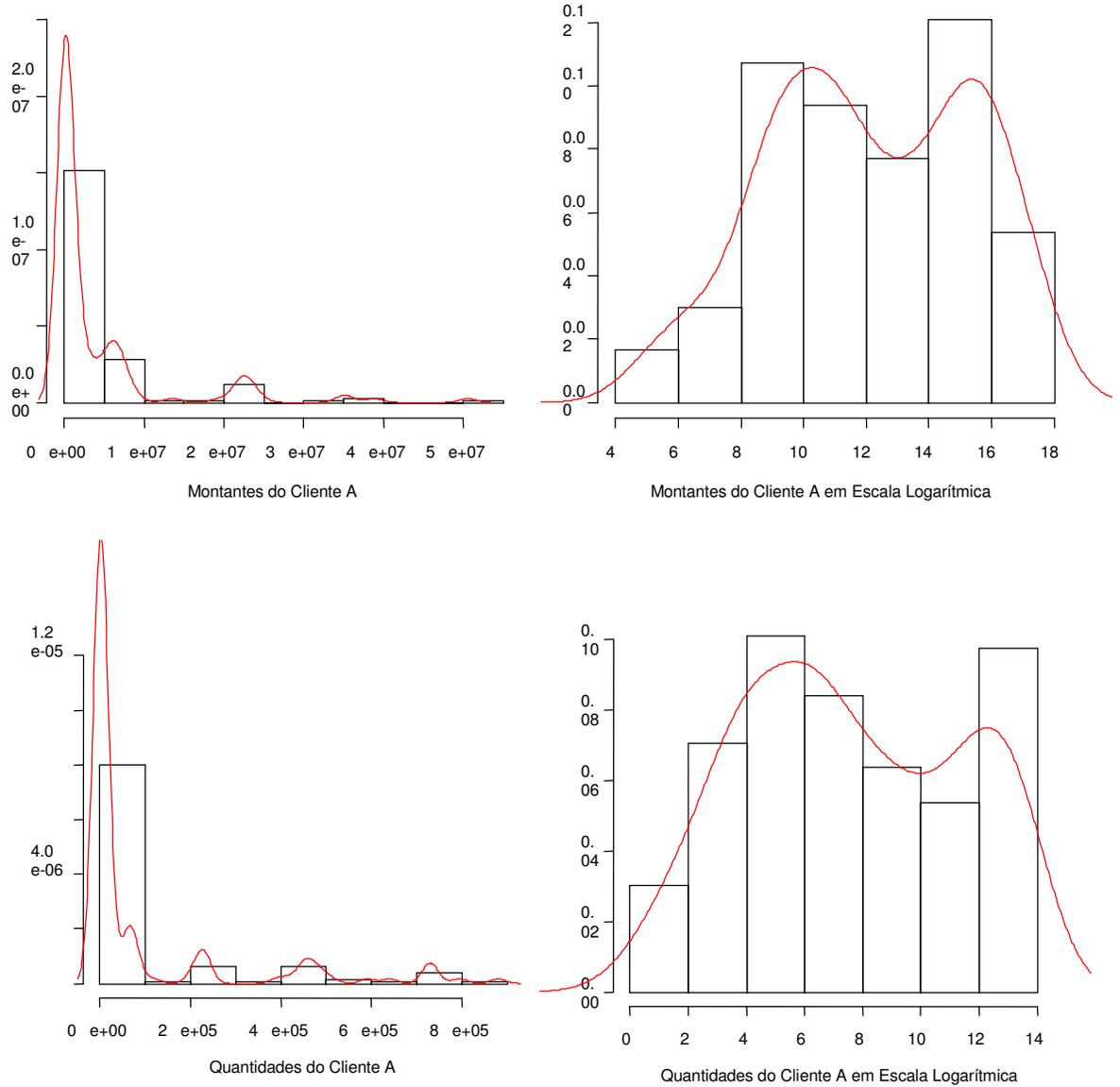


Figura 2.2 – Histogramas para os Dados do Cliente “B”

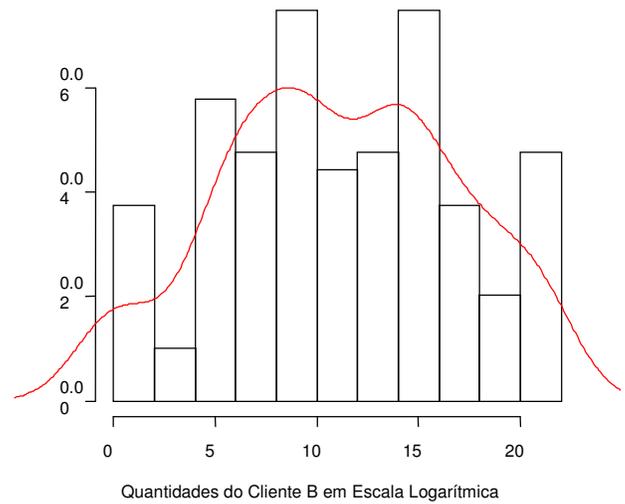
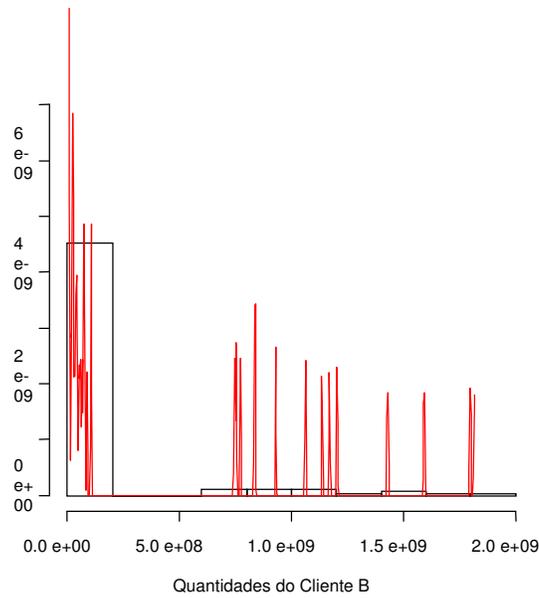
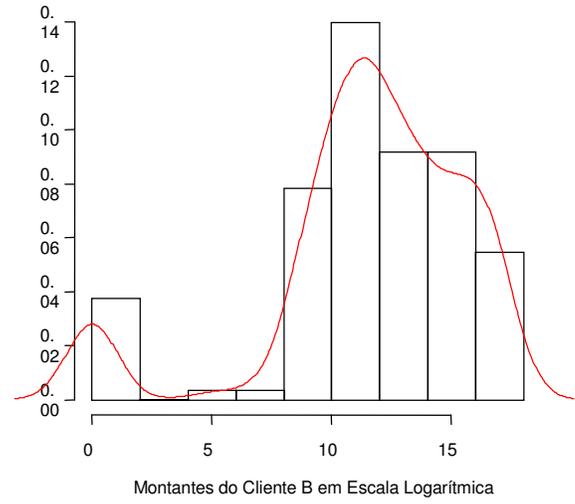
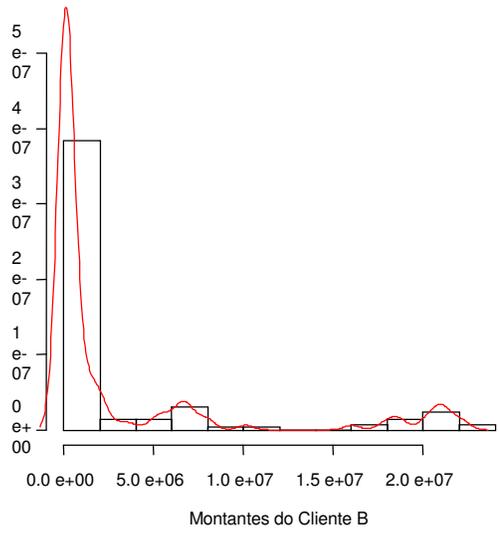


Figura 2.3 – Histogramas para os Dados do Cliente “C”

