

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS (CCET)
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

O Processo de Poisson Estendido e Aplicações

Luis Ernesto Bueno Salasar

Orientador: Prof. Dr. José Galvão Leite

Co-orientador: Prof. Dr. Francisco Louzada Neto

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

2007

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

S161pp

Salasar, Luis Ernesto Bueno.
O processo de Poisson estendido e aplicações / Luis
Ernesto Bueno. -- São Carlos : UFSCar, 2007.
151 f.

Dissertação (Mestrado) -- Universidade Federal de São
Carlos, 2007.

1. Processo estocástico. 2. Simulação. 3. Inferência
bayesiana. 4. Modelos discretos. I. Título.

CDD: 519.2 (20^a)

Dedicatória

Este trabalho é dedicado ao meu avô Clóvis Bueno Monteiro que completa neste ano seu centenário. Meu avô é para mim um grande exemplo de dignidade, coragem e luta em prol dos ideais de igualdade e justiça. Durante sua longa militância política, participou ativamente de momentos históricos importantes do nosso país, e sofreu na defesa de suas idéias inúmeras perseguições sendo preso pelos regimes totalitários que existiram no século passado. Apesar dessas inúmeras retaliações, sempre conservou seus ideais e conseguiu transmitir aos seus filhos esse legado de luta por uma sociedade mais justa e igualitária, os quais também tiveram uma atuação importante na construção de um mundo melhor.

*"Há aqueles que lutam um dia; e por isso são bons;
Há aqueles que lutam muitos dias; e por isso são muito bons;
Há aqueles que lutam anos; e são melhores ainda;
Porém há aqueles que lutam toda a vida; esses são os imprescindíveis"*

Bertolt Brecht

Agradecimentos

À minha família, pelo apoio, educação e exemplo que sempre me ofereceram.

Ao Professor Galvão, um exemplo de professor e orientador, que com sua paciência e excelente didática me ajudou muito na redação deste trabalho.

Ao Professor Louzada Neto pelo constante incentivo à pesquisa e por sempre me mostrar novas perspectivas e oportunidades de trabalhos acadêmicos.

Ao Professor Nelson Tanaka e ao Professor Dorival Leão Pinto, pela leitura cuidadosa deste trabalho e pelas importantes contribuições feitas.

À todos os professores do Departamento de Estatística da UFSCar, pelo acolhimento e disponibilidade que sempre tiveram comigo.

Aos colegas de mestrado, pela amizade e companheirismo.

À Tatiana Sato, por ter me ajudado nos momentos mais difíceis e pelo companheirismo nesta jornada em que caminhamos juntos.

Aos amigos de São Carlos, pela amizade, companheirismo e pelos momentos de descontração.

Resumo

Nesta dissertação veremos como o processo de Poisson estendido pode ser aplicado à construção de modelos probabilísticos discretos. Um processo de Poisson estendido é um processo estocástico a tempo contínuo com espaço de estados igual ao conjunto dos números naturais, obtido a partir de uma generalização do processo de Poisson homogêneo onde as taxas de transição dependem do estado atual do processo. A partir das taxas de transição e das equações diferenciais de Chapman-Kolmogorov pode-se determinar a distribuição de probabilidades para qualquer tempo fixado do processo. Reciprocamente, dada qualquer distribuição de probabilidades sobre o conjunto dos números naturais é possível determinar, de maneira única, uma seqüência de taxas de transição de um processo de Poisson estendido tal que, para algum instante, a distribuição unidimensional do processo coincide com a dada distribuição de probabilidades. Portanto, o processo de Poisson estendido se apresenta como uma ferramenta bastante flexível na análise de dados discretos, pois generaliza todos os modelos probabilísticos discretos.

Apresentaremos as taxas de transição dos processos de Poisson estendido que originam as distribuições de Poisson, Binomial e Binomial Negativa e determinaremos os estimadores de máxima verossimilhança, intervalos de confiança e testes de hipóteses dos parâmetros dos modelos propostos. Faremos também uma análise bayesiana destes modelos com *prioris* informativas e não informativas, apresentando os resumos a *posteriori* e comparando estes resultados com aqueles obtidos via inferência clássica.

Palavras-chave: processos estocásticos, processo de Poisson estendido, equações diferenciais de Chapman-Kolmogorov, modelos discretos, distribuição de Poisson, Binomial e Binomial Negativa, estimadores de máxima verossimilhança, intervalos de confiança e teste de hipóteses dos parâmetros.

Abstract

In this dissertation we will study how extended Poisson process can be applied to construct discrete probabilistic models. An Extended Poisson Process is a continuous time stochastic process with the state space being the natural numbers, it is obtained as a generalization of homogeneous Poisson process where transition rates depend on the current state of the process. From its transition rates and Chapman-Kolmogorov differential equations, we can determine the probability distribution at any fixed time of the process. Conversely, given any probability distribution on the natural numbers, it is possible to determine uniquely a sequence of transition rates of an extended Poisson process such that, for some instant, the unidimensional probability distribution coincides with the provided probability distribution. Therefore, we can conclude that extended Poisson process is as a very flexible framework on the analysis of discrete data, since it generalizes all probabilistic discrete models.

We will present transition rates of extended Poisson process which generate Poisson, Binomial and Negative Binomial distributions and determine maximum likelihood estimators, confidence intervals, and hypothesis tests for parameters of the proposed models. We will also perform a bayesian analysis of such models with informative and noninformative *prioris*, presenting *posteriori* summaries and comparing these results to those obtained by means of classic inference.

Key-words: stochastic processes; extended Poisson process; Chapman-Kolmogorov differential equations; discrete models; Poisson, Binomial and Negative Binomial distributions; maximum likelihood estimators, confidence intervals and hypothesis test of parameters.

Sumário

1	Introdução	1
1.1	Cadeias de Markov a Tempo Contínuo	3
1.2	Processo de Poisson Homogêneo	6
2	Processo de Poisson Estendido	11
2.1	Noções, Definições e Exemplos	11
2.2	Distribuição do Tempo de Espera entre Ocorrências	15
2.3	Processos de Nascimento Divergentes	17
2.4	Distribuições Discretas derivadas de Processos de Poisson Estendido	19
2.4.1	Distribuição de Poisson	20
2.4.2	Distribuição Binomial Negativa	21
2.4.3	Distribuição Binomial	23
2.5	Estudo da Dispersão do Processo de Poisson Estendido	25
2.6	Determinação das Distribuições de Probabilidades Unidimensionais do Processo	29
3	Modelo Estatístico e Estimação de Máxima Verossimilhança	32
3.1	Função de Verossimilhança	32
3.2	Formas Paramétricas para as Taxas de Transição	33
3.3	Aplicação: Dados sobre Abundância de Animais	34
3.4	Comparação com outros Modelos	38
4	Generalização da Distribuição de Poisson para Dados com Inflação ou Deflação de Zeros	41
4.1	Modelo Probabilístico	41
4.2	Função de Verossimilhança	48

4.3	Análise de Dados	49
5	Generalização da Distribuição Binomial Negativa para Dados com Inflação ou Deflação de Zeros	53
5.1	Modelo Probabilístico	53
5.2	Função de Verossimilhança	58
5.3	Análise de Dados	59
6	Estudo de Simulação	64
6.1	Metodologia	64
6.2	Resultados das Simulações para o Modelo PPE2	67
6.2.1	Probabilidades de Cobertura	67
6.2.2	Taxas de Rejeição para o Teste de Hipóteses	69
6.3	Resultados das Simulações para o Modelo PPE4	69
6.3.1	Probabilidades de Cobertura	69
6.3.2	Taxas de Rejeição dos Testes de Hipóteses	71
7	Inferência Bayesiana	73
7.1	Modelo PPE2	73
7.1.1	Prioris Não Informativas	75
7.1.2	Prioris Informativas	80
7.2	Modelo PPE4	81
7.2.1	Prioris Não Informativas	83
7.2.2	Prioris Informativas	89
8	Considerações Finais	91
	Referências Bibliográficas	94
	Apêndices	96
A	Rotinas do Capítulo 3	96
B	Rotinas do Capítulo 4	107

C Rotinas do Capítulo 5	111
D Rotinas do Capítulo 6	116
D.1 Modelo PPE2 - Probabilidade de Cobertura	116
D.2 Modelo PPE2 - Teste de Hipóteses	119
D.3 Modelo PPE2 - Decaimento da Variância	122
D.4 Modelo PPE4 - Probabilidade de Cobertura	125
D.5 Modelo PPE4 - Teste de Hipóteses	130
D.6 Modelo PPE4 - Decaimento da Variância	132
E Rotinas do Capítulo 7	137
E.1 PPE2 - Prioris Não Informativas	137
E.2 PPE2 - Prioris Informativas	140
E.3 PPE4 - Prioris Não Informativas	143
E.4 PPE4 - Prioris Informativas	147

Capítulo 1

Introdução

Dados de contagem surgem nos mais variados campos do conhecimento, tais como na biologia, medicina, ciências sociais e engenharia. Essa demanda fez com que surgisse uma variedade de métodos para a análise deste tipo de dados, destacando-se os modelos de Poisson, Binomial e Binomial Negativa.

Neste trabalho mostraremos como modelos probabilísticos discretos são originados a partir de um processo estocástico a tempo contínuo $\{X(t); t \geq 0\}$, que denominaremos de Processo de Poisson Estendido. Esse processo consiste numa generalização do Processo de Poisson Homogêneo onde a função de transição depende do estado atual do processo. Neste contexto, apresentaremos formas paramétricas das taxas de transição que originam as distribuições de Poisson, Binomial e Binomial Negativa. Serão propostas também formas paramétricas que generalizam os modelos anteriores para o caso de inflação ou deflação de zeros nos dados. Para estes modelos, discutiremos como é feita a inferência e os testes de hipóteses sobre os parâmetros.

Antes de introduzir os conceitos relacionados ao Processo de Poisson Estendido, faremos uma breve introdução de alguns conceitos básicos sobre as cadeias de Markov a tempo contínuo e sobre o Processo de Poisson Homogêneo, nas seções 1.1 e 1.2, respectivamente.

No Capítulo 2 definiremos o Processo de Poisson Estendido (seção 2.1), apresentaremos as principais características das distribuições do tempo de espera entre ocorrências do mesmo (seção 2.2) e uma condição necessária para obtermos um processo honesto, isto é, um processo não explosivo (seção 2.3). Na seção 2.4 mostraremos como são obtidas as seqüências das taxas de transição que originam as distribuições de Poisson, Binomial

e Binomial Negativa e, na seção 2.5, discutiremos como o comportamento da seqüência das taxas influencia a variabilidade do processo. Além disso, na seção 2.6 mostraremos como se determina a distribuição de probabilidades unidimensionais do processo a partir da seqüência das taxas de transição.

No Capítulo 3 apresentaremos o modelo estatístico e o método de estimação de máxima verossimilhança para os parâmetros que definem as taxas de transição do processo. Na seção 3.1 determinaremos a função de verossimilhança e proporemos na seção 3.2 algumas formas paramétricas para as taxas de transição. Na seção 3.3 faremos uma análise de dados da contagem de *Macoma lilliana* (pequeno molusco), disponíveis em Faddy (1997) e obteremos as estimativas de máxima verossimilhança para os parâmetros dos modelos propostos na seção 3.2. Na seção 3.4 faremos uma comparação de alguns modelos propostos baseados em Processo de Poisson Estendido com modelos usuais para dados com excesso de zeros e uns.

No Capítulo 4 será discutido um modelo baseado em um Processo de Poisson Estendido que generaliza a distribuição de Poisson e que tem bom ajuste quando há uma quantidade discrepante de zeros na amostra. Na seção 4.1 mostraremos como são obtidas a distribuição de probabilidades, média, variância e função geradora de probabilidades das variáveis do processo e, na seção 4.2, será apresentada a função de verossimilhança para o modelo em questão. Na seção 4.3 o modelo é ajustado a dados da contagem do número de acidentes entre operárias estudados por Greenwood & Yule (1920), sendo determinados as estimativas de máxima verossimilhança, os intervalos de confiança para os parâmetros envolvidos, e realizado um teste para verificar a adequação da distribuição de Poisson aos dados.

No Capítulo 5 será estudado um modelo baseado em um Processo de Poisson Estendido que generaliza a distribuição Binomial Negativa para dados com quantidade discrepante de zeros. Na seção 5.1 mostraremos como são obtidas a distribuição de probabilidades, média, variância e função geradora de probabilidades das variáveis do processo e, na seção 5.2, apresentaremos a função de verossimilhança para este modelo. Na seção 5.3 o modelo é ajustado a dados de contagem do número de visitas ao dentista, no período de um ano, disponível em Melkersson & Olsson (1999), sendo determinadas as estimativas de máxima verossimilhança, os intervalos de confiança para os parâmetros envolvidos, e realizado um

teste para verificar a adequação da distribuição Binomial Negativa a esses dados.

No Capítulo 6 será realizado um estudo de simulação para a verificação da eficiência dos estimadores intervalares assintóticos obtidos nos Capítulos 4 e 5. Esta simulação consistirá no cálculo das proporções dos intervalos de confiança que contém o valor verdadeiro do parâmetro, fixado previamente, para diferentes tamanhos amostrais.

No Capítulo 7 os modelos apresentados nos Capítulos 4 e 5 serão analisados sob o enfoque bayesiano, onde serão adotadas *prioris* normais independentes para cada um dos parâmetros dos modelos em questão. Além disso, será realizado um estudo de sensibilidade para verificar o quanto a variância das *prioris* influencia nas estimativas *a posteriori* obtidas.