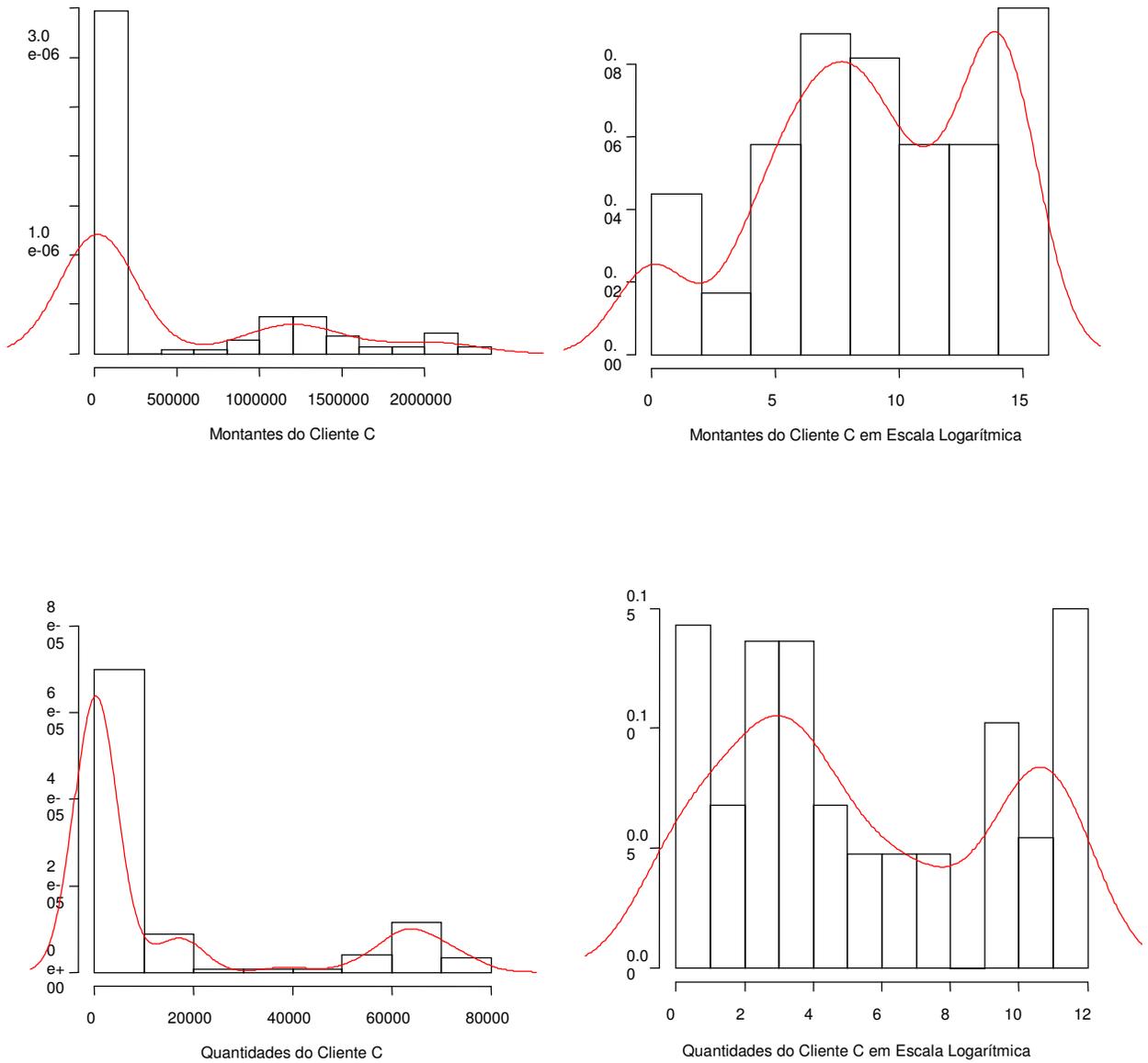
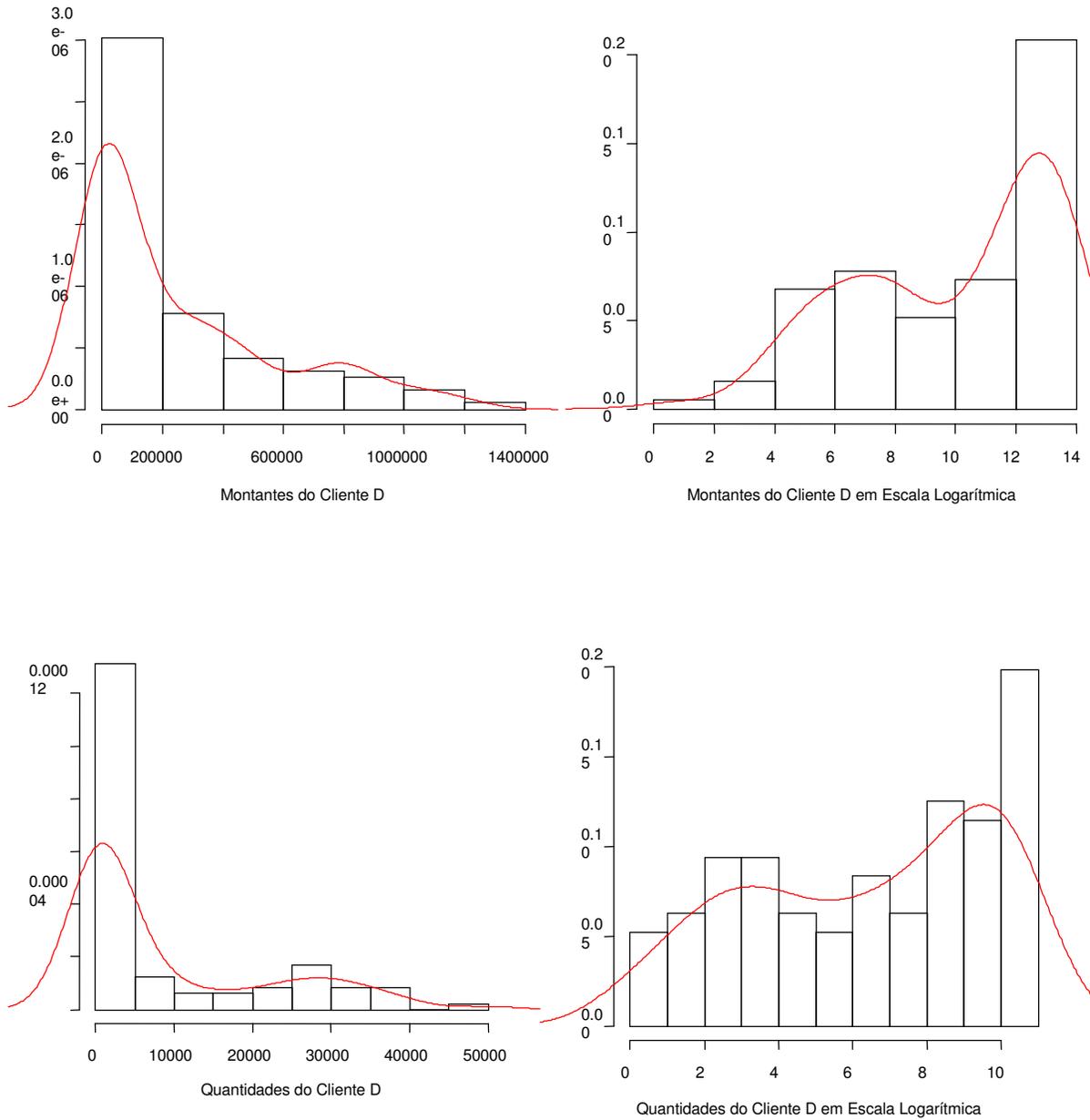


Figura 2.4 – Histogramas para os Dados do Cliente “D”



2.1. Modelos Dinâmicos

Quando os parâmetros de um modelo de regressão evoluem através do tempo, dizemos que este modelo é dinâmico, sendo ele definido através de duas equações:

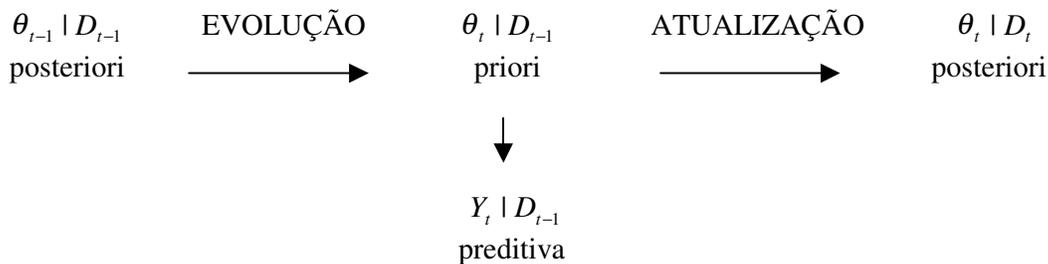
Equação de Observação: $Y_t = F_t' \theta_t + v_t$, com $v_t \sim N[0, V_t]$.

Equação de Sistema: $\theta_t = G_t \theta_{t-1} + \omega_t$, com $\omega_t \sim N[0, W_t]$.

Sendo que Y_t são os dados observados no tempo t , F_t' é um vetor de constantes conhecidas, θ é um vetor de parâmetros e v e ω são termos estocásticos (ruído branco) com distribuição Normal não correlacionados e serialmente não correlacionados, i.e.,

$$Cov(v_t, v_s) = Cov(\omega_t, \omega_s) = 0, \forall t \neq s \text{ e } Cov(v_t, \omega_s) = 0, \forall t, s.$$

O sistema de inferência funciona da seguinte forma:



sendo $D_t = \{ y_1, y_2, \dots, y_t \}$ o conjunto de observações até o tempo t .

Assim, a evolução é feita através da equação do sistema, a atualização é feita através da incorporação da informação obtida em y_t pelo Teorema de Bayes e a previsão é feita com a distribuição preditiva de y_t , dada a informação disponível antes de y_t ser observada. Esta distribuição preditiva é dada por

$$p(Y_t | D_{t-1}) = \int p(Y_t | \theta_{t-1}, D_{t-1}) \cdot p(\theta_{t-1} | D_{t-1}) d\theta_{t-1}.$$

Esta é a distribuição preditiva no tempo t , condicional aos dados observados até o tempo $t-1$. No contexto de MCMC, esta distribuição é facilmente obtida replicando a função de verossimilhança.

Quando temos $\theta_t | D_t$, nossa distribuição é sequencial, com todas as irregularidades características de uma série temporal. Se quisermos suavizá-la, entretanto, para estudar de que forma se dá a tendência, fazemos $\theta_t | D_n$, isto é, estudamos todas as observações conjuntamente, suavizando a tendência. Considerando, por exemplo, que queiramos suavizar a distribuição de θ_t dadas todas as observações, isso se dá pela expressão

$$p(\theta_{t-1} | D_t) = \int p(\theta_{t-1} | \theta_t, D_t) \cdot p(\theta_t | D_t) d\theta_t.$$

Percebemos claramente que o segundo termo do integrando é a posteriori mais atualizada. Já o primeiro termo é dado por

$$p(\theta_{t-1} | \theta_t, D_t) = p(\theta_{t-1} | \theta_t, y_t, D_{t-1}),$$

e aplicando o Teorema de Bayes, temos que

$$p(\theta_{t-1} | \theta_t, D_t) = \frac{p(y_t | \theta_{t-1}, \theta_t, D_{t-1}) p(\theta_{t-1} | \theta_t, D_{t-1})}{p(y_t | \theta_t, D_{t-1})}.$$

Mas para o caso em questão, o primeiro termo do numerador é exatamente o mesmo que o termo do denominador. Cancelando-os, ficamos com a expressão simplificada

$$p(\theta_{t-1} | \theta_t, D_t) = p(\theta_{t-1} | \theta_t, D_{t-1}).$$

Contudo, esta expressão simplificada é usada apenas para o caso Normal. Para casos não-Normais, teremos que efetuar o cálculo de todas as posteriores para cada tempo ($\theta_t | D_t$) e então, obter a distribuição suavizada $p(\theta_1, \dots, \theta_n | D_n)$.

No caso particular do Amostrador de Gibbs, a amostra da posteriori é obtida condicional aos tempos anteriores e aos dados até o tempo n , isto é

$$\theta_t | \theta_1, \dots, \theta_{t-1}, \theta_{t+1}, \dots, \theta_n, D_n$$

Maiores detalhes sobre modelos dinâmicos podem ser encontrados em Pole *et al* (1994) e West e Harrison (1997). Detalhes sobre métodos MCMC para estimação de parâmetros em modelos dinâmicos podem ser obtidos em Gamerman (1998).

Dois dos modelos propostos por Ntzoufras e Dellaportas(2002) levam em conta a natureza dinâmica dos dados. Na próxima seção é feita uma análise destes modelos e são propostas as devidas modificações.

2.2. Modelos Propostos

Em um artigo recente, Ntzoufras e Dellaportas(2002) propõem quatro modelos para estimar as incertezas futuras acerca dos montantes que as empresas de seguros devem ter em caixa para evitar a insolvência. Basicamente, os quatro modelos podem ser resumidos como:

- Modelo Estático levando em conta apenas os montantes de sinistros;
- Modelo Estático levando em conta tanto os montantes quanto as quantidades de sinistros;
- Modelo Dinâmico levando em conta apenas os montantes de sinistros;
- Modelo Dinâmico levando em conta tanto os montantes quanto as quantidades de sinistros;

Cabe salientar, no entanto, que estes modelos não estão levando em conta a natureza bimodal dos dados observados e, além disso, partem do pressuposto de que os totais das quantidades de sinistros são conhecidos. Quando estes totais não são conhecidos, segundo o artigo, pode ser usada a distribuição de Poisson (Verall, 2000), conforme utilizado pela Unimed Seguros (2004).

Poderia ser feita, portanto, a incorporação da idéia da ‘bimodalidade’ nos quatro modelos propostos, através de uma ‘mistura’ de distribuições, o que resultaria em o ito modelos a serem comparados. Entretanto, este estudo visa trabalhar com dados na escala de meses e não em anos, como no artigo de Ntzoufras e Dellaportas. Desta forma, as quantidades de sinistros na parte direita do triângulo de run-off acabam ficando demasiadamente baixas, permitindo que sejam estimados valores nulos ou iguais a 1 para as variáveis próximas desses valores. No citado artigo, a média da distribuição dos montantes é modelada como

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \log(n_{ij}),$$

sendo n_{ij} as quantidades. Isto acarreta um problema quando o algoritmo tenta obter o logaritmo natural de 0 ou de 1 e ele não é capaz de produzir as iterações. Desta forma, iremos trabalhar apenas com os modelos sem as quantidades de sinistros. Sugere-se que a variável “quantidades de sinistros” seja levada em conta em futuros estudos.

Assim sendo, serão comparados os quatro modelos a seguir:

- **Modelo 1: Modelo Estático e sem mistura**

$$\begin{aligned} Y_{ij} &\sim \text{LogN}(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \mu + \alpha_i + \beta_j \\ \mu &\sim N(0, \sigma_\mu^2), \quad \alpha \sim N(0, \sigma_\alpha^2), \quad \beta \sim N(0, \sigma_\beta^2), \end{aligned}$$

- **Modelo 2: Modelo Estático e com mistura**

$$\begin{aligned} Y_{ij} &\sim \text{LogN}(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \mu + \alpha_i + \beta_j \\ \mu &\sim P.N(\theta_1, \sigma_\mu^2) + (1-P).N(\theta_2, \sigma_\mu^2), \quad \alpha \sim N(0, \sigma_\alpha^2), \quad \beta \sim N(0, \sigma_\beta^2), \end{aligned}$$

- **Modelo 3: Modelo Dinâmico e sem mistura**

$$\begin{aligned} Y_{ij} &\sim \text{LogN}(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \mu + \alpha_i + \beta_j \\ \mu &\sim N(0, \sigma_\mu^2) \\ \alpha_i &= \alpha_{i-1} + h_i, \quad h_i \sim N(0, \sigma_h^2) \\ \beta_j &= \beta_{j-1} + v_j, \quad v_i \sim N(0, \sigma_v^2) \end{aligned}$$

- **Modelo 4: Modelo Dinâmico e com mistura**

$$\begin{aligned} Y_{ij} &\sim \text{LogN}(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \mu + \alpha_i + \beta_j \\ \mu &\sim P.N(\theta_1, \sigma_\mu^2) + (1-P).N(\theta_2, \sigma_\mu^2) \\ \alpha_i &= \alpha_{i-1} + h_i, \quad h_i \sim N(0, \sigma_h^2) \\ \beta_j &= \beta_{j-1} + v_j, \quad v_i \sim N(0, \sigma_v^2) \end{aligned}$$

Em todos os modelos temos que $i = 2, \dots, r$; $j = 2, \dots, r$ e é atribuída uma priori Gama não informativa para as variâncias das distribuições Normais. Para identificabilidade do modelo, é atribuída também a restrição $\alpha_1 = \beta_1 = 0$. Em cada modelo, temos que

Y_{ij} são os montantes dos sinistros;

μ é o valor esperado dos montantes de sinistros;

α_i é o efeito do i -ésimo mês;

β_j é o efeito do j -ésimo período de atraso;

P é o peso atribuído às parcelas da soma de distribuições.

Nota-se que, para comportar a idéia da "bimodalidade", foi considerada, ao invés de uma distribuição Log-Normal para os montantes de sinistros, uma mistura de distribuições Log-Normais, com médias diferentes e mesma variância.

Neste trabalho, esses quatro modelos serão implementados no pacote WinBUGS (*Bayesian Inference Using Gibbs Sampling*) a fim de serem comparados entre si e com a técnica clássica "Chain Ladder".

2.3. Critérios de Comparação

Um tipo de critério que vem sendo bastante utilizado para comparação de modelos é o DIC (*Deviance Information Criterion*), que já está implementado no WinBUGS. Contudo, há uma restrição quanto ao uso do DIC, uma vez que ele assume que a média a posteriori é um bom estimador dos parâmetros estocásticos, ou seja, não deve haver valores muito extremos ou bimodalidade nos dados. Como dois dos modelos propostos comportam esta bimodalidade, será utilizado outro critério bastante comum, o EPD (*Expected Predictive Deviance*), para avaliação das estimativas pontuais, sendo ele dado por:

$$EPD = \sum_{i=1}^k (\theta_i^p - \theta_i^o)^2 + \sum_{i=1}^k \sigma_i^2$$

Onde:

θ_i^p o parâmetro estimado pelo modelo (em MCMC, é obtido pela média dos valores simulados);

θ_i^v o verdadeiro valor do parâmetro, observado;

σ_i^2 a variância dos valores simulados para obter a estimativa do parâmetro;

O primeiro termo da soma pode ser vista como uma medida de “bondade de ajuste” do modelo, enquanto que a segunda é um termo de penalidade. O EPD é calculado para cada modelo e é escolhido aquele que produzir o menor valor.

Outra razão para o uso do EPD é que o foco deste estudo é na predição dos valores observáveis, ou seja, só faz sentido comparar modelos em termos de sua capacidade preditiva, já que a operadora está interessada nos valores futuros que ela necessitará ter em caixa para arcar com suas responsabilidades financeiras.

Maiores detalhes sobre critérios de comparação podem ser encontrados em Neves (2004).

2.4. Inferência Bayesiana e MCMC

Na Inferência Clássica, considera-se que os parâmetros do modelo são constantes, enquanto que na Inferência Bayesiana, esses parâmetros são tratados como variáveis aleatórias, seguindo uma determinada distribuição de probabilidade. É atribuída uma distribuição à priori para estes parâmetros, que reflete o conhecimento prévio que temos acerca deles, antes dos dados serem observados.

O objetivo é encontrar uma distribuição a posteriori para os parâmetros de interesse. Pelo Teorema de Bayes, temos que a posteriori é proporcional ao produto da função de verossimilhança (dados observados) pela distribuição a priori:

$$p(\theta | x) \propto p(x | \theta)p(\theta)$$

Contudo, efetuar esta operação analiticamente muitas vezes pode não ser possível ou ser demasiadamente trabalhoso. Deste modo, nos últimos anos, com o avanço dos recursos computacionais, tem-se pesquisado maneiras alternativas para contornar este problema.