1.1 Cadeias de Markov a Tempo Contínuo

Considere S um conjunto discreto e Ω um conjunto não vazio. Suponhamos que, para cada $t \in \mathbb{R}_+ = [0, \infty)$, a função $\omega \mapsto X(t)(\omega) \in S$, $\omega \in \Omega$, define uma variável aleatória. A família $\{X(t); t \in \mathbb{R}_+\}$ se diz uma cadeia (processo estocástico) a tempo contínuo e com espaço de estados S . Para cada $\omega \in \Omega$ fixado, a função $t \mapsto X(t)(\omega) \in S$, $t \in \mathbb{R}_+$, se diz a trajetória (função amostral) da cadeia. Se interpretarmos t como sendo o "tempo", então $X(t)$ representa o "estado da cadeia no instante t ".

As cadeias de Markov podem ser descritas, de maneira geral, como processos estocásticos para os quais dado o passado e o presente do processo, o estado futuro depende apenas do presente do processo. Esta característica é chamada de *Propriedade Markoviana*. Ligações telefônicas, desintegrações radioativas e quebras em cromossomos ao longo do tempo são exemplos de situações que podem ser modeladas por processos que satisfazem a esta propriedade. Diante disso, temos a seguinte definição [Çinlar, 1975].

Definição 1.1.1 *Um processo estocástico $\{X(t); t \in \mathbb{R}_+\}$ se diz uma cadeia de Markov a tempo contínuo e com espaço de estados S se*

- (i) $X(t)$ assume valores em um conjunto discreto S para todo $t \in \mathbb{R}_+$;

(ii) para quaisquer $t, s \geq 0$ e $j \in S$, tivermos

$$P\{X(t+s) = j \mid X(u); u \leq t\} = P\{X(t+s) = j \mid X(t)\}. \quad (1.1)$$

A propriedade (ii) da definição anterior é a chamada *Propriedade Markoviana*. Ela afirma que dado o conhecimento do processo até o instante (tempo) t , $t > 0$, a distribuição de $X(t+s)$ só depende de $X(t)$. Como essa probabilidade condicional só depende de t e s , denotamos por

$$p_{ij}(t, t+s) = P\{X(t+s) = j \mid X(t) = i\},$$

para quaisquer $i, j \in S$. A probabilidade $p_{ij}(t, t+s)$ se diz a probabilidade de transição do estado i no tempo t , para o estado j no tempo $t+s$.

Definição 1.1.2 *Uma cadeia de Markov $\{X(t); t \geq 0\}$ a tempo contínuo e com espaço de estados S se diz homogênea ou com probabilidades de transição estacionárias se, para todo t_1, t_2 e $s > 0$, tivermos $p_{ij}(t_1, t_1+s) = p_{ij}(t_2, t_2+s)$. Neste caso, usaremos a notação $p_{ij}(s) = p_{ij}(t, t+s)$, qualquer que seja $t > 0$ e $s > 0$.*

Doravante vamos considerar somente cadeias de Markov homogêneas com espaço de estados $S = \mathbb{N} = \{0, 1, \dots\}$. Para a caracterização de uma cadeia de Markov $\{X(t), t \in \mathbb{R}_+\}$ é necessário apenas conhecer sua distribuição inicial $p_i(0) = P\{X(0) = i\}$, para todo $i \in \mathbb{N}$, e as probabilidades de transição $p_{jk}(t)$, para todo $j, k \in \mathbb{N}$, como mostra o teorema abaixo.

Teorema 1.1.3 *Para quaisquer instantes $0 = t_0 < t_1 < \dots < t_n$ em $[0, +\infty)$ e quaisquer estados i_1, \dots, i_n em \mathbb{N} , temos*

$$P\{X(t_1) = i_1, \dots, X(t_n) = i_n\} = \sum_{i_0=0}^{\infty} p_{i_0}(0) p_{i_0, i_1}(t_1 - t_0) \cdots p_{i_{n-1}, i_n}(t_n - t_{n-1}).$$

Prova

Note que é suficiente mostrar que

$$P\{X(t_1) = i_1, \dots, X(t_n) = i_n \mid X(0) = i_0\} = p_{i_0, i_1}(t_1 - t_0) \cdots p_{i_{n-1}, i_n}(t_n - t_{n-1}),$$

para qualquer $i_0 \in \mathbb{N}$.

Observe que

$$\begin{aligned} P\{X(t_1) = i_1, \dots, X(t_n) = i_n \mid X(0) = i_0\} = \\ P\{X(t_n) = i_n \mid X(0) = i_0, \dots, X(t_{n-1}) = i_{n-1}\} P\{X(t_1) = i_1, \dots, X(t_{n-1}) = i_{n-1} \mid X(0) = i_0\}. \end{aligned}$$

Portanto, segue da propriedade Markoviana 1.1 que

$$P\{X(t_1) = i_1, \dots, X(t_n) = i_n \mid X(0) = i_0\} = p_{i_{n-1}, i_n}(t_n - t_{n-1}) P\{X(t_1) = i_1, \dots, X(t_{n-1}) = i_{n-1} \mid X(0) = i_0\}.$$

Aplicando recursivamente este raciocínio, provamos a igualdade desejada. ■

Uma relação fundamental com relação às probabilidades de transição de uma cadeia de Markov homogênea é a identidade de Chapman-Kolmogorov

$$p_{ik}(t+h) = \sum_{j=0}^{\infty} p_{ij}(t) p_{jk}(h), \quad (1.2)$$

para quaisquer $i, k \in \mathbb{N}$ e quaisquer $t \geq 0$ e $h \geq 0$.

Para mostrar tal identidade escrevemos o primeiro membro de 1.2 como

$$\begin{aligned} p_{ik}(t+h) &= P\{X(t+h) = k \mid X(0) = i\} \\ &= \sum_{j=0}^{\infty} P\{X(t+h) = k \mid X(0) = i, X(t) = j\} P\{X(t) = j \mid X(0) = i\}, \end{aligned}$$

o que implica pela propriedade Markoviana 1.1 que

$$\begin{aligned} p_{ik}(t+h) &= \sum_{j=0}^{\infty} P\{X(t+h) = k \mid X(t) = j\} P\{X(t) = j \mid X(0) = i\} \\ &= \sum_{j=0}^{\infty} p_{ij}(t) p_{jk}(h). \end{aligned}$$

A identidade de Chapman-Kolmogorov será útil na obtenção de equações diferenciais, cujas soluções serão as probabilidades de transição do processo.

Definição 1.1.4 *Uma característica importante das cadeias de Markov é a chamada Pro-*

propriedade Forte de Markov:

$$P\{X(T+h) = j \mid [X(t) : t < T, X(T) = i]\} = p_{ij}(h), \quad (1.3)$$

onde que T é uma regra de parada, isto é, T é uma variável aleatória tal que para todo $t > 0$ o evento $[T \leq t]$ só depende de $\{X(s); s \leq t\}$.

Note que esta propriedade é uma generalização da propriedade Markoviana dada em 1.1.

1.2 Processo de Poisson Homogêneo

Suponha que $X(t)$ seja o número de ocorrências de um evento até o instante t , $t > 0$. O interesse é obter a distribuição de probabilidades de $X(t)$ a partir de algumas suposições.

Suponha $X(0) = 0$, isto é, $P\{X(0) = 0\} = 1$ e defina

$$p_n(t) = P\{X(t) = n\} = P\{X(t+s) - X(s) = n\},$$

para todo $n \in \mathbb{N}$ e quaisquer $s \geq 0$ e $t > 0$.

O Processo de Poisson é um processo de contagem em que as ocorrências do evento não são simultâneas e uma vez que o processo está no estado n , a única transição possível é para o estado $n + 1$.

Considere um intervalo de tempo de comprimento 1. Particionemos este intervalo em M subintervalos disjuntos de comprimento $h = 1/M$, onde M é suficientemente grande para que em cada subintervalo haja no máximo uma ocorrência do evento. Portanto, a probabilidade de uma ocorrência do evento em qualquer subintervalo é igual a $1 - p_0(h)$ e, supondo que os números de ocorrências do evento nos subintervalos sejam independentes, segue que o número esperado de subintervalos contendo apenas um evento é $\frac{1-p_0(h)}{h}$ (note que $M = 1/h$). Intuitivamente podemos esperar que, quando $h \rightarrow 0$, este número convirja para o número esperado de ocorrências num intervalo de tempo de comprimento 1. Portanto, quando $h \rightarrow 0$, existe $\lambda > 0$ tal que

$$\frac{1 - p_0(h)}{h} \rightarrow \lambda. \quad (1.4)$$

Da mesma maneira, o número esperado de subintervalos com duas ou mais ocorrências tende a zero quando h diminui, ou seja, quando $h \rightarrow 0$ temos

$$\frac{1 - p_0(h) - p_1(h)}{h} \rightarrow 0. \quad (1.5)$$

Baseado nessas idéias, apresentamos a definição a seguir.

Definição 1.2.1 *Um processo estocástico $\{X(t); t \geq 0\}$ com espaço de estados $\mathbb{N} = \{0, 1, \dots\}$ é um processo de Poisson homogêneo com parâmetro (taxa) λ , $\lambda > 0$, se valerem as condições*

- (i) $X(0) = 0$;
- (ii) se $s < t$, então $X(s) \leq X(t)$;
- (iii) o processo tem incrementos independentes e estacionários, isto é, se $0 < t_1 < t_2 < \dots < t_n$, $n \geq 2$, então $X(t_1), X(t_2) - X(t_1), \dots, X(t_n) - X(t_{n-1})$ são variáveis independentes e para quaisquer s_1, s_2 e t , tais que $s_1 \geq 0, s_2 \geq 0$ e $t > 0$, as variáveis aleatórias $X(s_1 + t) - X(s_1)$ e $X(s_2 + t) - X(s_2)$ são identicamente distribuídas;
- (iv) $P\{X(t+h) - X(t) = 1\} = \lambda h + o(h)$, para quaisquer t e h , tais que $t \geq 0, h > 0$;
- (v) $P\{X(t+h) - X(t) > 1\} = o(h)$, para quaisquer t e h , tais que $t \geq 0, h > 0$;

Observações

1. Na definição acima $o(h)$ denota uma função real de h que decresce para zero mais rápido que h , isto é,

$$\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0.$$

2. Naturalmente, $X(t)$ pode ser interpretado como o número de ocorrências de um evento durante o intervalo de tempo $[0, t]$, $t > 0$. A condição (iii) implica que o processo é uma cadeia de Markov a tempo contínuo homogênea. De fato, para todo $n \geq 2$, quaisquer $0 = t_0 < t_1 < t_2 < \dots < t_n$ e quaisquer $0 \leq i_1 \leq i_2 \leq \dots \leq i_n$ em

\mathbb{N} , tem-se

$$\begin{aligned}
& P\{X(t_n) = i_n \mid X(t_1) = i_1, \dots, X(t_{n-1}) = i_{n-1}\} \\
&= P\{X(t_n) - X(t_{n-1}) = i_n - i_{n-1} \mid X(t_1) = i_1, X(t_2) - X(t_1) = i_2 - i_1, \dots, \\
&\quad X(t_{n-1}) - X(t_{n-2}) = i_{n-1} - i_{n-2}\} \\
&= P\{X(t_n) - X(t_{n-1}) = i_n - i_{n-1}\} \\
&= P\{X(t_n) - X(t_{n-1}) = i_n - i_{n-1} \mid X(t_{n-1}) = i_{n-1}\} \\
&= P\{X(t_n) = i_n \mid X(t_{n-1}) = i_{n-1}\}.
\end{aligned}$$

Por outro lado, para quaisquer $t_1, t_2, s > 0$ e quaisquer $i, j \in \mathbb{N}$ com $i \leq j$, tem-se

$$\begin{aligned}
P\{X(t_1 + s) = j \mid X(t_1) = i\} &= P\{X(t_1 + s) - X(t_1) = j - i\} \\
&= P\{X(t_2 + s) - X(t_2) = j - i\} \\
&= P\{X(t_2 + s) = j \mid X(t_2) = i\}
\end{aligned}$$

A partir das condições da definição 1.2.1 pode-se obter a distribuição de probabilidades de $X(t)$, $t > 0$. Observe que

$$p_n(t) = P\{X(t) = n\} = P\{X(t + s) - X(s) = n\} = P\{X(t) = n \mid X(0) = 0\} = p_{0n}(t),$$

para qualquer $t, t > 0$.

Logo, utilizando as propriedades (iii), (iv) e (v) determina-se a probabilidade $p_n(t+h)$. Para $n \geq 1$, pode haver n ocorrências do evento até o tempo $t+h$ de três maneiras mutuamente exclusivas.

Primeiro, no tempo t o processo se encontra no estado n e não há ocorrências entre t e $t+h$, o que tem probabilidade igual a

$$p_n(t)p_0(h) = p_n(t)[1 - \lambda h] + o(h).$$

Segundo, no tempo t o processo está no estado $n-1$ e há uma ocorrência entre t e

$t + h$, que tem probabilidade

$$p_{n-1}(t)p_1(h) = p_{n-1}(t)[\lambda h] + o(h).$$

A terceira possibilidade é que o processo esteja em um estado menor que $n - 1$ no tempo t e haja duas ou mais ocorrências entre o tempo t e $t + h$, o que tem probabilidade $o(h)$. Portanto, somando-se estas três probabilidades, conclui-se que

$$p_n(t + h) = p_n(t)[1 - \lambda h] + p_{n-1}(t)[\lambda h] + o(h),$$

o que implica que

$$\frac{p_n(t + h) - p_n(t)}{h} = -\lambda p_n(t) + \lambda p_{n-1}(t) + \frac{o(h)}{h}.$$

Passando ao limite quando $h \downarrow 0$, obtém-se a seguinte equação diferencial para $n \geq 1$:

$$p'_n(t) = -\lambda p_n(t) + \lambda p_{n-1}(t). \quad (1.6)$$

Para $n = 0$, temos que não há ocorrências até o tempo $t + h$ de uma única maneira, ou seja, não há ocorrências nos intervalos $(0, t]$ e $(t, t + h]$. Portanto,

$$p_0(t + h) = p_0(t)p_0(h) = p_0(t)[1 - \lambda h] + o(h).$$

Logo,

$$\frac{p_0(t + h) - p_0(t)}{h} = -\lambda p_0(t) + \frac{o(h)}{h},$$

o que implica

$$p'_0(t) = -\lambda p_0(t). \quad (1.7)$$

O sistema de equações diferenciais 1.6 e 1.7 é chamado sistema de equações diferenciais de Chapman-Kolmogorov para frente, que pode ser resolvido recursivamente usando as condições iniciais $p_0(0) = 1$ e $p_n(0) = 0$ para $n \geq 1$, obtidas de (i) da definição 1.2.1. Então, resolvendo este sistema tem-se que a variável aleatória $X(t)$ tem distribuição de

Poisson com parâmetro λt , isto é,

$$p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n \in \mathbb{N}.$$

Vimos que o Processo de Poisson Homogêneo $\{X(t); t \geq 0\}$ é uma cadeia de Markov homogênea a tempo contínuo, com distribuição inicial degenerada no zero ($P\{X(0) = 0\} = 1$). As probabilidades de transição são dadas por

$$\begin{aligned} p_{jk}(t) &= P\{X(s+t) = k \mid X(s) = j\} \\ &= \frac{P\{X(s+t) = k, X(s) = j\}}{P\{X(s) = j\}} \\ &= \frac{P\{X(s+t) - X(s) = k - j, X(s) = j\}}{P\{X(s) = j\}} \\ &= P\{X(s+t) - X(s) = k - j\} \\ &= P\{X(t) = k - j\} \\ &= \frac{(\lambda t)^{k-j} e^{-\lambda t}}{(k-j)!}, \end{aligned}$$

para quaisquer $s, t, s \geq 0, t > 0$ e quaisquer $j, k, j \leq k$.

Capítulo 2

Processo de Poisson Estendido

2.1 Noções, Definições e Exemplos

Uma generalização do Processo de Poisson Homogêneo é obtida neste capítulo com a suposição de que a taxa de transição do processo seja uma função do estado em que ele se encontra. Tal generalização é formalizada na seguinte definição.

Definição 2.1.1 *Diremos que uma cadeia de Markov homogênea a tempo contínuo $\{X(t); t \geq 0\}$ com espaço de estados \mathbb{N} é um Processo de Poisson Estendido (PPE) se valerem as condições*

- (i) $X(0) = 0$;
- (ii) se $s < t$ então $X(s) \leq X(t)$;
- (iii) $P\{X(t+h) = n+1 | X(t) = n\} = \lambda(n)h + o(h)$, para quaisquer t, h e $n, t \geq 0, h > 0$ e $n \in \mathbb{N}$;
- (iv) $P\{X(t+h) > n+1 | X(t) = n\} = o(h)$, para quaisquer t, h e $n, t \geq 0, h > 0$ e $n \in \mathbb{N}$.

As probabilidades de transição da cadeia são dadas por

$$p_{in}(t) = P\{X(s+t) = n | X(s) = i\},$$

para quaisquer s, t, i e $n, s \geq 0, t > 0, i, n \in \mathbb{N}, i \leq n$.

Usando a identidade de Chapman-Kolmogorov 1.2 e a definição 2.1.1, obtém-se para $n = i$,

$$\begin{aligned} p_{ii}(t+h) &= p_{ii}(t)p_{ii}(h) \\ &= p_{ii}(t)[1 - \lambda(i)h] + o(h) \\ &= p_{ii}(t) - \lambda(i)p_{ii}(t)h + o(h), \end{aligned}$$

o que implica

$$\frac{p_{ii}(t+h) - p_{ii}(t)}{h} = -\lambda(i)p_{ii}(t) + \frac{o(h)}{h}.$$

Passando ao limite para $h \downarrow 0$, obtém-se a equação diferencial

$$p'_{ii}(t) = -\lambda(i)p_{ii}(t), \quad (2.1)$$

que combinada à condição inicial $p_{ii}(0) = 1$, fornece a solução

$$p_{ii}(t) = e^{-\lambda(i)t}. \quad (2.2)$$

Novamente, usando a identidade de Chapman-Kolmogorov 1.2 e a definição 2.1.1, obtém-se para $n > i$,

$$\begin{aligned} p_{in}(t+h) &= p_{i,n-1}(t)p_{n-1,n}(h) + p_{in}(t)p_{nn}(h) + o(h) \\ &= [\lambda(n-1)h]p_{i,n-1}(t) + [1 - \lambda(n)h]p_{in}(t) + o(h), \end{aligned}$$

o que implica que

$$\frac{p_{in}(t+h) - p_{in}(t)}{h} = -\lambda(n)p_{in}(t) + \lambda(n-1)p_{i,n-1}(t) + \frac{o(h)}{h}.$$

Passando ao limite para $h \downarrow 0$, obtém-se a equação diferencial

$$p'_{in}(t) = -\lambda(n)p_{in}(t) + \lambda(n-1)p_{i,n-1}(t), \quad (2.3)$$

que combinada à condição inicial $p_{in}(0) = 0$ para $n > i$, fornece a seguinte solução recursiva:

$$p_{in}(t) = \lambda(n-1)e^{-\lambda(n)t} \int_0^t e^{\lambda(n)s} p_{i,n-1}(s) ds. \quad (2.4)$$

Exemplo 2.1.2 *Transmutação radioativa.* Um átomo radioativo, por exemplo urânio, pode mudar para um átomo de outro tipo pela emissão de partículas ou raios-gama. Cada tipo de átomo representa um estado do processo. No decorrer do tempo teremos uma sucessão de estados $E_0 \rightarrow E_1 \rightarrow E_2 \rightarrow \dots \rightarrow E_m$. De acordo com teorias físicas aceitas, a probabilidade de uma transição $E_n \rightarrow E_{n+1}$ é constante enquanto o átomo está no estado E_n . Portanto, as equações diferenciais de Chapman-Kolmogorov 2.1 e 2.3 descrevem a dinâmica deste processo. Se E_m for o estado final do processo devemos tomar $\lambda(m) = 0$, tornando nula a probabilidade de uma transição $E_m \rightarrow E_{m+1}$.

Exemplo 2.1.3 *O processo de Yule.* O processo de Yule não é propriamente um PPE porque seu estado inicial é diferente de zero. Contudo as condições (ii), (iii) e (iv) da definição 2.1.1 são satisfeitas. Considere uma população cujos elementos podem dar origem a novos elementos, mas não podem morrer. Assuma que durante qualquer intervalo de tempo suficientemente pequeno de comprimento h , cada elemento pode dar origem a um novo elemento, dois ou mais novos elementos ou a nenhum elemento com probabilidades $\lambda h + o(h)$, $o(h)$ e $1 - \lambda h + o(h)$, respectivamente. Se não houver interação entre os elementos e se no instante t o tamanho populacional for n , então a probabilidade de que ocorra um nascimento, dois ou mais nascimentos ou nenhum nascimento no intervalo de tempo entre os instantes t e $t + h$ é igual a $n\lambda h + o(h)$, $o(h)$ e $1 - n\lambda h + o(h)$, respectivamente. Logo, a probabilidade condicional do tamanho populacional ser igual a n no instante t , dado que no instante inicial há i elementos na população, satisfaz as equações 2.1 e 2.3 com $\lambda(n) = \lambda n$, isto é,

$$\begin{aligned} p'_{in}(t) &= -n\lambda p_{in}(t) + (n-1)\lambda p_{i,n-1}(t), & n > i, \\ p'_{ii}(t) &= -i\lambda p_{ii}(t). \end{aligned} \quad (2.5)$$

Portanto, resolvendo recursivamente o sistema diferencial 2.5 sujeito às condições iniciais

$p_{ii}(0) = 1$ e $p_{in}(0) = 0$ para $n > i$, conclui-se que para $n \geq i > 0$ vale

$$p_{in}(t) = \binom{n-1}{n-i} (e^{-\lambda t})^i (1 - e^{-\lambda t})^{n-i}.$$

Isto é, se denotarmos por $X(t)$ o tamanho populacional no instante t , então a distribuição de probabilidades condicional de $X(t)$, dado $X(0) = i$, é Binomial Negativa com parâmetros i e $e^{-\lambda t}$.

Seja $\{X(t); t \geq 0\}$ um PPE e denotemos por

$$p_n(t) = P\{X(t) = n\} = P\{X(t) = n | X(0) = 0\} = p_{0n}(t), \quad n \in \mathbb{N}.$$

Então, para a distribuição de probabilidades $\{p_n(t); n \in \mathbb{N}\}$ de $X(t)$, as equações 2.1 e 2.3 são escritas como

$$p'_0(t) = -\lambda(0)p_0(t), \quad (2.6)$$

$$p'_n(t) = -\lambda(n)p_n(t) + \lambda(n-1)p_{n-1}(t), \quad n = 1, 2, \dots \quad (2.7)$$

Lembrando que $p_0(0) = 1$ e $p_n(0) = 0$ para $n > 0$, obtém-se a seguinte solução recursiva para este sistema de equações diferenciais:

$$p_0(t) = e^{-\lambda(0)t}, \quad (2.8)$$

$$p_n(t) = \lambda(n-1)e^{-\lambda(n)t} \int_0^t e^{\lambda(n)s} p_{n-1}(s) ds, \quad n \geq 1. \quad (2.9)$$

2.2 Distribuição do Tempo de Espera entre Ocorrências

Suponha $\{X(t) : t \geq 0\}$ um PPE. Seja

$$\begin{aligned} T_0 &= \inf\{s > 0 : X(s) = 1\}, \\ T_n &= \inf\{s > 0 : X(s) = n + 1\} - \sum_{i=0}^{n-1} T_i(\omega), \quad n = 1, 2, \dots \end{aligned}$$

A variável aleatória T_n é o tempo de permanência do processo no estado n , $n \in \mathbb{N}$.

Teorema 2.2.1 *Se não houver estados absorventes nem instantâneos do processo, isto é, $0 < \lambda(n) < \infty$, $n \in \mathbb{N}$, então para todo $u_n > 0$,*

$$P\{T_n > u_n | T_0, \dots, T_{n-1}\} = e^{-\lambda(n)u_n},$$

implicando que T_n tem distribuição exponencial com taxa $\lambda(n)$.

Prova

$$P\{T_n > u_n | T_0, \dots, T_{n-1}\} = P\{T_n > u_n | [X(t) : t < T_0 + \dots + T_{n-1}, X(T_0 + \dots + T_{n-1}) = n]\} =$$

$$P\{X(T_0 + \dots + T_{n-1} + u_n) = n | [X(t) : t < T_0 + \dots + T_{n-1}, X(T_0 + \dots + T_{n-1}) = n]\}.$$

Como $T_0 + \dots + T_{n-1}$ é uma regra de parada, isto é, o evento $[T_0 + \dots + T_{n-1} \leq t]$ só depende de $\{X(s) : s \leq t\}$, conclui-se usando a propriedade forte de Markov 1.3, que

$$P\{T_n > u_n | T_0, \dots, T_{n-1}\} = p_{nn}(u_n).$$

Portanto, usando 2.2 tem-se que

$$P\{T_n > u_n | T_0, \dots, T_{n-1}\} = e^{-\lambda(n)u_n}. \quad (2.10)$$

Por outro lado,

$$\begin{aligned} P\{T_n > u_n\} &= \int_0^\infty \cdots \int_0^\infty P\{T_n > u_n | T_0 = t_0, T_1 = t_1, \dots, T_{n-1} = t_{n-1}\} \cdot f(t_0, \dots, t_{n-1}) dt_0 \cdots dt_{n-1} \\ &= e^{-\lambda(n)u_n}, \end{aligned}$$

onde f representa a função densidade de probabilidade de (T_0, \dots, T_{n-1}) . ■

Corolário 2.2.2 *Com a mesma condição do teorema anterior, a função densidade conjunta de T_0, T_1, \dots, T_n , $n \geq 1$, é dada por*

$$\begin{aligned} f_{(T_0, \dots, T_n)}(t_0, t_1, \dots, t_n) &= \prod_{j=0}^n \lambda(j) e^{-\lambda(j)t_j} \\ &= \prod_{j=0}^n f_{T_j}(t_j). \end{aligned} \quad (2.11)$$

Prova

Note que a densidade conjunta de T_0, T_1, \dots, T_n é obtida por

$$\begin{aligned} f_{(T_0, \dots, T_n)}(t_0, t_1, \dots, t_n) &= \frac{\partial}{\partial t_0} \cdots \frac{\partial}{\partial t_n} P\{T_0 \leq t_0, \dots, T_n \leq t_n\} \\ &= \frac{\partial}{\partial t_0} \cdots \frac{\partial}{\partial t_n} P\{T_n \leq t_n | T_0 \leq t_0, \dots, T_{n-1} \leq t_{n-1}\} \cdot \\ &\quad P\{T_0 \leq t_0, \dots, T_{n-1} \leq t_{n-1}\} \\ &= \frac{\partial}{\partial t_0} \cdots \frac{\partial}{\partial t_n} [1 - P\{T_n > t_n | T_0 \leq t_0, \dots, T_{n-1} \leq t_{n-1}\}] \cdot \\ &\quad P\{T_0 \leq t_0, \dots, T_{n-1} \leq t_{n-1}\}, \end{aligned}$$

o que implica por 2.10 que

$$\begin{aligned} f_{(T_0, \dots, T_n)}(t_0, t_1, \dots, t_n) &= \frac{\partial}{\partial t_0} \cdots \frac{\partial}{\partial t_n} [1 - e^{-\lambda(n)t_n}] P\{T_0 \leq t_0, \dots, T_{n-1} \leq t_{n-1}\} \\ &= \lambda(n) e^{-\lambda(n)t_n} \frac{\partial}{\partial t_0} \cdots \frac{\partial}{\partial t_{n-1}} P\{T_0 \leq t_0, \dots, T_{n-1} \leq t_{n-1}\}. \end{aligned}$$

Repetindo este raciocínio recursivamente, demonstra-se 2.11. ■

De 2.11 segue que $\{T_n; n \in \mathbb{N}\}$ é uma seqüência de variáveis aleatórias independentes.