Uma solução sugerida que tem sido bastante utilizada é encontrar uma posteriori aproximada ao invés de calcular-se a expressão exata da distribuição de interesse. Nesse sentido, o Método de Monte Carlo tem sido estudado e muitas alternativas surgiram, dentre elas o Método de Monte Carlo via Cadeias de Markov (MCMC), que pode ser posto em prática através de dois algoritmos principais: O Algoritmo de Metropolis-Hastings e o Algoritmo de Gibbs.

O software WinBUGS trabalha com o Algoritmo de Gibbs, que basicamente gera amostras aleatórias de uma distribuição de interesse, a partir de cadeias de Markov. Estas cadeias são obtidas através das distribuições condicionais completas dos parâmetros, por exemplo:

$$\begin{aligned}\theta_1^t &\sim \pi(\theta_1 | \theta_2^{t-1}, \theta_3^{t-1}, \dots, \theta_d^{t-1}) \\ \theta_2^t &\sim \pi(\theta_2 | \theta_1^{t-1}, \theta_3^{t-1}, \dots, \theta_d^{t-1}) \\ &\vdots \\ \theta_d^t &\sim \pi(\theta_d | \theta_1^t, \theta_2^t, \dots, \theta_{d-1}^t)\end{aligned}$$

Desta forma, ao final das iterações, tem-se uma série de valores simulados dos parâmetros, que formam uma amostra aproximada da distribuição a posteriori. O WinBUGS retorna as médias destes parâmetros, que podem ser utilizadas como estimadores para os parâmetros de interesse.

É preciso ter certo cuidado, contudo, pois quando se lida com cadeias de Markov, os valores podem ser correlacionados. É sempre necessário verificar não apenas se a cadeia convergiu para a distribuição de interesse, como também averiguar se os valores gerados são aproximadamente não-correlacionados. Além disso, é interessante certificar-se de que é possível amostrar das condicionais completas que estão sendo propostas.

Maiores detalhes sobre o Amostrador de Gibbs podem ser encontrados em Gamerman (1997, 2006).

2.5. Exemplo de funcionamento do WinBUGS

Este exemplo foi extraído do Volume 01 dos exemplos implementados no próprio pacote. Trata-se de modelos hierárquicos utilizados nas priors conjugadas Gamma-Poisson.

Neste caso, é discutida uma análise Bayesiana para modelos hierárquicos onde a priori conjugada é adotada no primeiro nível, mas para qualquer outra distribuição a priori dada para os hiperparâmetros, suas respectivas posteriors não serão fechadas por amostragem. Assim sendo, devemos obter uma posteriori aproximada para estes parâmetros. O WinBUGS sempre calcula uma distribuição aproximada por extrair uma amostra da densidade. O exemplo abaixo considera 10 bombas de usina de força, onde o número de falhas x_i tem distribuição de Poisson.

$$X_i | \theta_i \sim \text{Poisson}(\theta_i, t_i), \quad i = 1, \dots, 10,$$

onde θ_i é a taxa de falha por bomba i e t_i é o tempo de operação da bomba (em 1000 segundos de hora). Os dados são resumidos abaixo:

Pump	t_i	x_i
1	94.5	5
2	15.7	1
3	62.9	5
4	126	14
5	5.24	3
6	31.4	19
7	1.05	1
8	1.05	1
9	2.1	4
10	10.5	22

Deseja-se estimar a taxa média de falhas para cada bomba. Assim, como a conjugada da Poisson é a Distribuição Gama, assumiu-se uma Distribuição Gama para θ_i , ou seja

$$\theta_i \sim \text{Gama}(\alpha, \beta), \quad i = 1, \dots, 10.$$

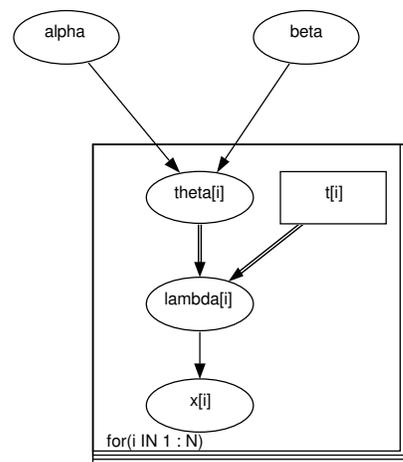
Como trata-se de um modelo hierárquico, devemos estabelecer prioris também para os hiperparâmetros. Então, no segundo nível da hierarquia, assumiu-se que:

$$\alpha \sim \text{Exponencial}(1.0)$$

$$\beta \sim \text{Gama}(0.1, 1.0)$$

Desta forma, teríamos uma posteriori Gama para β , mas não conhecemos uma posteriori padrão para α . Conseqüentemente, usamos o Amostrador de Gibbs para simular uma densidade a posteriori aproximada para os hiperparâmetros, conseguindo, assim, uma posteriori (também aproximada) para $\theta_i | X_i$.

Diagrama de Influência para o Exemplo das Bombas



Fonte: WinBUGS Manual.

O WinBUGS trabalha com três arquivos, sendo o primeiro deles para especificação do modelo, o segundo para entrada dos dados (que irão compor a verossimilhança) e por fim, os valores iniciais atribuídos para que se possa rodar as iterações. Para este exemplo, os arquivos conteriam as seguintes informações:

```

model
{
  for (i in 1 : N) {
    theta[i] ~ dgamma(alpha, beta)
    lambda[i] <- theta[i] * t[i]
    x[i] ~ dpois(lambda[i])
  }
  alpha ~ dexp(1)
  beta ~ dgamma(0.1, 1.0)
}

data
list(t = c(94.3, 15.7, 62.9, 126, 5.24, 31.4, 1.05, 1.05, 2.1, 10.5),
     x = c( 5,  1,  5, 14,  3, 19,  1,  1,  4, 22), N = 10)

initial
list(alpha = 1, beta = 1)

```

É preciso gerar iterações de aquecimento, chamadas de “burn-in”, para evitar a autocorrelação da cadeia. Fazendo um aquecimento de 1000 iterações, seguido de um processo de 10.000, obtemos os seguintes resultados:

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
alpha	0.6867	0.265	0.004712	0.2886	0.6519	1.297	1001	10000
beta	0.9024	0.529	0.009881	0.1863	0.8023	2.201	1001	10000
theta[1]	0.05986	0.02488	2.26E-4	0.02134	0.0565	0.1176	1001	10000
theta[2]	0.1015	0.07971	8.03E-4	0.0085	0.0817	0.3045	1001	10000
theta[3]	0.08899	0.0375	3.893E-4	0.03056	0.0836	0.1774	1001	10000
theta[4]	0.1156	0.03018	3.267E-4	0.06408	0.1134	0.1808	1001	10000
theta[5]	0.6043	0.3181	0.003656	0.1478	0.5518	1.348	1001	10000
theta[6]	0.6121	0.14	0.001406	0.3686	0.6009	0.9133	1001	10000
theta[7]	0.899	0.7265	0.00671	0.08038	0.7092	2.758	1001	10000
theta[8]	0.9095	0.7333	0.008292	0.07935	0.7213	2.803	1001	10000
theta[9]	1.587	0.7769	0.008188	0.4597	1.454	3.461	1001	10000
theta[10]	1.995	0.427	0.004109	1.254	1.965	2.913	1001	10000

Não é difícil concluir que, apesar de não conhecermos a distribuição a posteriori de $\theta_i | x_j$, conseguimos obter as esperanças estimadas para os parâmetros de interesse. Assim, sabemos que α e β têm, em média, os valores de 0,6867 e 0,9024, respectivamente. Além disso, obtivemos a média de taxa de falhas para cada uma das 10 bombas. É interessante notar que, como se trata das médias das distribuições para cada parâmetro, temos estimadores de bayes por perda quadrática. Caso desejássemos um estimador por perda absoluta, bastaria utilizarmos a mediana a posteriori.

O WinBUGS também fornece o desvio-padrão da estimativa (sd) e o erro para a média estimada via MCMC (MC error). Se os valores da média forem não correlacionados, este erro será σ / \sqrt{n} . Mas, por definição da Cadeia de Markov, temos que estes valores serão correlacionados. Para contornar este problema, Roberts (1996, pág. 50) propõe que sejam usados lotes (batchs) de valores simulados e se calcule as médias desses lotes. Estas médias serão aproximadamente não correlacionadas e sendo n o número de lotes, pode-se obter uma estimativa para o desvio-padrão e conseqüentemente, para o erro.

Os valores “val2.5pc” e “val97.5pc” nada mais são do que os limites inferiores e superiores para o Intervalo de Credibilidade de 95% de confiança, respectivamente. A coluna “start” nos diz simplesmente qual o número da iteração inicial daqueles resultados, enquanto “sample” é o tamanho da amostra (número de iterações produzidas).