2.3 Processos de Nascimento Divergentes

O sistema de equações diferenciais de Chapman-Kolmogorov 2.6 e 2.7, juntamente com as condições iniciais $p_j(0) = I_{\{0\}}(j)$ pode ser resolvido recursivamente, dando origem a uma única solução $\{p_n(t); n \in \mathbb{N}\}$ para a distribuição de probabilidades de $X(t)$, dada por 2.8 e 2.9. Segue também das equações 2.8 e 2.9 que $p_n(t) > 0$, para todo $t > 0$. A única questão que resta é saber se $\{p_n(t); n \in \mathbb{N}\}$, $t > 0$, dá origem a distribuições de probabilidade próprias, ou seja, se $\sum_{j=0}^{\infty} p_j(t) = 1$ para todo $t > 0$ fixado. Quando isto ocorre dizemos que o processo é honesto.

Veremos que nem sempre isto ocorre, pois tomando $\{\lambda(n); n \in \mathbb{N}\}$ com um crescimento muito rápido pode-se ter $\sum_{j=0}^{\infty} p_j(t) < 1$. O primeiro membro desta inequação representa a probabilidade de ocorrer um número finito de eventos em um intervalo de tempo de comprimento t . Neste caso, a diferença entre o segundo membro e o primeiro membro da inequação representa a probabilidade de haver uma explosão, ou seja, um número infinito de ocorrências em determinado intervalo finito de tempo ($P\{X(t) = \infty\} > 0$).

Teorema 2.3.1 *Para que $\sum_{j=0}^{\infty} p_j(t) = 1$, para todo $t > 0$, é necessário e suficiente que a série $\sum_{j=0}^{\infty} \frac{1}{\lambda(j)}$ diverja.*

Prova

Seja $S_k(t) = p_0(t) + p_1(t) + \dots + p_k(t)$, $k = 0, 1, \dots$, a seqüência das somas parciais da série em questão. Como esta seqüência é não-decrescente e limitada superiormente por 1, segue que existe o limite

$$\mu(t) = \lim_{k \rightarrow \infty} [1 - S_k(t)].$$

Somando membro a membro as equações 2.6 e 2.7 de $n = 0$ até k , temos

$$S'_k(t) = -\lambda(k)p_k(t),$$

que juntamente com a condição inicial $S_k(0) = 1$ implica

$$S_k(t) = -\lambda(k) \int_0^t p_k(s) ds + 1.$$

Como $1 - S_k(t) \searrow \mu(t)$, quando $k \rightarrow \infty$, e $1 - S_k(t) \leq 1$, então

$$\mu(t) \leq 1 - S_k(t) \leq 1,$$

o que implica

$$\mu(t) \leq \lambda(k) \int_0^t p_k(s) ds \leq 1$$

ou

$$\frac{\mu(t)}{\lambda(k)} \leq \int_0^t p_k(s) ds \leq \frac{1}{\lambda(k)}. \quad (2.12)$$

Somando membro a membro 2.12 de $k = 0$ até n , temos

$$\mu(t) \left[\frac{1}{\lambda(0)} + \cdots + \frac{1}{\lambda(n)} \right] \leq \int_0^t S_n(s) ds \leq \frac{1}{\lambda(0)} + \cdots + \frac{1}{\lambda(n)}. \quad (2.13)$$

Suponhamos que $S_n(t) \rightarrow 1$ para todo t , quando $n \rightarrow \infty$. Então $\mu(t) = 0$ e a integral em 2.13 converge para t , quando $n \rightarrow \infty$. Logo, para valores arbitrariamente grandes de t , essa integral assume valores arbitrariamente grandes, para todo n suficientemente grande. Portanto, não podemos ter $\sum_{j=0}^{\infty} \frac{1}{\lambda(j)} < \infty$, pois isto implicaria que a integral em 2.13 é limitada para todo t e para todo n . Então, conclui-se que a série $\sum_{j=0}^{\infty} \frac{1}{\lambda(j)}$ diverge sempre que $S_n(t) \rightarrow 1$ quando $n \rightarrow \infty$, para todo t .

Reciprocamente, suponhamos que a série $\sum_{j=0}^{\infty} \frac{1}{\lambda(j)}$ diverja. Então, supondo que $\mu(t) > 0$ ou equivalentemente $\sum_{j=0}^{\infty} p_j(t) < 1$, para algum t , conclui-se que a integral em 2.13 é maior ou igual ao termo geral de uma seqüência divergente para $+\infty$, o que é um absurdo, pois para cada t fixado, a integral em 2.13 é limitada superiormente por t , para todo n . Portanto, a conclusão é que se $\sum_{j=0}^{\infty} \frac{1}{\lambda(j)}$ for divergente então $S_n(t) \rightarrow 1$ quando $n \rightarrow \infty$, para todo t . ■

O teorema acima se torna plausível quando interpretado de maneira probabilística. Lembremos que T_j é o tempo que o processo gasta no estado j . Já vimos que T_j tem distribuição exponencial com média $\frac{1}{\lambda(j)}$. Portanto, o tempo médio que o processo gasta passando por todos os estados é $\sum_{j=0}^{\infty} \frac{1}{\lambda(j)}$. Se esta série for finita significa que, em média, o processo passa por todos os estados possíveis em um tempo finito. O estado ocupado

pelo processo além deste tempo será infinito.

2.4 Distribuições Discretas derivadas de Processos de Poisson Estendido

Anteriormente vimos que para uma dada seqüência de taxas de transição $\{\lambda(n); n \in \mathbb{N}\}$ de um PPE honesto, existe uma única distribuição de probabilidades satisfazendo as equações diferenciais de Chapman-Kolmogorov para frente. Outro resultado importante é que dada qualquer distribuição de probabilidades sobre \mathbb{N} é possível determinar um PPE honesto, cuja distribuição unidimensional, para algum instante, coincide com a dada distribuição de probabilidades, como mostram os seguintes teoremas.

Teorema 2.4.1 *Se $\{\pi_i, i = 0, 1, \dots\}$ for uma distribuição de probabilidades com suporte em \mathbb{N} , isto é, $\pi_i > 0$ para todo $i \in \mathbb{N}$ e $\sum_{i=0}^{\infty} \pi_i = 1$, então existe uma única seqüência de taxas de transição $\{\lambda(n), n \in \mathbb{N}\}$ correspondente a um PPE honesto $\{X(t); t \geq 0\}$ para o qual, para algum $s > 0$, $p_i(s) = P\{X(s) = i\} = \pi_i$, para $i \in \mathbb{N}$.*

Prova

Fixado $s = s_0$, conclui-se da equação 2.8 que $p_0(s_0) = \pi_0$ se e somente se $\lambda(0) = -\ln(\pi_0)/s_0$. Agora, suponhamos que para algum m , existam únicos $\lambda(0), \lambda(1), \dots, \lambda(m)$ tais que a partir das equações diferenciais de Chapman-Kolmogorov 2.6 e 2.7 obtemos $p_i(s_0) = \pi_i$, para $i = 0, \dots, m$. Provaremos que existe um único $\lambda(m+1)$ tal que $p_{m+1}(s_0) = \pi_{m+1}$. Observe que, a partir de 2.9 obtém-se a seguinte relação

$$p_{m+1}(s_0) = \lambda(m) \int_0^{s_0} e^{-\lambda(m+1)[s_0-x]} p_m(x) dx.$$

Observando a relação acima como uma função de $\lambda(m+1)$, concluímos que essa função é contínua em $(0, +\infty)$. Entretanto, observe que quando $\lambda(m+1) \rightarrow \infty$ temos que $p_{m+1}(s_0) \rightarrow 0$ e quando $\lambda(m+1) = 0$, o estado $m+1$ é absorvente, com $p_{m+1}(s_0) = 1 - p_0(s_0) - p_1(s_0) - \dots - p_m(s_0)$. Como $p_{m+1}(s)$ é uma função estritamente monótona e contínua de $\lambda(m+1)$, concluímos que para qualquer número π entre 0 e $1 - p_0(s_0) -$

$p_1(s_0) - \cdots - p_m(s_0) = 1 - \pi_0 - \pi_1 - \cdots - \pi_m$ existe um único $\lambda(m+1) > 0$ tal que $p_{m+1}(s_0) = \pi$. Como $0 < \pi_{m+1} < 1 - \pi_0 - \pi_1 - \cdots - \pi_m$, existe um único $\lambda(m+1)$, tal $p_{m+1}(s_0) = \pi_{m+1}$. ■

Teorema 2.4.2 *Se $\{\pi_i, i = 0, 1, \dots, N\}$ for uma distribuição de probabilidades, isto é, $\pi_i > 0$ para todo $i = 1, \dots, N$ e $\sum_{i=0}^N \pi_i = 1$, então existe uma única seqüência finita de taxas de transição $\{\lambda(n), n = 0, 1, \dots, N\}$ que corresponde a um PPE honesto $\{X(t); t \geq 0\}$ para o qual, para algum $s > 0$, $p_i(s) = \pi_i$, para $i = 0, 1, \dots, N$. Além disso, $\lambda(N) = 0$.*

Prova

Fixando $s = s_0$ e utilizando o mesmo argumento anterior podemos provar que para cada π_i existe um único $\lambda(i)$ tal que $p_i(s_0) = \pi_i$, $i = 0, \dots, N$. Além disso, pode-se provar que $\lambda(N) = 0$. De fato, note que de 2.9 temos

$$p_N(s_0) = \lambda(N-1) \int_0^{s_0} e^{-\lambda(N)[s_0-x]} p_{N-1}(x) dx.$$

Analisando $p_N(s_0)$ como função de $\lambda(N)$ e utilizando a mesma argumentação da prova anterior concluímos que $p_N(s_0)$ assume valores entre 0 (quando $\lambda(N) \rightarrow \infty$) e $1 - p_0(s_0) - p_1(s_0) - \cdots - p_{N-1}(s_0) = 1 - \pi_0 - \pi_1 - \cdots - \pi_{N-1} = \pi_N$ (quando $\lambda(N) = 0$). Portanto, para que $p_N(s_0) = \pi_N$ devemos ter necessariamente $\lambda(N) = 0$. ■

2.4.1 Distribuição de Poisson

Teorema 2.4.3 *Se $\{X(t); t \geq 0\}$ for um PPE com taxas de transição $\lambda(n) = \lambda I_{\mathbb{N}}(n)$ com $\lambda > 0$, então a variável aleatória $X(t)$, $t > 0$, tem distribuição de Poisson com média λt .*

Prova

Supondo que as taxas de transição são dadas por $\lambda(n) = \lambda I_{\mathbb{N}}(n)$, vamos mostrar que $X(t)$ tem distribuição de Poisson com média λt , isto é

$$p_n(t) = P\{X(t) = n\} = \frac{e^{-\lambda t} (\lambda t)^n}{n!}, \quad n = 0, 1, 2, \dots \quad (2.14)$$

Provaremos a tese 2.14 por indução finita sobre n .

Para $n = 0$, temos por 2.8 que $p_0(t) = e^{-\lambda t}$. Portanto, 2.14 se verifica para $n = 0$.

Vamos supor por indução que a tese valha para $n = k$, isto é,

$$p_k(t) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}. \quad (2.15)$$

De 2.9, temos para $n = k + 1$ que

$$p_{k+1}(t) = \lambda e^{-\lambda t} \int_0^t e^{\lambda s} p_k(s) ds.$$

Substituindo 2.15 na última equação, temos

$$\begin{aligned} p_{k+1}(t) &= \lambda e^{-\lambda t} \int_0^t e^{\lambda s} \frac{e^{-\lambda s} (\lambda s)^k}{k!} ds \\ &= \frac{\lambda^{k+1} e^{-\lambda t}}{k!} \int_0^t s^k ds \\ &= \frac{\lambda^{k+1} e^{-\lambda t}}{k!} \frac{t^{k+1}}{(k+1)} \\ &= \frac{(\lambda t)^{k+1} e^{-\lambda t}}{(k+1)!}, \end{aligned}$$

o que implica que a expressão 2.14 vale para $n = k + 1$, desde que valha para $n = k$.

Portanto, 2.14 vale para todo n natural. ■

Corolário 2.4.4 *A variável $X(1)$ de um PPE honesto $\{X(t); t \geq 0\}$ tem distribuição de Poisson com média λ se as taxas de transição forem dadas por $\lambda(n) = \lambda I_{\mathbb{N}}(n)$.*

2.4.2 Distribuição Binomial Negativa

Teorema 2.4.5 *Se $\{X(t); t \geq 0\}$ for um PPE com taxas de transição $\lambda(n) = a(b + n)I_{\mathbb{N}}(n)$ com a e b números reais estritamente positivos, então a variável aleatória $X(t)$, $t > 0$, tem distribuição Binomial Negativa com parâmetros b e e^{-at} .*

Prova

Supondo que as taxas de transição são dadas por $\lambda(n) = a(b+n)I_{\mathbb{N}}(n)$, mostraremos que $X(t)$ tem distribuição Binomial Negativa com parâmetros b e e^{-at} , isto é,

$$p_n(t) = P\{X(t) = n\} = \binom{b+n-1}{n} (e^{-at})^b (1 - e^{-at})^n, \quad n = 0, 1, 2, \dots \quad (2.16)$$

Provaremos a tese 2.16 por indução finita sobre n .

Para $n = 0$, temos por 2.8 que $p_0(t) = e^{-abt}$. Portanto, 2.16 se verifica para $n = 0$.

Vamos supor por indução que a tese valha para $n = k$, isto é,

$$p_k(t) = \binom{b+k-1}{k} (e^{-at})^b (1 - e^{-at})^k. \quad (2.17)$$

De 2.9, temos para $n = k + 1$ que

$$p_{k+1}(t) = a(b+k)e^{-a(b+k+1)t} \int_0^t e^{a(b+k+1)s} p_k(s) ds.$$

Substituindo 2.17 na última equação, temos

$$\begin{aligned} p_{k+1}(t) &= a(b+k)e^{-a(b+k+1)t} \int_0^t e^{a(b+k+1)s} \binom{b+k-1}{k} (e^{-as})^b (1 - e^{-as})^k ds \\ &= \binom{b+k-1}{k} (b+k) a e^{-a(b+k+1)t} \int_0^t e^{a(k+1)s} (1 - e^{-as})^k ds \\ &= \binom{b+k}{k+1} (k+1) a e^{-a(b+k+1)t} \int_0^t e^{as} [e^{as} (1 - e^{-as})]^k ds \\ &= \binom{b+k}{k+1} (k+1) a e^{-a(b+k+1)t} \int_0^t e^{as} [e^{as} - 1]^k ds. \end{aligned}$$

Fazendo a substituição $u = e^{as} - 1$ na integral acima, obtém-se:

$$\begin{aligned}
 p_{k+1}(t) &= \binom{b+k}{k+1} (k+1) a e^{-a(b+k+1)t} \frac{1}{a} \int_0^{e^{at}-1} u^k du \\
 &= \binom{b+k}{k+1} (k+1) e^{-a(b+k+1)t} \frac{1}{k+1} u^{k+1} \Big|_0^{e^{at}-1} \\
 &= \binom{b+k}{k+1} e^{-a(b+k+1)t} (e^{at} - 1)^{k+1} \\
 &= \binom{b+k}{k+1} [e^{-at}]^b [e^{-at}(e^{at} - 1)]^{k+1} \\
 &= \binom{b+k}{k+1} [e^{-at}]^b [1 - e^{-at}]^{k+1}.
 \end{aligned}$$

Portanto, 2.16 vale para todo n natural. ■

Corolário 2.4.6 *A variável $X(1)$ de um PPE $\{X(t); t \geq 0\}$ tem distribuição Binomial Negativa com parâmetros r e p , com r real positivo e $0 < p < 1$, se as taxas de transição forem dadas por $\lambda(n) = -\ln(p)(r+n)I_{\mathbb{N}}(n)$.*

2.4.3 Distribuição Binomial

Teorema 2.4.7 *Se $\{X(t); t \geq 0\}$ for um PPE com taxas de transição $\lambda(n) = a(N-n)I_{\{0,\dots,N\}}(n)$, com a um número real estritamente positivo e N um número natural maior ou igual a 1, então a variável aleatória $X(t)$, $t > 0$, tem distribuição Binomial com parâmetros N e $1 - e^{-at}$.*

Prova

Supondo que as taxas de transição são dadas por $\lambda(n) = a(N-n)I_{\{0,\dots,N\}}(n)$, vamos mostrar que $X(t)$ tem distribuição Binomial com parâmetros N e $1 - e^{-at}$, isto é,

$$p_n(t) = P\{X(t) = n\} = \binom{N}{n} (1 - e^{-at})^n (e^{-at})^{N-n}, \quad n = 0, 1, 2, \dots, N. \quad (2.18)$$

Provaremos a tese 2.18 por indução finita sobre n .

Para $n = 0$ temos por 2.8 que $p_0(t) = e^{-aNt}$. Portanto, 2.18 se verifica para $n = 0$.

Vamos supor por indução que a tese valha para $n = k$, $k = 0, \dots, N - 1$, isto é

$$p_k(t) = \binom{N}{k} (1 - e^{-at})^k (e^{-at})^{N-k}. \quad (2.19)$$

De 2.9, temos para $n = k + 1$ que

$$p_{k+1}(t) = a(N - k)e^{-a(N-k-1)t} \int_0^t e^{a(N-k-1)s} p_k(s) ds.$$

Substituindo 2.19 na última equação, temos

$$\begin{aligned} p_{k+1}(t) &= a(N - k)e^{-a(N-k-1)t} \int_0^t e^{a(N-k-1)s} \binom{N}{k} (1 - e^{-as})^k (e^{-as})^{N-k} ds \\ &= \binom{N}{k} (N - k)a e^{-a(N-k-1)t} \int_0^t e^{-as} (1 - e^{-as})^k ds. \end{aligned}$$

Fazendo a substituição $u = 1 - e^{-as}$ na integral acima, obtém-se

$$\begin{aligned} p_{k+1}(t) &= \binom{N}{k+1} (k+1) e^{-a(N-k-1)t} \int_0^{1-e^{-at}} u^k du \\ &= \binom{N}{k+1} (k+1) e^{-a(N-k-1)t} \frac{1}{k+1} u^{k+1} \Big|_0^{1-e^{-at}} \\ &= \binom{N}{k+1} e^{-a(N-k-1)t} (1 - e^{-at})^{k+1} \\ &= \binom{N}{k+1} (e^{-at})^{N-(k+1)} (1 - e^{-at})^{k+1}, \end{aligned}$$

o que prova o teorema. ■

Corolário 2.4.8 *A variável $X(1)$ de um PPE $\{X(t); t \geq 0\}$ tem distribuição Binomial com parâmetros N e p se as taxas de transição forem dadas por $\lambda(n) = -\ln(1 - p)(N - n)I_{\{0, \dots, N\}}(n)$.*

2.5 Estudo da Dispersão do Processo de Poisson Estendido

Nesta seção estudaremos a dispersão de um PPE honesto em função da seqüência de suas taxas de transição $\{\lambda(n), n \in \mathbb{N}\}$. A distribuição de probabilidades unidimensionais de um PPE honesto, cujas taxas são não constantes não será mais de Poisson. Todavia, podemos comparar o nível de variabilidade da distribuição resultante dessas taxas com o nível de variabilidade da distribuição de Poisson. Seja $\{X(t); t \geq 0\}$ um PPE honesto. A distribuição de probabilidades $\{p_n(t); n \in \mathbb{N}\}$, $t > 0$, onde $p_n(t) = P\{X(t) = n\}$, $n \in \mathbb{N}$, se diz superdispersa relativamente à distribuição de Poisson se $Var\{X(t)\} > E\{X(t)\}$ e subdispersa se $Var\{X(t)\} < E\{X(t)\}$. Antes de enunciarmos o próximo teorema, que estabelece a relação entre seqüências não decrescentes e não crescentes de taxas de transição $\{\lambda(n), n \in \mathbb{N}\}$ de um PPE honesto e o correspondente nível de variabilidade comparado ao da distribuição de Poisson, faremos alguns comentários sobre ordenação estocástica. Sejam X e Y variáveis aleatórias assumindo valores em \mathbb{N} . A variável Y é dita estocasticamente maior do que X , denotado por $Y \stackrel{st}{\geq} X$, se $P\{Y \geq n\} \geq P\{X \geq n\}$, para todo $n \in \mathbb{N}$. Um importante resultado neste contexto é que, se $Y \stackrel{st}{\geq} X$ e f for uma função real definida em \mathbb{N} tal que $E\{f(Y)\}$ e $E\{f(X)\}$ sejam finitas, então $E\{f(Y)\} \geq E\{f(X)\}$ se f for não decrescente e $E\{f(Y)\} \leq E\{f(X)\}$ se f for não crescente. O resultado do teorema enunciado a seguir foi conjecturado por Faddy (1994) e provado por Ball (1995).

Teorema 2.5.1 *Seja $\{X(t); t \geq 0\}$ um PPE honesto. Se a seqüência das taxas de transição $\{\lambda(n); n \in \mathbb{N}\}$ for não decrescente, então $Var\{X(t)\} \geq E\{X(t)\}$ para todo t , $t > 0$. Por outro lado, se a seqüência das taxas de transição $\{\lambda(n); n \in \mathbb{N}\}$ for não crescente, então $Var\{X(t)\} \leq E\{X(t)\}$ para todo t , $t > 0$.*

Prova

Consideremos $D(t) = Var\{X(t)\} - E\{X(t)\}$, $t \geq 0$. Como $X(0) = 0$, temos que $D(0) = 0$ e se mostrarmos que $D'(t) > 0$ ou $D'(t) < 0$ dependendo se $\{\lambda(n); n \in \mathbb{N}\}$ é não decrescente ou não crescente, respectivamente, então $D(t) > 0$ para $\{\lambda(n); n \in \mathbb{N}\}$ não decrescente e $D(t) < 0$ para $\{\lambda(n); n \in \mathbb{N}\}$ não crescente.

Note que

$$\mu(t) = E\{X(t)\} = \sum_{n=0}^{\infty} np_n(t), \quad t \geq 0.$$

Usando as equações diferenciais de Chapman-Kolmogorov 2.6 e 2.7 temos

$$\begin{aligned} \mu'(t) &= \sum_{n=0}^{\infty} np'_n(t) \\ &= \sum_{n=1}^{\infty} n\{\lambda(n-1)p_{n-1}(t) - \lambda(n)p_n(t)\} \\ &= \sum_{n=0}^{\infty} (n+1)\lambda(n)p_n(t) - \sum_{n=1}^{\infty} n\lambda(n)p_n(t) \\ &= \sum_{n=0}^{\infty} \lambda(n)p_n(t). \end{aligned}$$

Definamos também,

$$\begin{aligned} \mu_2(t) &= E\{X(t)[X(t) - 1]\} \\ &= \sum_{n=0}^{\infty} n(n-1)p_n(t). \end{aligned}$$

Logo, usando novamente as equações 2.6 e 2.7 pode-se mostrar que

$$\mu'_2(t) = 2 \sum_{n=0}^{\infty} n\lambda(n)p_n(t),$$

o que implica

$$\begin{aligned} D(t) &= \text{Var}\{X(t)\} - E\{X(t)\}^2 \\ &= \mu_2(t) - \mu^2(t). \end{aligned}$$

Então,

$$\begin{aligned} D'(t) &= \mu'_2(t) - 2\mu(t)\mu'(t) \\ &= 2\left\{\sum_{n=0}^{\infty} n\lambda(n)p_n(t) - \mu(t) \sum_{n=0}^{\infty} \lambda(n)p_n(t)\right\}. \end{aligned}$$

O restante da prova se reduz a mostrar que

$$\sum_{n=0}^{\infty} n\lambda(n)p_n(t) \geq \mu(t)\left\{\sum_{n=0}^{\infty} \lambda(n)p_n(t)\right\} \quad (2.20)$$

ou

$$\sum_{n=0}^{\infty} n\lambda(n)p_n(t) \leq \mu(t)\left\{\sum_{n=0}^{\infty} \lambda(n)p_n(t)\right\}. \quad (2.21)$$

caso $\{\lambda(n); n \in \mathbb{N}\}$ seja não decrescente ou não crescente, respectivamente. Para cada t , $t > 0$, definamos as variáveis aleatórias Y_t com distribuição $P\{Y_t = n\} = (\mu(t))^{-1}np_n(t)$ e $P\{X_t = n\} = p_n(t)$, $n \in \mathbb{N}$. Ball (1995) provou que $Y_t \stackrel{st}{\geq} X_t$ e definindo a função f em \mathbb{N} por $f(n) = \lambda(n)$, $n \in \mathbb{N}$, segue que

$$E\{f(Y_t)\} = (\mu(t))^{-1} \sum_{n=0}^{\infty} n\lambda(n)p_n(t)$$

e

$$E\{f(X(t))\} = \sum_{n=0}^{\infty} \lambda(n)p_n(t).$$

Como $Y_t \stackrel{st}{\geq} X_t$, então vale a relação 2.20 ou 2.21 caso $\{\lambda(n); n \in \mathbb{N}\}$ seja não decrescente ou não crescente, respectivamente, isto é, $Var\{X(t)\} \geq E\{X(t)\}$ para $\{\lambda(n); n \in \mathbb{N}\}$ não decrescente e $Var\{X(t)\} \leq E\{X(t)\}$ para $\{\lambda(n); n \in \mathbb{N}\}$ não crescente. Pode-se mostrar também que estas últimas desigualdades são estritas quando $\{\lambda(n); n \in \mathbb{N}\}$ é estritamente crescente ou estritamente decrescente. ■

Em aplicações, seqüências não decrescentes ou não crescentes podem fornecer explicações para alguma super ou subdispersão apresentadas. Seqüências não decrescentes implicam que a chance de um nascimento aumenta à medida que mais nascimentos ocorrem, enquanto seqüências não crescentes implicam que esta chance diminui. Por exemplo, no contexto de abundância de espécies, onde observamos os números de animais de uma amostra de lugares, seqüências não decrescentes sugerem um comportamento social, implicando que os indivíduos escolhem morar próximos uns aos outros, enquanto seqüências não crescentes sugerem um comportamento anti-social, implicando que os indivíduos preferem morar sozinhos. A consideração de contagens com distribuição de Poisson, correspondente a uma seqüências de taxas de transição constante, sugere que os indivíduos

são indiferentes uns aos outros.