Agora, suponha que quiséssemos encontrar a distribuição preditiva de X . Para isso, basta replicar a função de verossimilhança, acrescentando, no modelo:

```
for (i in 1 : N){  
  xrep[i] ~ dpois(lambda[i])}
```

E em seguida, monitorar a variável “xrep”. Esse procedimento é feito porque o WinBUGS não permite que se monitore diretamente a variável “x”, pois isso obviamente não faz sentido, já que “x” são os dados observados. Fazendo esse monitoramento e rodando novas iterações, obtemos os seguintes resultados:

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
xrep[1]	5.663	3.382	0.03245	1.0	5.0	14.0	1	11000
xrep[2]	1.602	1.765	0.01565	0.0	1.0	6.0	1	11000
xrep[3]	5.655	3.396	0.03247	1.0	5.0	14.0	1	11000
xrep[4]	14.57	5.396	0.05116	5.0	14.0	26.0	1	11000
xrep[5]	3.162	2.431	0.02436	0.0	3.0	9.0	1	11000
xrep[6]	19.15	6.185	0.06288	8.0	19.0	33.0	1	11000
xrep[7]	0.9308	1.226	0.01107	0.0	1.0	4.0	1	11000
xrep[8]	0.9447	1.233	0.01207	0.0	1.0	4.0	1	11000
xrep[9]	3.311	2.471	0.02678	0.0	3.0	9.0	1	11000
xrep[10]	20.87	6.41	0.05768	10.0	20.0	35.0	1	11000

A interpretação, para este caso (usando a mediana como estimador), é que para a bomba 1, o número de falhas esperado é 5. Para a bomba 2, o número esperado seria 1, e assim por diante.

2.6. Implementação para estimação da IBNR

Para implementação dos modelos no WinBUGS, foram utilizados três conjuntos de dados de diferentes operadoras de planos de saúde fornecidos pela *Actuarial Assessoria e Consultoria Atuarial Ltda.* Utilizou-se um período tal que tanto o triângulo superior quanto o inferior fossem conhecidos, a fim de se verificar quão bem o modelo está fazendo suas predições para o triângulo inferior e, conseqüentemente, para os totais estimados. Desta forma, retiramos as informações do triângulo inferior e iremos utilizar os dados do triângulo superior para estimá-las e obter a diferença entre o total observado e o total estimado.

Os dados utilizados para os três exemplos podem ser visualizados nas tabelas 1 a

3.

Tabela 1 - Montantes de sinistros observados para o Exemplo 01 - Maio/2003 a Abril/2004

Data / Ocorrência	Meses de Atraso (em Relação ao Aviso)											
	0	1	2	3	4	5	6	7	8	9	10	11
maio-03	4802850,44	17441804,59	4187133,72	986868,99	162719,47	43455,92	54711,78	17247,51	1838,40	22931,88	1505,94	1429,38
junho-03	5754643,14	16393109,71	3466015,88	476135,99	123730,16	151586,32	37286,08	7258,55	9053,96	11472,11	1118,54	NA
julho-03	6345038,34	18993649,18	4483928,68	519284,86	243013,63	101922,59	33312,53	39304,75	27304,29	3203,13	NA	NA
agosto-03	5685354,99	18714061,71	4045084,57	798964,94	119793,58	32068,08	16992,17	60210,91	4113,60	NA	NA	NA
setembro-03	5232596,56	18709616,07	4149276,34	751410,28	165674,81	273579,24	25147,54	15592,55	NA	NA	NA	NA
outubro-03	6028734,64	18908539,86	4082696,08	981156,37	288234,90	210038,25	39179,13	NA	NA	NA	NA	NA
novembro-03	4900277,19	16810779,81	3934517,91	690947,00	127896,57	67192,65	NA	NA	NA	NA	NA	NA
dezembro-03	5285268,37	13870982,04	3437329,48	735041,12	132101,02	NA	NA	NA	NA	NA	NA	NA
janeiro-04	4218038,58	16093191,27	3502019,43	1072454,57	NA	NA	NA	NA	NA	NA	NA	NA
fevereiro-04	4833407,81	14054768,13	3293114,14	NA	NA	NA	NA	NA	NA	NA	NA	NA
março-04	6341794,00	18918359,50	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
abril-04	5566210,76	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Fonte: Actuarial Cons. e Asses. Atuarial Ltda.

Tabela 2 - Montantes de sinistros observados para o Exemplo 02 - Maio/2003 a Abril/2004

Data / Ocorrência	Meses de Atraso (em Relação ao Aviso)											
	0	1	2	3	4	5	6	7	8	9	10	11
maio-03	104405,55	18882505,20	5339018,30	1082301,53	533148,36	227604,22	188074,51	220422,70	60436,90	71729,64	13551,60	7200,61
junho-03	63732,48	21092090,21	5691012,37	1734074,65	937720,00	588785,88	387067,98	136283,04	149759,49	43361,55	28692,32	NA
julho-03	76120,70	17785830,16	6753158,28	1924892,33	1087263,59	473464,89	128460,39	12260,14	13626,80	14993,46	NA	NA
agosto-03	13151,32	21360902,01	7485911,14	2159697,64	729506,61	54226,33	146871,35	5501,10	1294,60	NA	NA	NA
setembro-03	64390,79	18448094,70	6867912,09	1991785,47	384388,03	76510,11	47590,02	5148,88	NA	NA	NA	NA
outubro-03	113827,64	21354326,82	6447792,73	1662407,54	304739,23	130013,17	135079,95	NA	NA	NA	NA	NA
novembro-03	64712,62	18349482,31	6986931,58	1806099,98	702130,58	366987,19	NA	NA	NA	NA	NA	NA
dezembro-03	67306,19	20401934,03	6484635,35	1287153,77	422677,60	NA	NA	NA	NA	NA	NA	NA
janeiro-04	88808,34	20807287,69	5393905,90	1565329,83	NA	NA	NA	NA	NA	NA	NA	NA
fevereiro-04	71331,24	20594792,10	6191115,29	NA	NA	NA	NA	NA	NA	NA	NA	NA
março-04	66015,30	22072511,81	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
abril-04	91701,38	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Fonte: Actuarial Cons. e Asses. Atuarial Ltda.

Tabela 3 - Montantes de sinistros observados para o Exemplo 03 - Maio/2003 a Abril/2004

Data / Ocorrência	Meses de Atraso (em Relação ao Aviso)											
	0	1	2	3	4	5	6	7	8	9	10	11
maio-03	985228,22	1679670,00	1000307,93	117199,42	8375,17	9132,42	760,63	1888,12	101,00	1017,20	497,84	4355,13
junho-03	1474402,49	1564465,83	663527,93	116740,24	27003,64	2094,35	350,12	362,49	175,70	16599,59	52,00	NA
julho-03	1314415,61	2198484,20	1020437,65	97263,06	9994,01	13901,69	645,55	2956,10	1011,18	1291,98	NA	NA
agosto-03	1220559,21	1810571,45	963354,05	82864,75	10059,22	2694,70	8366,95	769,26	48,39	NA	NA	NA
setembro-03	1314524,22	1947226,74	1173037,10	114469,31	15596,70	7818,35	4920,10	75,60	NA	NA	NA	NA
outubro-03	1252848,48	2038015,46	1343848,25	30380,96	4506,83	1076,65	348,23	NA	NA	NA	NA	NA
novembro-03	1242319,94	2044887,18	1215045,00	95925,03	26962,70	2150,36	NA	NA	NA	NA	NA	NA
dezembro-03	1347404,70	2242265,52	1196271,19	129551,03	2969,86	NA	NA	NA	NA	NA	NA	NA
janeiro-04	1072224,39	1784386,29	1229406,41	137478,22	NA	NA	NA	NA	NA	NA	NA	NA
fevereiro-04	1107892,29	2062246,88	1170858,50	NA	NA	NA	NA	NA	NA	NA	NA	NA
março-04	1103378,37	2044992,34	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
abril-04	1155161,57	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Fonte: Actuarial Cons. e Asses. Atuarial Ltda.

Depois de efetuadas 20.000 iterações de aquecimento, seguidas de outras 50.000, obtivemos os resultados que são resumidos na tabela 4 e nos gráficos 1 a 3.