Além disso, já vimos que uma seqüência de taxas de transição constante gera a distribuição de Poisson, enquanto que uma seqüência com taxas de transição linearmente crescentes gera a distribuição Binomial Negativa. A partir disto, parece razoável conjecturar que uma seqüência de taxas de transição cujo crescimento está “entre” constante e linearmente crescente corresponde a uma distribuição cuja dispersão está, em algum sentido, “entre” a distribuição de Poisson e a Binomial Negativa, enquanto que uma seqüência de taxas de transição cujo crescimento é maior que linearmente crescente corresponde a uma distribuição cuja dispersão é “maior” do que aquela associada à Binomial Negativa.

Nesse sentido, Faddy (1997) conjecturou que qualquer seqüência $\{\lambda(n); n \in \mathbb{N}\}$ concavamente crescente, isto é, $\lambda(n+1) - \lambda(n) < \lambda(n) - \lambda(n-1)$, para todo $n \in \mathbb{N} - \{0\}$, corresponde a uma distribuição cuja dispersão está “entre” a de Poisson e a Binomial Negativa, enquanto que uma seqüência $\{\lambda(n); n \in \mathbb{N}\}$ convexamente crescente, isto é, $\lambda(n+1) - \lambda(n) > \lambda(n) - \lambda(n-1)$, para todo $n \in \mathbb{N} - \{0\}$, corresponde a uma distribuição cuja dispersão é maior que aquela associada a da Binomial Negativa. Esta conjectura ainda não foi provada, embora análise de dados e aproximações assintóticas dêem algum suporte a isto.

Como ilustração, consideremos $\{X(t); t \geq 0\}$ um PPE com taxas de transição $\{\lambda(n); n \in \mathbb{N}\}$ linearmente crescentes, dadas por $\lambda(n) = a(b+n)$, $n \in \mathbb{N}$, com $a > 0$ e $b > 0$. Como vimos $X(t)$, $t > 0$, tem distribuição Binomial Negativa com parâmetros b e e^{-at} , cuja variância $\sigma^2(t)$ é igual a

$$\sigma^2(t) = \frac{b(1 - e^{-at})}{e^{-2at}}$$

e média $\mu(t)$ é igual a

$$\mu(t) = \frac{b(1 - e^{-at})}{e^{-at}}.$$

Com isso, a variância $\sigma^2(t)$ pode ser relacionada com a média $\mu(t)$ pela expressão

$$\sigma^2(t) = \mu(t) \left(1 + \frac{\mu(t)}{b} \right) = \mu(t) \frac{\lambda(\mu(t))}{\lambda(0)}.$$

Portanto, dado um PPE $\{X(t); t \geq 0\}$ com taxas de transição $\{\lambda(n); n \in \mathbb{N}\}$, para o qual a média e a variância de $X(t)$ são $\mu(t)$ e $\sigma^2(t)$, $t > 0$, respectivamente, a conjectura

de Faddy afirma que se $\{\lambda(n); n \in \mathbb{N}\}$ for convexamente crescente, então a variância $\sigma^2(t)$ será maior que $\mu(t) \frac{\lambda(\mu(t))}{\lambda(0)}$, enquanto que se a seqüência $\{\lambda(n); n \in \mathbb{N}\}$ for côncavamente crescente, então $\mu(t) < \sigma^2(t) < \mu(t) \frac{\lambda(\mu(t))}{\lambda(0)}$, para todo $t > 0$.

Os resultados acima evidenciam a abrangência de propriedades de dispersão admitidas por modelos obtidos a partir de um Processo de Poisson Estendido. Estas propriedades de dispersão, determinadas pela correspondente seqüência das taxas de transição, são melhor resumidas graficamente, como representado na Figura 2.1.

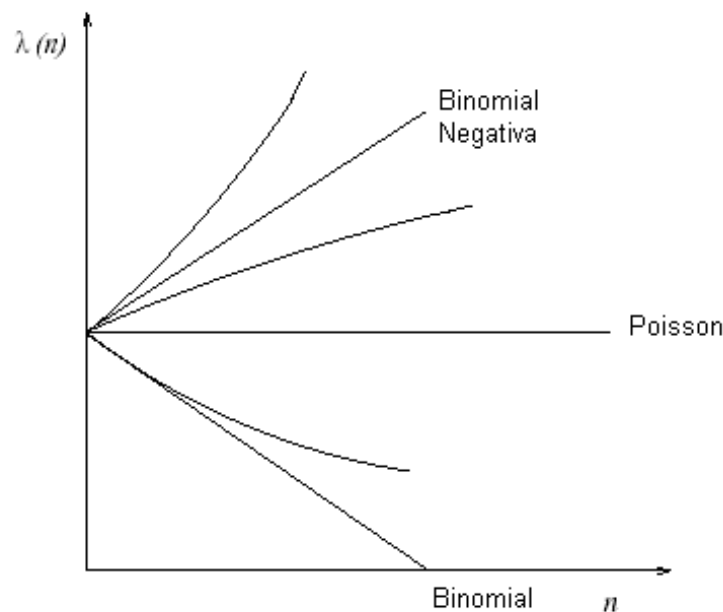


Figura 2.1. Taxas de Transição em função de n

2.6 Determinação das Distribuições de Probabilidades Unidimensionais do Processo

As distribuições de probabilidades unidimensionais de um PPE honesto são obtidas por meio da solução do sistema de equações diferenciais de Chapman-Kolmogorov para frente. Mostramos, resolvendo este sistema recursivamente, que os casos particulares com seqüências de taxas de transição constantes, linearmente decrescentes e linearmente crescentes geraram as distribuições de Poisson, Binomial e Binomial Negativa, respec-

tivamente. É possível também resolver essas equações para quaisquer seqüências de taxas de transição.

Uma expressão geral para $p_n(t)$ quando as taxas de transição $\lambda(n)$ são todas diferentes entre si é encontrada em Feller (1971):

$$p_n(t) = \frac{1}{\lambda(n)} \sum_{i=0}^n \left[\lambda(i) e^{-\lambda(i)t} \prod_{j \in \{0, \dots, n\} \setminus \{i\}} \frac{\lambda(j)}{\lambda(j) - \lambda(i)} \right]. \quad (2.22)$$

Uma outra forma de se determinar $p_n(t)$ é reescrever as equações diferenciais de Chapman-Kolmogorov sob a forma matricial. Definindo \mathbf{Q} como a matriz abaixo,

$$\mathbf{Q} = \begin{pmatrix} -\lambda(0) & \lambda(0) & 0 & \cdots & \cdots & 0 \\ 0 & -\lambda(1) & \lambda(1) & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & 0 & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \cdots & \ddots & -\lambda(n-1) & \lambda(n-1) \\ 0 & 0 & \cdots & \cdots & 0 & -\lambda(n) \end{pmatrix}, \quad (2.23)$$

pode-se reescrever o sistema de equações diferenciais de Chapman-Kolmogorov 2.6 e 2.7 como

$$\frac{d}{dt} \mathbf{p}(t) = \mathbf{p}(t) \mathbf{Q}, \quad (2.24)$$

sendo $\mathbf{p}(t) = (p_0(t), p_1(t), \dots, p_n(t))$ o vetor de probabilidades de $X(t)$ e \mathbf{Q} a matriz das taxas de transição do processo.

A solução de 2.24 é dada por

$$\mathbf{p}(t) = \mathbf{p}(0) \exp(\mathbf{Q}t) = \mathbf{p}(0) \sum_{i=0}^{\infty} \frac{t^i}{i!} \mathbf{Q}^i, \quad (2.25)$$

onde $\mathbf{p}(0)$ é o vetor de probabilidades de $X(0)$. Como $X(0) = 0$, temos que $\mathbf{p}(0)$ é um vetor linha com 1 na primeira coluna e 0 nas demais.

A equação 2.25 fornece a distribuição de probabilidades unidimensionais para um tempo fixado t . Sem perda de generalidade, podemos tomar $t = 1$.

Do ponto de vista computacional a solução na forma de exponenciação de matriz 2.25

é mais conveniente que a solução analítica 2.22, uma vez que esta última toma a forma de uma série alternada e só vale quando todas as taxas de transição são diferentes.

Capítulo 3

Modelo Estatístico e Estimação de Máxima Verossimilhança

3.1 Função de Verossimilhança

Sabemos que para um dado PPE honesto $\{X(t); t \geq 0\}$ corresponde uma distribuição de probabilidades em \mathbb{N} dada por $p_n(1) = P\{X(1) = n\}$, $n \in \mathbb{N}$. Seja Y_1, Y_2, \dots, Y_n uma amostra aleatória de $X(1)$, o que implica que a função de verossimilhança do modelo é dada por

$$\begin{aligned} L &= \prod_{i=1}^n P\{Y = y_i\} \\ &= \prod_{i=1}^n p_{y_i}. \end{aligned}$$

Considere $N = \max\{Y_i; i = 1, \dots, n\}$ o valor máximo observado e $n_i = \sum_{j=1}^n I_{\{i\}}(Y_j)$, $i = 0, 1, \dots, N$, ou seja, n_i é o número de observações iguais a i na amostra. Com estas novas variáveis a função de verossimilhança pode ser escrita como

$$L = \prod_{i=0}^N (p_i)^{n_i}.$$

Denotando $\mathbf{n} = (n_0, \dots, n_N)$ o vetor de frequências observadas na amostra e $\mathbf{p} = (p_0, \dots, p_N)$ o vetor de probabilidades da variável $X(1)$, o logaritmo da função de veros-

similhança pode ser escrito na forma matricial como

$$l = \sum_{i=0}^N n_i \ln(p_i) = \mathbf{n} \cdot \ln(\mathbf{p}), \quad (3.1)$$

onde $(\ln(\mathbf{p}))^T = (\ln(p_0), \dots, \ln(p_N))$ e $\mathbf{p} = \mathbf{e}_1 \cdot \exp(\mathbf{Q})$, com \mathbf{e}_1 sendo o vetor $(1, 0, \dots, 0)$ de tamanho $N + 1$ e \mathbf{Q} a matriz das taxas de transição dada em 2.23 com $n = N + 1$. Para determinarmos as estimativas de máxima verossimilhança dos parâmetros envolvidos utilizaremos um procedimento de otimização para maximizar o logaritmo da função de verossimilhança dada em 3.1.

3.2 Formas Paramétricas para as Taxas de Transição

Baseado nas considerações da seção 2.5 sobre a relação entre a média e a variância da variável $X(t)$, Faddy (1997) sugeriu a seguinte forma paramétrica para as taxas de transição: $\lambda(n) = a(b + n)^c$, em que $a > 0$, $b > 0$ e $c \leq 1$. Esta parametrização proporciona grande generalidade para a seqüência das taxas de transição. Para $c = 1$ esta parametrização dá origem à distribuição Binomial Negativa, para $c = 0$ temos a distribuição de Poisson, enquanto que para $0 < c < 1$ temos uma distribuição cuja variabilidade está entre a de Poisson e a Binomial Negativa. Para valores de $c < 0$ temos uma distribuição subdispersa relativamente à distribuição de Poisson.

Além da questão da variabilidade nos dados, é comum em análise de dados discretos haver uma quantidade excessiva de zeros. Para contornar esta situação, propomos algumas parametrizações das taxas de transição que são obtidas pela adição de um parâmetro λ_0 . Compararemos algumas parametrizações com e sem este parâmetro. Definamos os modelos:

$$\text{PPE1} : \lambda(n) = \lambda, \quad n \in \mathbb{N}, \lambda > 0;$$

$$\text{PPE2} : \lambda(n) = \begin{cases} \lambda_0, & n = 0, \\ \lambda, & n \geq 1, \end{cases} \quad n \in \mathbb{N}, \lambda_0 > 0, \lambda > 0;$$

$$\text{PPE3} : \lambda(n) = a(b+n), \quad n \in \mathbb{N}, a > 0, b > 0;$$

$$\text{PPE4} : \lambda(n) = \begin{cases} \lambda_0, & n = 0, \\ a(b+n), & n \geq 1, \end{cases} \quad n \in \mathbb{N}, \lambda_0 > 0, a > 0, b > -1;$$

$$\text{PPE5} : \lambda(n) = a(b+n)^c, \quad n \in \mathbb{N}, a > 0, b > 0, c \leq 1;$$

$$\text{PPE6} : \lambda(n) = \begin{cases} \lambda_0, & n = 0, \\ a(b+n)^c, & n \geq 1, \end{cases} \quad n \in \mathbb{N}, \lambda_0 > 0, a > 0, b > -1, c \leq 1;$$

Observe que os modelos PPE1 e PPE3 correspondem aos modelo de Poisson e Binomial Negativa, respectivamente. O modelo PPE5 corresponde a uma generalização de ambos proposta por Faddy (1997). Os modelos PPE2, PPE4 e PPE6 são as respectivas generalizações dos modelos PPE1, PPE3 e PPE5 para o caso com excesso de zeros.

3.3 Aplicação: Dados sobre Abundância de Animais

Os dados abaixo são contagens de *Macoma Liliiana*, um pequeno molusco, obtidas por amostragem em uma entre quatro localidades observadas por Faddy (1997). Uma alta frequência de zeros foi observada na localidade A, como pode se notar pela Tabela 3.1. Além disso, os dados mostraram uma grande superdispersão (variância > média).

Tabela 3.1. Dados de contagem de *Macoma Liliana*

Contagem	Frequências Observadas
0	9
1	2
2	1
3	2
4	1
5	1
7	4
8	1
9	4
10	1
11	3
12	2
13	1
14	3
16	1
17	1
18	1
20	1
25	1

$$\bar{y} = 7,725 \quad s^2 = 42,666$$

\bar{y} : média amostral s^2 : variância amostral

A tabela 3.2 a seguir fornece uma comparação destes modelos segundo as frequências observadas e esperadas segundo cada modelo. As rotinas computacionais usadas para ajustar os modelos deste capítulo se encontram no Apêndice A.