Tabela 4 - Resumo de Resultados para os três conjuntos de dados - 2006

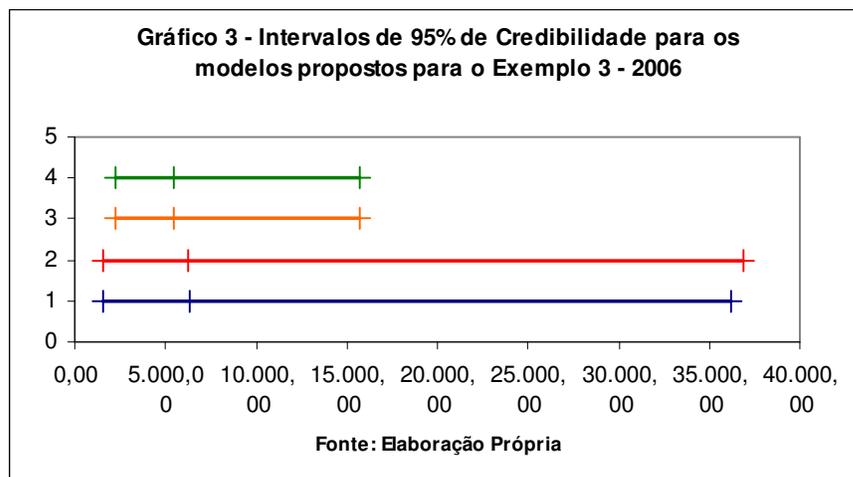
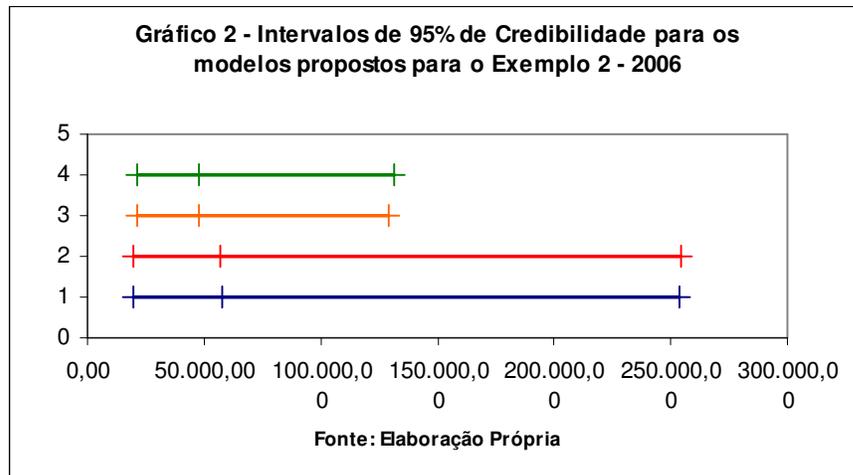
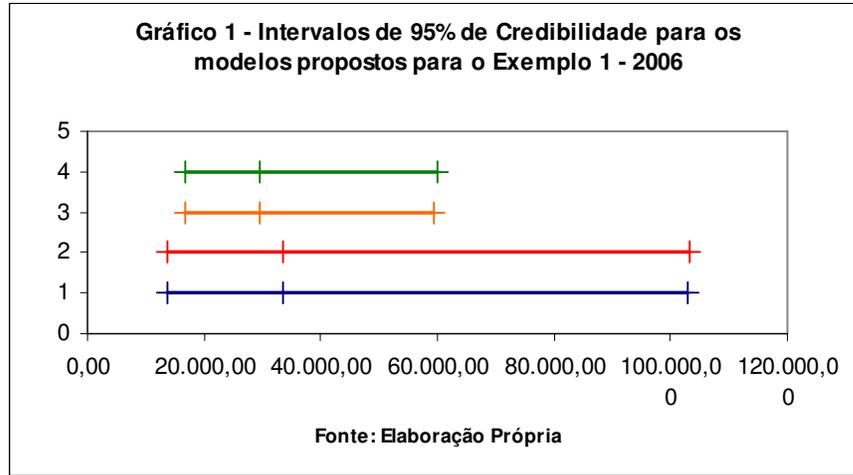
	Critério de Estimação	IBNR Estimada	Diferença (Valor Observado)	EPD	EQM
Exemplo 01	Modelo 1	34.535.557,38	3.656.710,70	71,47	145.461,37
	Modelo 2	34.528.650,96	3.649.804,28	71,47	145.622,77
	Modelo 3	30.095.914,81	-782.931,87	58,06	134.573,36
	Modelo 4	30.401.423,78	-477.422,90	58,15	134.864,63
	Chain Ladder	30.250.595,89	-628.250,79	-	174.871,48
	Critério de Estimação	IBNR Estimada	Diferença (Valor Observado)	EPD	EPD
Exemplo 02	Modelo 1	60.408.476,96	9.692.730,55	139,80	328.910,30
	Modelo 2	60.251.016,41	9.535.270,00	139,63	328.048,62
	Modelo 3	48.991.628,77	-1.724.117,64	125,37	345.738,02
	Modelo 4	49.055.359,30	-1.660.387,11	124,97	351.366,42
	Chain Ladder	47.818.496,01	-2.897.250,40	-	386.513,36
	Critério de Estimação	IBNR Estimada	Diferença (Valor Observado)	EPD	EPD
Exemplo 03	Modelo 1	6.609.179,01	1.557.262,27	197,19	7.775,94
	Modelo 2	6.614.534,61	1.562.617,87	197,38	7.645,84
	Modelo 3	5.549.841,91	497.925,17	157,83	7.898,11
	Modelo 4	5.565.625,86	513.709,12	158,39	7.885,29
	Chain Ladder	4.513.614,45	-538.302,29	-	15.498,23

Fonte: Elaboração Própria

A título de informação, segue o tempo de operação do WinBUGS para cada modelo, em um PC de 733 Mhz com 256MB de memória, segundo o primeiro conjunto de dados. Os demais conjuntos tiveram tempos bastante semelhantes.

Modelo	Tempo (em segundos)
1	57
2	58
3	349
4	353

Nos anexos, podem ser encontrados os arquivos detalhados do WinBUGS utilizados para implementação, bem como alguns gráficos para análise da convergência da cadeia e a saída do WinBUGS para o primeiro conjunto de dados.



Capítulo 3

Resultados e Conclusões

A fim de encontrar uma nova maneira para estimação da provisão de IBNR, foram propostos modelos Bayesianos fundamentados em estudos anteriores para estimação da parte inferior do triângulo de run-off.

Em primeira análise, os dois primeiros modelos Bayesianos (estáticos) parecem superestimar demasiadamente o valor dos montantes de sinistros a serem pagos pela operadora. Se olharmos para os intervalos de credibilidade associados a esses dois modelos, verificamos que sua larga amplitude pode ser um problema. Contudo, cabe salientar que a distribuição dos totais estimados é assimétrica e concentrada à esquerda, ou seja, em valores mais baixos. Estes modelos podem ser ideais caso a operadora deseje manter em caixa um montante maior do que o necessário, a fim de trabalhar com uma margem de segurança e evitar riscos eventuais.

Os modelos dinâmicos parecem se sair melhor não apenas na estimação de um intervalo de credibilidade mais restrito, como também nas estimativas dos totais (soma do triângulo inferior). Sobretudo o modelo 4, que incorpora a idéia da bimodalidade, já que, nos casos analisados, foi o modelo que melhor conseguiu se aproximar do valor observado da provisão IBNR, superando as estimativas da técnica “Chain Ladder”, que parece subestimar o verdadeiro valor dos montantes. Desta forma, dentre os modelos e exemplos estudados, os resultados analisados indicaram que o quarto modelo proposto (dinâmico e com mistura de densidades) acabou se mostrando o melhor preditor para a estimativa da IBNR.

Apesar disso, se compararmos os valores de EPD entre os quatro modelos, verificamos que não houve diferença significativa entre os modelos com mistura e os modelos sem mistura. Isto se deve ao fato de que, nas estimativas pontuais, não parece fazer diferença se o modelo comporta a bimodalidade ou não. O que realmente influencia é a grande variabilidade dos dados, pois, na parte esquerda da tabela, onde os valores são maiores e a variabilidade é baixa, a diferença entre observado e estimado não é alta. Por outro lado, na parte direita, onde os valores são menores e a variabilidade é maior, esta diferença é mais alta. Ao se calcular o Erro Quadrático Médio das estimativas pontuais com relação aos seus verdadeiros valores, os resultados obtidos são bastante similares aos do EPD, pelo mesmo motivo. O que ocorre é que, ao somarmos as estimativas pontuais, a ineficiência do modelo nas estimativas de maior variabilidade é compensada por sua eficácia nas estimativas de menor variabilidade, o que resulta na estimação de uma IBNR mais condizente com a realidade da seguradora.

No que diz respeito às estimativas do método Chain Ladder nos três exemplos, a análise aponta para a possibilidade de que este método estimará os totais dos montantes de sinistros dentro do intervalo de credibilidade estimado pelo método Bayesiano e na concentração de maior probabilidade, já que ele subestima o valor do montante, o que pode ser uma vantagem, dada a sua praticidade. Contudo, não há meios de se garantir que isso sempre ocorra e as técnicas Bayesianas são mais aconselháveis para estimação não apenas da provisão total, mas também das parcelas separadas. Além disso, sempre é interessante poder verificar a amplitude do Intervalo de Credibilidade, já que a operadora pode estar interessada em saber dentro de quais limites estariam os valores dos montantes de que ela necessita para evitar a insolvência.

Cabe salientar que é sempre necessário verificar qual modelo melhor se adapta à natureza de cada operadora, uma vez que uma IBNR superestimada pode comprometer os lucros, mas uma IBNR subestimada põe em risco a solvência da operadora.

Outro problema bastante comum nesse tipo de dados é o fato das cédulas da parte superior do triângulo de run-off nem sempre poderem vir todas preenchidas. Como não há meios de saber se estes “buracos” nas informações são informações faltantes ou, simplesmente, valores nulos, isso acarreta uma gigantesca variabilidade nos dados, o que, conforme já foi demonstrado, certamente influenciará as estimativas finais.

Em trabalhos futuros, sugere-se que seja feita a comparação entre o modelo aqui proposto com outros já citados na literatura, ou mesmo com outras técnicas tradicionais, além de se tratar do problema dos “buracos” nas informações. Além disso, cabe salientar que o modelo aqui proposto ainda é passível de melhorias, como a inclusão das quantidades de sinistros no modelo, ou mesmo a utilização de prioris diferentes das abordadas neste estudo.

Desta forma, espera-se encontrar um modelo cada vez melhor para estimação da IBNR ou de outras provisões que utilizem esta mesma metodologia.

Referências Bibliográficas

BOULTER, Anthony; GRUBBS, Dawson, 2000. **Late Claims Reserves in Reinsurance**. Swiss Reinsurance Company, Zurich.

CASTIGLIONE, Luiz Roberto, 1997. **Seguros: Conceitos e Critérios de Avaliação de Resultados**, Editora Manuais Técnicos de Seguros, São Paulo, SP.

EHLERS, Ricardo S., 2003. **Apostila: Análise de Séries Temporais**, Departamento de Estatística – UFPR, Curitiba, Paraná. Disponível em <http://leg.ufpr.br/~ehlers/stemp>

EHLERS, Ricardo S., 2003. **Apostila: Introdução à Inferência Bayesiana**, Departamento de Estatística – UFPR, Curitiba, Paraná.

EHLERS, Ricardo S. e MELO, Leonardo G, 2006. **Fully Bayesian Approach for Computing Claim Amounts of Occuring but Not Reported Events**.

FONSECA, Thais C. O.; VALLE, Carlos A.A., 2006. **Fluxo de Pagamento de Sinistros: Aspectos Computacionais e Aplicações**. Relatório Técnico n. 4, Rio de Janeiro – RJ.

GAMERMAN, Dani, 1997. **Curso de Modelos Dinâmicos**, Instituto de Pesquisa Econômica Aplicada, Rio de Janeiro, RJ.

GAMERMAN, Dani, 1997. **Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference**. Chapman and Hall, London, U. K.

GAMERMAN, Dani, 1998. **Monte Carlo Markov Chain for Dynamic Generalized Linear Models**. *Biometrika* 85, 215-227.

MOREIRA, William., 2005. **Uma Proposta para Avaliação do Capital Baseado no Risco de subscrição de Provisões**. Tese de Mestrado submetida à Universidade do Rio de Janeiro, RJ – Brasil.

NASCIMENTO, Fernando F., 2006. **Uma metodologia semi-paramétrica para IBNR(Incurred but not Reported)**. Tese de Mestrado submetida à Universidade Estadual de Campinas, Campinas – SP.

NEVES, César, 2004. **Graduação Bayesiana de Taxas de Mortalidade**, Tese de Mestrado submetida à Universidade do Rio de Janeiro, RJ – Brasil.

NEVES, César, 2004. **Graduação Bayesiana de Taxas de Mortalidade: Uma Aplicação na Cobertura de Sobrevivência e na Avaliação da Provisão Matemática.** Revista Brasileira de Risco e Seguro, p. 90-111, Rio de Janeiro, RJ.

NTZOUFRAS, Ioannis, 1999. **Aspects of Bayesian Model and Variable Selection via MCMC.** Tese de Doutorado. Athens University of Economics and Business, Atenas – Grécia.

NTZOUFRAS, Ioannis e DELLAPORTAS, Petros, 2002. **Bayesian Modelling of Outstanding Liabilities Incorporating Claim Count Uncertainty.** North American Actuarial Journal, 6(1), 113-136.

OLIVEIRA JR., Gilmar J., 2004. **Método Bayesiano no Cálculo das Provisões de Sinistros Ocorridos Mas Não Avisados (IBNR) e Provisões de Sinistros a Liquidar (PSL),** FUNENSEG / UFRJ, Rio de Janeiro, RJ.

POLE, Andy, WEST, Mike e JEFF, Harrison, 1994. **Applied Bayesian Forecasting and Time Series Analysis,** Chapman & Hall, New York, NY.

ROBERTS, G. O., 1996. **Markov chain concepts related to sampling algorithms.** Em W. R. Gilks, S. Richardson e D. J. Spiegelhalter (Eds.). *Markov Chain Monte Carlo in Practice.* Chapman and Hall, London, U.K.

SPIEGELHATTER, David; THOMAS, Andrew; BEST, Nicky; LUNN, David, 2003. **WinBUGS User Manual,** <http://www.mrc-bsu.cam.uk/bugs>, Version 1.4.

STEVENSON, William J., 1981. **Estatística Aplicada à Administração,** Harbra, São Paulo, SP.

STRAUB, Erwin, 1980. **Non-Life Insurance Mathematics.** Association of Swiss Actuaries, Zurich.

UNIMED Seguros, 2004. **IBNR, Sob a Visão Estocástica.** Responsável: Lara Cristina Facchini, lara@unimedseguros.com.br, (11) 3265-9856.

VERRALL, R.J., 2000. **An investigation into stochastic claims reserving models and the chain-ladder technique.** Insurance: Mathematics & Economics 26, p. 91-99.

WEST, M ; HARRISON, P.J., 1997. **Bayesian Forecasting and Dynamic Models.** Springer Verlag, New York.

Anexo 1 – Comandos do R utilizados para construção dos histogramas

```
hist(v,prob=T,ylim=range(c(0,2.5*10^(-07))),xlab='Montantes do Cliente A')
lines(density(v),col=2)
hist(log(v),prob=T,xlab='Montantes do Cliente A em Escala Logarítima')
lines(density(log(v)),col=2)
hist(q,prob=T,ylim=range(c(0,13*10^(-06))),xlab='Quantidades do Cliente A')
lines(density(q),col=2)
hist(log(q),prob=T,xlab='Quantidades do Cliente A em Escala Logarítima')
lines(density(log(q)),col=2)
```

Anexo 2 – Arquivos do WinBUGS utilizados para implementação dos modelos

```
## Modelo 01
```

```
model;
{
  for(i in 1:r){
    for(j in 1:r){
      y[i,j] ~ dlnorm(mu[i,j],tau)
      mu[i,j] <- mu.y + alpha.y[i] + beta.y[j]
    }
  }
  alpha.y[1] <- 0
  beta.y[1] <- 0
  for(i in 2:r){
    alpha.y[i] ~ dnorm(0.0,tau.alpha)
    beta.y[i] ~ dnorm(0.0,tau.beta)
  }
  mu.y ~ dnorm(0.0,tau.y)
  tau ~ dgamma(0.001,0.001)
  tau.y <- 1 / 1000
  tau.alpha <- 1 / 100
  tau.beta <- 1 / 100
  for(i in 2:r){
    for(j in (r+2-i):r){
      z[i,j] <- log(y[i,j])
    }
  }
  for(i in 1:r){
    for(j in 1:(r-i+1)){
      zrep[i,j] ~ dnorm(mu[i,j],tau)
      z[i,j] <- log(y[i,j])
      d[i,j] <- zrep[i,j] - z[i,j]
      d2[i,j] <- pow(d[i,j],2)
    }
  }
}

list(tau = 0.001)
```

Dados para o exemplo 01:

```
list(r=12,
y=structure(.Data = c(
4802850.44, 17441804.59, 4187133.72, 986868.99, 162719.47, 43455.92, 54711.78, 17247.51, 1838.4, 22931.88, 1505.94, 1429.38,
5754643.14, 16393109.71, 3466015.88, 476135.99, 123730.16, 151586.32, 37286.08, 7258.55, 9053.96, 11472.11, 1118.54, NA,
6345038.34, 18993649.18, 4483928.68, 519284.86, 243013.63, 101922.59, 33312.53, 39304.75, 27304.29, 3203.13, NA, NA,
5685354.99, 18714061.71, 4045084.57, 798964.94, 119793.58, 32068.08, 16992.17, 60210.91, 4113.6, NA, NA, NA,
5232596.56, 18709616.07, 4149276.34, 751410.28, 165674.81, 273579.24, 25147.54, 15592.55, NA, NA, NA, NA,
6028734.64, 18908539.86, 4082696.08, 981156.37, 288234.9, 210038.25, 39179.13, NA, NA, NA, NA, NA,
4900277.19, 16810779.81, 3934517.91, 690947.00, 127896.57, 67192.65, NA, NA, NA, NA, NA, NA, NA,
5285268.37, 13870982.04, 3437329.48, 735041.12, 132101.02, NA, NA, NA, NA, NA, NA, NA, NA,
4218038.58, 16093191.27, 3502019.43, 1072454.57, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
4833407.81, 14054768.13, 3293114.14, NA, NA,
6341794.00, 18918359.5, NA, NA,
5566210.76, NA, NA),
.Dim=c(12,12)))
```

Saída parcial do WinBUGS para o exemplo 01:

mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
d2[1,1]	0,2772241	0,4024783	0,0025501922,472934E-40,1230293	1,411802	1,411802	20001	50000
d2[1,2]	0,2665079	0,3880416	0,0020572212,556857E-40,1178249	1,356742	1,356742	20001	50000
.
y[2,12]	1805,438	1386,523	9,518671	386,1626	1436,38	5394,258	20001 50000
y[3,11]	1791,477	1161,664	8,758651	477,3343	1507,931	4793,25	20001 50000
y[3,12]	1821,289	1391,646	10,13834	395,6578	1456,772	5438,235	20001 50000
y[4,10]	9964,162	6066,341	44,92767	2825,031	8514,333	25512,02	20001 50000
.
y[11,10]	9972,957	6120,973	49,00316	2798,619	8552,573	25685,71	20001 50000
y[11,11]	1795,049	1186,911	9,051778	472,7832	1502,532	4840,357	20001 50000
y[11,12]	1813,304	1415,854	9,352569	394,2054	1440,011	5429,083	20001 50000
y[12,2]	18666740,0	10373560,0	87680,21	5874175,0	16332870,0	45104960,0	20001 50000
y[12,3]	4354191,0	2446775,0	21392,27	1376390,0	3802373,0	10569750,0	20001 50000
y[12,4]	864729,5	486559,9	4581,318	274546,8	756583,0	2104811,0	20001 50000
y[12,5]	191061,5	108895,8	994,9057	59071,67	165892,1	472793,8	20001 50000
y[12,6]	110621,0	63971,2	550,9789	34329,21	96220,91	272167,3	20001 50000
y[12,7]	37979,89	22045,41	192,2521	11389,5	32913,39	93906,7	20001 50000
y[12,8]	24288,59	14170,11	118,5068	7289,972	21061,44	60884,96	20001 50000
y[12,9]	8125,372	4927,517	44,84365	2353,827	6973,548	20662,96	20001 50000
y[12,10]	10019,83	6196,179	49,82833	2810,307	8546,006	25938,37	20001 50000
y[12,11]	1793,669	1198,046	9,812343	469,5813	1495,99	4822,191	20001 50000
y[12,12]	1830,203	1416,952	10,38874	386,7693	1450,43	5470,471	20001 50000