Tabela 3.2. Freqüências Observadas e Esperadas

Contagem	Observada	Freqüências					
		PPE1	PPE2	PPE3	PPE4	PPE5	PPE6
0	9	0,02	9,57	5,50	8,99	8,89	8,85
1	2	0,14	1,00	4,27	1,26	1,50	1,96
2	1	0,53	1,10	3,58	1,38	1,48	1,38
3	2	1,36	1,21	3,07	1,50	1,54	1,44
4	1	2,62	1,34	2,67	1,62	1,62	1,52
5	1	4,05	1,48	2,34	1,75	1,71	1,62
6	0	5,21	1,62	2,05	1,86	1,81	1,72
7	4	5,75	1,78	1,81	1,96	1,89	1,82
8	1	5,56	1,92	1,60	2,03	1,95	1,90
9	4	4,77	2,05	1,42	2,06	1,98	1,95
10	1	3,68	2,13	1,26	2,04	1,98	1,97
11	3	2,59	2,16	1,12	1,98	1,93	1,95
≥ 12	11	3,73	12,56	7,34	11,37	11,56	11,79

A Tabela 3.3 a seguir apresenta as estimativas de máxima verossimilhança (EMV) dos parâmetros, o logaritmo da função de verossimilhança calculado nos EMV, $l(\hat{\theta})$, a medida χ^2 e os critérios de seleção de modelos AIC(Akaike, 1974) e BIC (Schwarz, 1978) para cada um dos modelos.

Tabela 3.3. Estimativas de Máxima Verossimilhança.

	PPE1	PPE2	PPE3	PPE4	PPE5	PPE6
$\widehat{\lambda}_0$		1,43		1,49		1,51
$\widehat{\lambda}$	7,725	15,16				
\widehat{a}			2,30	0,45	10,42	13,21
\widehat{b}			0,86	26,23	0,00	-0,99
\widehat{c}					0,20	0,10
$l(\widehat{\theta})$	-192,04	-116,96	-124,13	-116,37	-116,37	-116,30
χ^2	60636,4	35,91	34,52	22,60	23,54	25,13
<i>AIC</i>	386,08	237,93	252,25	238,75	238,74	240,60
<i>BIC</i>	387,77	241,31	255,63	243,81	243,81	247,35

Na tabela acima, os critérios de seleção de modelos AIC e BIC são determinados por

$$AIC = -2 \ln(L(\widehat{\theta}; y)) + 2d,$$

$$BIC = -2 \ln(L(\widehat{\theta}; y)) + d \ln(n),$$

onde d é o número de parâmetros do modelo, n é o tamanho amostral e $\widehat{\theta}$ a estimativa de máxima verossimilhança de θ , vetor de parâmetros do modelo.

A medida χ^2 é definida por

$$\chi^2 = \sum_{i=1}^n \frac{(fe_i - fo_i)^2}{fe_i},$$

em que fe_i representa a frequência esperada e fo_i representa a frequência observada para a observação i .

Pela observação da tabela acima, podemos concluir que os modelos PPE2, PPE4 e PPE5 foram aqueles com melhor ajuste pelos critérios de seleção AIC e BIC, enquanto o modelo PPE1, que corresponde à distribuição de Poisson, foi o que pior se ajustou. Além disso, nota-se que a inserção do parâmetro λ_0 para acomodar o excesso de zeros, em geral, melhora de forma significativa o ajuste dos modelos, segundo estes critérios e também pela análise da Tabela 3.2.

3.4 Comparação com outros Modelos

Um modelo bastante difundido na modelagem de dados discretos com excesso de zeros é o modelo de Poisson com zeros inflacionados (ZIP). A função de probabilidades para este modelo é dada por

$$P\{Y = y | \theta, \omega\} = p(y | \theta, \omega) = \omega I_{\{0\}}(y) + (1 - \omega) \frac{\theta^y e^{-\theta}}{y!} I_{\{0,1,\dots\}}(y),$$

em que $0 \leq \omega \leq 1$ e $\theta > 0$. A esperança e a variância para este modelo são dadas por

$$\begin{aligned} E(Y) &= (1 - \omega)\theta = \tau, \\ \text{Var}(Y) &= \tau + \left(\frac{\omega}{1 - \omega}\right) \tau^2. \end{aligned}$$

Recentemente, Melkersson & Olsson (1999) (citado em Saito, 2005) apresentaram uma aplicação onde o modelo ZIP é adaptado também para o excesso de uns dos dados. Nesta aplicação, a distribuição ZOIP (Zeros and Ones Inflated Poisson) foi utilizada num estudo de contagem do número de visitas ao dentista de cidadãos suecos. A função de probabilidade para este modelo é

$$P\{Y = y | \theta, \omega_0, \omega_1, \omega_2\} = p(y | \theta, \omega) = \omega_0 I_{\{0\}}(y) + \omega_1 I_{\{1\}}(y) + \omega_2 \frac{\theta^y e^{-\theta}}{y!} I_{\{0,1,\dots\}}(y),$$

onde $\omega = (\omega_0, \omega_1, \omega_2)$, $\omega_0, \omega_1, \omega_2 \geq 0$ e $\omega_0 + \omega_1 + \omega_2 = 1$. Neste caso, a esperança e a variância são dadas por:

$$\begin{aligned} E(Y) &= \omega_1 + \omega_2 \theta, \\ \text{Var}(Y) &= \omega_1 + \omega_2 \theta (\theta + 1) - (\omega_1 + \omega_2 \theta)^2. \end{aligned}$$

A seguir, comparamos estes dois modelos (ZIP e ZOIP) com o modelo PPE2 e com

mais dois modelos baseados em Processos de Poisson Estendido, a saber

$$\text{PPE7} : \lambda(n) = \begin{cases} \lambda_0, & n = 0, \\ \lambda_1, & n = 1, \\ \lambda, & n \geq 2, \end{cases} \quad \lambda_0, \lambda_1, \lambda > 0;$$

$$\text{PPE8} : \lambda(n) = \begin{cases} \lambda_0, & n = 0, \\ \lambda_1, & n = 1, \\ a(b+n), & n \geq 2, \end{cases} \quad \lambda_0, \lambda_1, a, b > 0.$$

Os dados de Melkersson & Olsson (1999) estão dispostos na Tabela 3.4 abaixo.

Tabela 3.4. Dados sobre Número de Visitas ao Dentista

Número de Visitas	Frequência Observada
0	134
1	314
2	149
3	69
4	32
5	26
6	14
7	6
8	1
10	11
12	3
15	3
20	4
Total	766

$$\bar{y} = 1,935 \quad s^2 = 5,875$$

\bar{y} : média amostral s^2 : variância amostral

A Tabela 3.5 apresenta as frequências observadas e esperadas segundo os modelos em questão, o logaritmo da função de verossimilhanças calculado nos EMV, $l(\hat{\theta})$, a medida χ^2 e os critérios de seleção de modelos AIC(Akaike, 1974) e BIC (Schwarz, 1978) para

cada um dos modelos. Com relação à frequência de zeros, os modelos que tiveram melhor desempenho foram o ZIP, ZOIP e PPE8.

Tabela 3.5. Frequências Esperadas e Observadas

Contagem	Observada	Frequências				
		ZIP	PPE2	ZOIP	PPE7	PPE8
0	134	134,6	160,13	134,4	139,98	134,06
1	314	192,9	173,92	314	321,42	314,06
2	149	196,3	177,76	81,98	76,49	152,08
3	69	133,3	129,55	86,06	69,92	62,82
4	32	67,99	72,84	67,85	57,26	34,05
5	26	27,76	33,23	42,85	41,86	20,74
6	14	9,45	12,74	22,58	27,36	13,51
7	6	2,76	4,21	10,21	16,07	9,19
≥ 8	22	0,91	1,63	6,11	15,64	24,44
$l(\hat{\theta})$		-1587	-1571,9	-1480	-1437,5	-1338
χ^2		633,53	430,39	130,10	101,75	97,93
AIC		3178	3147,8	2966	2881	2684
BIC		3187	3157,08	2980	2895	2703

Pela Tabela 3.5, observa-se que o modelo PPE8, uma generalização da Binomial Negativa para dados com excesso de zeros e uns, tem melhor ajuste que os demais pelos critérios AIC e BIC. Nota-se também por esses critérios que os modelos PPE2 e PPE7 têm bom desempenho quando comparados com os modelos ZIP e ZOIP, respectivamente.

Capítulo 4

Generalização da Distribuição de Poisson para Dados com Inflação ou Deflação de Zeros

4.1 Modelo Probabilístico

Já sabemos do Teorema 2.4.3 que se $\{X(t); t \geq 0\}$ for um Processo de Poisson Estendido com taxas de transição $\lambda(n) = \lambda$, $\lambda > 0$, então a distribuição de $X(t)$ é de Poisson com média λt . Isto é,

$$P\{X(t) = n\} = f_{POIS}(n; \lambda t) = \frac{e^{-\lambda t} (\lambda t)^n}{n!}, \quad n \in \mathbb{N}, \quad (4.1)$$

sendo $f_{POIS}(\cdot; \lambda t)$ a função de probabilidades da distribuição de Poisson com parâmetro λt .

Uma generalização deste processo proposta por Faddy (1994) é obtida considerando um Processo de Poisson Estendido $\{Y(t); t \geq 0\}$ com taxas de transição dadas por

$$\lambda(n) = \begin{cases} \lambda_0, & \text{se } n = 0, \\ \lambda_1, & \text{se } n \geq 1, \end{cases} \quad (4.2)$$

onde $\lambda_0 > 0$, $\lambda_1 > 0$. Como no Capítulo 3, denotaremos este modelo por PPE2.

Utilizando a variável T_0 , como definida na seção 2.2, que representa o tempo de per-

manência do processo no estado 0 ou o tempo até a primeira ocorrência do evento, podemos determinar a distribuição de probabilidades da variável $Y(t)$, $P\{Y(t) = n\}$, $n \in \mathbb{N}$. De fato, observe que

$$P\{Y(t) = n\} = \int_0^{\infty} P\{Y(t) = n | T_0 = \tau\} f_{T_0}(\tau) d\tau$$

e como T_0 tem distribuição exponencial com parâmetro $\lambda(0)$, segue que

$$P\{Y(t) = n\} = \int_0^{\infty} P\{Y(t) = n | T_0 = \tau\} \lambda(0) e^{-\lambda(0)\tau} d\tau. \quad (4.3)$$

Observe que

$$P\{Y(t) = 0 | T_0 = \tau\} = \begin{cases} 1, & \text{se } \tau > t, \\ 0, & \text{se } \tau \leq t, \end{cases}$$

o que substituído na equação 4.3 implica

$$P\{Y(t) = 0\} = \int_t^{\infty} \lambda(0) e^{-\lambda(0)\tau} d\tau = e^{-\lambda_0 t}. \quad (4.4)$$

Para $n \geq 1$,

$$P\{Y(t) = n | T_0 = \tau\} = \begin{cases} 0, & \text{se } \tau > t, \\ P\{Y(t) = n | [Y(s) = 0, s < \tau], Y(\tau) = 1\}, & \text{se } \tau \leq t, \end{cases}$$

o que pela propriedade Markoviana 1.1 implica

$$P\{Y(t) = n | T_0 = \tau\} = \begin{cases} 0, & \text{se } \tau > t, \\ p_{1n}(t - \tau), & \text{se } \tau \leq t. \end{cases}$$

Substituindo este resultado em 4.3, temos para $n \geq 1$,

$$P\{Y(t) = n\} = \int_0^t p_{1n}(t - \tau) \lambda(0) e^{-\lambda(0)\tau} d\tau.$$

Como as funções $p_{1n}(x)$ obedecem às equações diferenciais 2.1 e 2.3, segue que

$$\begin{aligned} p'_{11}(x) &= -\lambda(1)p_{11}(x), \\ p'_{1n}(x) &= -\lambda(n)p_{1n}(x) + \lambda(n-1)p_{1,n-1}(x), \quad n = 2, 3, \dots \end{aligned}$$

Logo, considerando as transformações

$$\begin{aligned} g_n(x) &= p_{1,n+1}(x), \quad n = 0, 1, 2, \dots, \\ \xi(n) &= \lambda(n+1), \quad n = 0, 1, 2, \dots, \end{aligned}$$

obtem-se o seguinte sistema de equações diferenciais para as funções $\{g_n(x), n = 0, 1, \dots\}$:

$$\begin{aligned} g'_0(x) &= -\xi(0)g_0(x) \\ g'_n(x) &= -\xi(n)g_n(x) + \xi(n-1)g_{n-1}(x), \quad n = 1, 2, \dots, \end{aligned}$$

onde $\xi(n) = \lambda(n+1) = \lambda_1$ para $n = 0, 1, 2, \dots$. Este sistema pode ser resolvido de maneira análoga à prova do Teorema 2.4.3, obtendo-se como soluções as funções

$$g_n(x) = f_{POIS}(n; \lambda_1 x), \quad n = 0, 1, 2, \dots,$$

o que implica

$$p_{1n}(x) = g_{n-1}(x) = f_{POIS}(n-1; \lambda_1 x), \quad n = 1, 2, 3, \dots,$$

isto é,

$$p_{1n}(x) = \frac{e^{-\lambda_1 x} (\lambda_1 x)^{n-1}}{(n-1)!}, \quad n = 1, 2, 3, \dots$$

Substituindo este resultado em 4.3, temos para $n \geq 1$

$$P\{Y(t) = n\} = \int_0^t \frac{e^{-\lambda_1(t-\tau)} [\lambda_1(t-\tau)]^{n-1}}{(n-1)!} \lambda_0 e^{-\lambda_0 \tau} d\tau.$$