Modelo 02

```

model;
{
  for (i in 1 : r) {
    for(j in 1 : r){
      y[i , j] ~ dlnorm(mu[i , j], tau)
      mu[i , j]<- mu.y + alpha.y[i] + beta.y[j]
    }
  }
  alpha.y[1]<- 0
  beta.y[1]<- 0
  for (i in 2 : r){
    alpha.y[i] ~ dnorm(0.0, tau.alpha)
    beta.y[i] ~ dnorm(0.0, tau.beta)
  }
  mu.y ~ dnorm(theta[T], tau.y)
  T ~ dcat(P[])
  P[1:2] ~ ddirch(delta[])
  delta[1]<- 1
  delta[2]<- 1
  theta[2] <- theta[1] + k
  theta[1] ~ dnorm(0.0, tau.theta)
  k ~ dnorm(0.0, 1.0E-3)|0.0,
  tau ~ dgamma(0.001 , 0.001)
  tau.y<- 1 / 1000
  tau.alpha <- 1 / 100
  tau.beta <- 1 / 100
  tau.theta<- 1/100
  for( i in 2 : r) {
    for (j in (r + 2 - i): r){
      z[i , j]<- log(y[i , j])
    }
  }
  for (i in 1 : r){
    for (j in 1 : (r - i + 1)){
      zrep[i , j] ~ dnorm(mu[i , j], tau)
      z[i , j]<- log(y[i , j])
      d[i , j] <- zrep[i , j] - z[i , j]
      d2[i , j]<- pow(d[i , j], 2)
    }
  }
}

```

```
list(mu.y = 0.0, theta=c(10, NA), tau = 0.001, k=5)
```

```
## Modelo 03
```

```
model;
{
  for( i in 1 : r ) {
    for( j in 1 : r ) {
      y[ i, j ] ~ dlnorm( mu[ i, j ], tau )
      mu[ i, j ] <- mu.y + alpha.y[ i ] + beta.y[ j ]
    }
  }
  alpha.y[1] <- 0
  beta.y[1] <- 0
  for( i in 2 : r ) {
    alpha.y[i] <- alpha.y[i-1] + h[i]
    beta.y[i] <- beta.y[i-1] + v[i]
    h[i] ~ dnorm(0.0, tau.h)
    v[i] ~ dnorm(0.0, tau.v)
  }
  mu.y ~ dnorm(0.0, tau.y)
  tau ~ dgamma(0.001, 0.001)
  tau.v ~ dgamma(0.00001, 0.00001)
  tau.h ~ dgamma(0.00001, 0.00001)
  tau.y <- 1 / 1000
  tau.alpha <- 1 / 100
  tau.beta <- 1 / 100
  for( i in 2 : r ) {
    for( j in (r + 2 - i): r){
      z[i , j]<- log(y[i , j])
    }
  }
  for( i in 1 : r){
    for( j in 1 : (r - i + 1)){
      zrep[i , j] ~ dnorm(mu[i , j], tau)
      z[i , j]<- log(y[i , j])
      d[i , j] <- zrep[i , j] - z[i , j]
      d2[i , j]<- pow(d[i , j], 2)
    }
  }
}

list(tau = 0.001, tau.v = 1, tau.h = 1)
```

Modelo 04

```

model;
{
  for (i in 1 : r) {
    for(j in 1 : r){
      y[i , j] ~ dlnorm(mu[i , j], tau)
      mu[i , j]<- mu.y + alpha.y[i] + beta.y[j]
    }
  }
  alpha.y[1]<- 0
  beta.y[1]<- 0
  for (i in 2 : r){
    alpha.y[i] <- alpha.y[i-1] + h[i]
    beta.y[i] <- beta.y[i-1] + v[i]
    h[i] ~ dnorm(0.0, tau.h)
    v[i] ~ dnorm(0.0, tau.v)
  }
  mu.y ~ dnorm(theta[T], tau.y)
  T ~ dcat(P[])
  P[1:2] ~ ddirch(delta[])
  delta[1]<- 1
  delta[2]<- 1
  theta[2] <- theta[1] + k
  theta[1] ~ dnorm(0.0, tau.theta)
  k ~ dnorm(0.0, 1.0E-3)|(0.0,)
  tau ~ dgamma(0.001 , 0.001)
  tau.v ~ dgamma(0.00001, 0.00001)
  tau.h ~ dgamma(0.00001, 0.00001)
  tau.y<- 1 / 1000
  tau.alpha <- 1 / 100
  tau.beta <- 1 / 100
  tau.theta<- 1/100
  for( i in 2 : r) {
    for (j in (r + 2 - i): r){
      z[i , j]<- log(y[i , j])
    }
  }
  for (i in 1 : r){
    for (j in 1 : (r - i + 1)){
      zrep[i , j] ~ dnorm(mu[i , j], tau)
      z[i , j]<- log(y[i , j])
      d[i , j] <- zrep[i , j] - z[i , j]
      d2[i , j]<- pow(d[i , j], 2)
    }
  }
}

list(mu.y = 0.0, theta=c(10, NA), tau = 0.001, tau.v=1, tau.h=1, k=5)

```

Anexo 3 – Distribuições Condicionais completas para cada modelo

Conforme pode ser verificado em Ntzoufras(1999), as distribuições condicionais completas para os modelos 1 e 3 (sem mistura) são dadas por:

Modelo 1:

$$f(\mu_{ij} | \cdot) = N \left(\frac{Y_{..}}{r^2 + \sigma^2 / \sigma_\mu^2}, \frac{\sigma^2}{r^2 + \sigma^2 / \sigma_\mu^2} \right)$$

$$f(\alpha_i | \cdot) = N \left(\frac{Y_{i.} - Y_{1.} - r \sum_{k \neq 1, i} \alpha_k}{2r + \sigma^2 / \sigma_{\alpha_i}^2}, \frac{\sigma^2}{2r + \sigma^2 / \sigma_{\alpha_i}^2} \right)$$

$$f(\beta_j | \cdot) = N \left(\frac{Y_{.j} - Y_{.1} - r \sum_{k \neq 1, j} \beta_k}{2r + \sigma^2 / \sigma_{\beta_i}^2}, \frac{\sigma^2}{2r + \sigma^2 / \sigma_{\beta_i}^2} \right)$$

$$f(\sigma^{-2} = \tau | \cdot) = G \left(a_\tau + \frac{r^2}{2}, b_\tau + \frac{\sum_{i=1, j=1}^r (Y_{ij} - \mu_{ij})^2}{2} \right)$$

Modelo 3:

$$f(\mu_{ij} | \cdot) = N \left(\frac{Y_{..} + r^2 \bar{\alpha}_i - r^2 \bar{\beta}_j}{r^2 + \sigma^2 / \sigma_\mu^2}, \frac{\sigma^2}{r^2 + \sigma^2 / \sigma_\mu^2} \right)$$

$$f(\alpha_i | \cdot) = N \left(\frac{Y_{i.} - r\mu_{ij} - \beta_j + (\alpha_{i+1} + \alpha_{i-1})\sigma^2 / \sigma_h^2}{r + 2\sigma^2 / \sigma_h^2}, \frac{\sigma^2}{r + 2\sigma^2 / \sigma_h^2} \right)$$

$$f(\beta_j | \cdot) = N \left(\frac{Y_{.j} - r\mu_{ij} - \alpha_i + (\beta_{j+1} + \beta_{j-1})\sigma^2 / \sigma_v^2}{r + 2\sigma^2 / \sigma_v^2}, \frac{\sigma^2}{r + 2\sigma^2 / \sigma_v^2} \right)$$

$$f(\sigma_h^2 | \cdot) = G \left(a_h + \frac{r-1}{2}, b_h + \frac{\sum_{i=2}^r (\alpha_i - \alpha_{i-1})^2}{2} \right)$$

$$f(\sigma_v^2 | \cdot) = G \left(a_v + \frac{r-1}{2}, b_v + \frac{\sum_{j=2}^r (\beta_j - \beta_{j-1})^2}{2} \right)$$

As distribuições condicionais completas para os modelos 2 e 4 são especificadas detalhadamente em Melo e Ehlers(2006), sendo denotadas por:

Modelo 2:

$$f(\mu_{ij} | \cdot) = N \left(\frac{Y_{..} - \sum_{ij} (\alpha_i + \beta_j) - n_2 \theta}{r^2 \sigma^{-2} + \sigma_{\mu}^{-2}}, (r^2 \sigma^{-2} + \sigma_{\mu}^{-2})^{-1} \right)$$

$$f(\alpha_i | \cdot) = N \left(\frac{Y_{i.} - r\mu - \sum_j \beta_j - r_2 \theta}{r \sigma^{-2} + \sigma_{\alpha}^{-2}}, (r \sigma^{-2} + \sigma_{\alpha}^{-2})^{-1} \right)$$

$$f(\beta_j | \cdot) = N \left(\frac{Y_{.j} - Y_{.1} - r \sum_{k \neq 1, j} \beta_k}{2r + \sigma^2 / \sigma_{\beta_i}^2}, \frac{\sigma^2}{2r + \sigma^2 / \sigma_{\beta_i}^2} \right)$$

$$f(\sigma^{-2} = \tau | \cdot) = G \left(a_{\tau} + \frac{r^2}{2}, b_{\tau} + \frac{\sum_{i=1, j=1}^r (Y_{ij} - \mu_{ij})^2}{2} \right)$$

$$f(p | \cdot) = \text{Beta}(n_1 + 1, n_2 + 1)$$

Modelo 4:



Anexo 4 – Gráficos para análise da convergência