Fazendo a substituição $x = t - \tau$, temos para $n \geq 1$,

$$\begin{aligned} P\{Y(t) = n\} &= \int_0^t \frac{e^{-\lambda_1 x} [\lambda_1 x]^{n-1}}{(n-1)!} \lambda_0 e^{-\lambda_0(t-x)} dx \\ &= \frac{\lambda_0 \lambda_1^{n-1} e^{-\lambda_0 t}}{(n-1)!} \int_0^t x^{n-1} e^{-(\lambda_1 - \lambda_0)x} dx. \end{aligned} \quad (4.5)$$

Usando 4.4 e 4.5, pode-se determinar a função geradora de probabilidades de $Y(t)$, dada por

$$\begin{aligned} G(s) &= E\{s^{Y(t)}\} \\ &= \sum_{n=0}^{\infty} s^n P\{Y(t) = n\} \\ &= P\{Y(t) = 0\} + \lambda_0 e^{-\lambda_0 t} \sum_{n=1}^{\infty} \int_0^t \frac{(\lambda_1 x)^{n-1} s^n}{(n-1)!} e^{(\lambda_0 - \lambda_1)x} dx \\ &= e^{-\lambda_0 t} + s \lambda_0 e^{-\lambda_0 t} \int_0^t \left[\sum_{n=1}^{\infty} \frac{(\lambda_1 x)^{n-1} s^{n-1}}{(n-1)!} \right] e^{(\lambda_0 - \lambda_1)x} dx \\ &= e^{-\lambda_0 t} + s \lambda_0 e^{-\lambda_0 t} \int_0^t \left[\sum_{n=0}^{\infty} \frac{(\lambda_1 x s)^n}{n!} \right] e^{(\lambda_0 - \lambda_1)x} dx \\ &= e^{-\lambda_0 t} + s \lambda_0 e^{-\lambda_0 t} \int_0^t e^{\lambda_1 x s} e^{(\lambda_0 - \lambda_1)x} dx \\ &= e^{-\lambda_0 t} + s \lambda_0 e^{-\lambda_0 t} \int_0^t e^{[\lambda_0 + \lambda_1(s-1)]x} dx, \quad s \in \mathbb{R}, \end{aligned}$$

ou seja,

$$G(s) = \begin{cases} e^{-\lambda_0 t} + s \lambda_0 t e^{-\lambda_0 t}, & \text{se } s = 1 - \lambda_0/\lambda_1, \\ e^{-\lambda_0 t} + \frac{s \lambda_0 e^{-\lambda_0 t} \{e^{[\lambda_0 + \lambda_1(s-1)]t} - 1\}}{\lambda_0 + \lambda_1(s-1)}, & \text{se } s \neq 1 - \lambda_0/\lambda_1. \end{cases}$$

Naturalmente, podemos determinar a média e a variância de $Y(t)$ a partir da função geradora $G(s)$, contudo elas são mais facilmente determinadas diretamente a partir da

distribuição de probabilidade de $Y(t)$, isto é,

$$\begin{aligned}
E\{Y(t)\} &= \sum_{n=0}^{\infty} nP\{Y(t) = n\} \\
&= \sum_{n=1}^{\infty} n \int_0^t \frac{e^{-\lambda_1 x} [\lambda_1 x]^{n-1}}{(n-1)!} \lambda_0 e^{-\lambda_0(t-x)} dx \\
&= \int_0^t \left[\sum_{n=0}^{\infty} (n+1) \frac{e^{-\lambda_1 x} [\lambda_1 x]^n}{n!} \right] \lambda_0 e^{-\lambda_0(t-x)} dx \\
&= \int_0^t (\lambda_1 x + 1) \lambda_0 e^{-\lambda_0(t-x)} dx \\
&= \int_0^t \lambda_0 e^{-\lambda_0(t-x)} dx + \int_0^t \lambda_1 \lambda_0 x e^{-\lambda_0(t-x)} dx \\
&= \lambda_0 e^{-\lambda_0 t} \int_0^t e^{\lambda_0 x} dx + \lambda_1 \lambda_0 e^{-\lambda_0 t} \int_0^t x e^{\lambda_0 x} dx \\
&= (1 - e^{-\lambda_0 t}) + \lambda_1 t - \frac{\lambda_1}{\lambda_0} (1 - e^{-\lambda_0 t}) \\
&= \lambda_1 t + (e^{-\lambda_0 t} - 1) \left(\frac{\lambda_1}{\lambda_0} - 1 \right),
\end{aligned}$$

e de maneira similar obtemos a seguinte expressão para $E\{Y(t)[Y(t) - 1]\}$:

$$\begin{aligned}
 E\{Y(t)[Y(t) - 1]\} &= \sum_{n=2}^{\infty} n(n-1)P\{Y(t) = n\} \\
 &= \sum_{n=2}^{\infty} n(n-1) \int_0^t \frac{e^{-\lambda_1 x} [\lambda_1 x]^{n-1}}{(n-1)!} \lambda_0 e^{-\lambda_0(t-x)} dx \\
 &= \int_0^t \left[\sum_{n=2}^{\infty} n(n-1) \frac{e^{-\lambda_1 x} [\lambda_1 x]^{n-1}}{(n-1)!} \right] \lambda_0 e^{-\lambda_0(t-x)} dx \\
 &= \int_0^t \left[\lambda_1 x \sum_{n=2}^{\infty} n \frac{e^{-\lambda_1 x} [\lambda_1 x]^{n-2}}{(n-2)!} \right] \lambda_0 e^{-\lambda_0(t-x)} dx \\
 &= \int_0^t \left[\lambda_1 x \sum_{n=0}^{\infty} (n+2) \frac{e^{-\lambda_1 x} [\lambda_1 x]^n}{n!} \right] \lambda_0 e^{-\lambda_0(t-x)} dx \\
 &= \int_0^t \lambda_1 x (\lambda_1 x + 2) \lambda_0 e^{-\lambda_0(t-x)} dx \\
 &= \lambda_1^2 \lambda_0 \int_0^t x^2 e^{-\lambda_0(t-x)} dx + 2\lambda_1 \lambda_0 \int_0^t x e^{-\lambda_0(t-x)} dx \\
 &= \lambda_1^2 \lambda_0 e^{-\lambda_0 t} \int_0^t x^2 e^{\lambda_0 x} dx + 2\lambda_1 \lambda_0 e^{-\lambda_0 t} \int_0^t x e^{\lambda_0 x} dx \\
 &= (\lambda_1 t)^2 + 2 \frac{\lambda_1}{\lambda_0} \left(1 - \frac{\lambda_1}{\lambda_0} \right) [e^{-\lambda_0 t} + \lambda_0 t - 1] \\
 &= (\lambda_1 t)^2 + \left(\frac{\lambda_1}{\lambda_0} - 1 \right) \left[2 \frac{\lambda_1}{\lambda_0} (1 - e^{-\lambda_0 t}) - 2\lambda_1 t \right],
 \end{aligned}$$

o que implica

$$\begin{aligned}
 & Var\{Y(t)\} - E\{Y(t)\} = E\{Y(t)[Y(t) - 1]\} - E^2\{Y(t)\} \\
 = & (\lambda_1 t)^2 + \left(\frac{\lambda_1}{\lambda_0} - 1\right) \left[2\frac{\lambda_1}{\lambda_0}(1 - e^{-\lambda_0 t}) - 2\lambda_1 t \right] - \left[\lambda_1 t + (e^{-\lambda_0 t} - 1) \left(\frac{\lambda_1}{\lambda_0} - 1\right) \right]^2 \\
 = & (\lambda_1 t)^2 + \left(\frac{\lambda_1}{\lambda_0} - 1\right) \left[2\frac{\lambda_1}{\lambda_0}(1 - e^{-\lambda_0 t}) - 2\lambda_1 t \right] - (\lambda_1 t)^2 - 2\lambda_1 t \left(\frac{\lambda_1}{\lambda_0} - 1\right) (e^{-\lambda_0 t} - 1) \\
 & - (e^{-\lambda_0 t} - 1)^2 \left(\frac{\lambda_1}{\lambda_0} - 1\right)^2 \\
 = & \left(\frac{\lambda_1}{\lambda_0} - 1\right) \left\{ 2\frac{\lambda_1}{\lambda_0}(1 - e^{-\lambda_0 t}) - 2\lambda_1 t - 2\lambda_1 t(e^{-\lambda_0 t} - 1) - \left(\frac{\lambda_1}{\lambda_0} - 1\right) (e^{-2\lambda_0 t} - 2e^{-\lambda_0 t} + 1) \right\} \\
 = & \left(\frac{\lambda_1}{\lambda_0} - 1\right) \left\{ 2\frac{\lambda_1}{\lambda_0}(1 - e^{-\lambda_0 t}) - 2\lambda_1 t e^{-\lambda_0 t} + \frac{\lambda_1}{\lambda_0}(-e^{-2\lambda_0 t} + 2e^{-\lambda_0 t} - 1) + e^{-2\lambda_0 t} - 2e^{-\lambda_0 t} + 1 \right\} \\
 = & \left(\frac{\lambda_1}{\lambda_0} - 1\right) e^{-2\lambda_0 t} \left\{ \frac{\lambda_1}{\lambda_0}(e^{2\lambda_0 t} - 1) - 2(\lambda_1 t + 1)e^{\lambda_0 t} + e^{2\lambda_0 t} + 1 \right\}. \tag{4.6}
 \end{aligned}$$

Desta relação, pode-se mostrar que o fator entre colchetes é positivo para todo $t > 0$.

Para isso, basta considerarmos a função

$$g(t) = \frac{\lambda_1}{\lambda_0}(e^{2\lambda_0 t} - 1) - 2(\lambda_1 t + 1)e^{\lambda_0 t} + e^{2\lambda_0 t} + 1$$

e mostrarmos que $g(0) = 0$ e $g'(t) > 0$ para todo $t > 0$.

Portanto, de 4.6 pode-se concluir que: (i) se $\lambda_1 > \lambda_0$ então $Var\{Y(t)\} > E\{Y(t)\}$ e (ii) se $\lambda_1 < \lambda_0$ então $Var\{Y(t)\} < E\{Y(t)\}$, para todo $t > 0$. Isto é, se $\lambda_1 > \lambda_0$ temos um modelo superdisperso relativamente à distribuição de Poisson, enquanto que se $\lambda_1 < \lambda_0$ temos um modelo subdisperso.

A expressão para a variância de $Y(t)$ é dada por

$$Var\{Y(t)\} = \lambda_1 t + \left(\frac{\lambda_1}{\lambda_0} - 1\right) e^{-2\lambda_0 t} \left\{ \frac{\lambda_1}{\lambda_0}(e^{2\lambda_0 t} - 1) - 2(\lambda_1 t + 1)e^{\lambda_0 t} + e^{2\lambda_0 t} + 1 \right\}$$

4.2 Função de Verossimilhança

Consideremos Y_1, Y_2, \dots, Y_n uma amostra aleatória de tamanho n da distribuição de $Y(1)$ dada pelas equações 4.4 e 4.5 com $t = 1$. Então, a função de verossimilhança é dada por

$$\begin{aligned}
 L(\lambda_0, \lambda_1) &= \prod_{i=1}^n P\{Y_i = y_i\} \\
 &= \prod_{i: y_i=0} P\{Y_i = y_i\} \prod_{i: y_i>0} P\{Y_i = y_i\} \\
 &= e^{-n_0\lambda_0} \prod_{i: y_i>0} \left[\frac{\lambda_0 \lambda_1^{y_i-1} e^{-\lambda_0}}{(y_i-1)!} \int_0^1 x^{y_i-1} e^{-(\lambda_1-\lambda_0)x} dx \right] \\
 &= \frac{e^{-n_0\lambda_0} \lambda_0^{n-n_0} \lambda_1^{(\sum_{i=1}^n y_i) - (n-n_0)}}{\prod_{i: y_i>0} (y_i-1)!} \prod_{i: y_i>0} \int_0^1 x^{y_i-1} e^{-(\lambda_1-\lambda_0)x} dx,
 \end{aligned}$$

onde y_i , $i = 1, \dots, n$, são os valores observados e $n_0 = \sum_{j=1}^n I_{\{0\}}(y_j)$ o número de observações iguais a 0 na amostra.

O logaritmo da função de verossimilhança é

$$\begin{aligned}
 l(\lambda_0, \lambda_1) &= -n\lambda_0 + (n - n_0) \log(\lambda_0) + [\sum_{i=1}^n y_i - (n - n_0)] \log(\lambda_1) \\
 &\quad + \sum_{i: y_i>0} \log \left(\int_0^1 x^{y_i-1} e^{-(\lambda_1-\lambda_0)x} dx \right) - \sum_{i: y_i>0} \log(y_i - 1)!.
 \end{aligned}$$

Como os algoritmos iterativos utilizados na obtenção das estimativas de máxima verossimilhança têm melhor desempenho quando o espaço paramétrico não tem restrição, adotaremos a seguinte reparametrização

$$l_o = \log(\lambda_0),$$

$$l_1 = \log(\lambda_1).$$

Com essa reparametrização, a função de verossimilhança é dada por

$$L(l_0, l_1) = \frac{\exp\{-ne^{l_0} + (n - n_0)l_0 + [\sum_{i=1}^n y_i - (n - n_0)]l_1\}}{\prod_{i: y_i > 0} (y_i - 1)!} \prod_{i: y_i > 0} \int_0^1 x^{y_i-1} \exp\{-(e^{l_1} - e^{l_0})x\} dx \quad (4.7)$$

e conseqüentemente o logaritmo da função de verossimilhança é dado por

$$l(l_0, l_1) = -ne^{l_0} + (n - n_0)l_0 + [\sum_{i=1}^n y_i - (n - n_0)]l_1 + \sum_{i: y_i > 0} \log\left(\int_0^1 x^{y_i-1} \exp\{-(e^{l_1} - e^{l_0})x\} dx\right) - \sum_{i: y_i > 0} \log(y_i - 1)!. \quad (4.8)$$

4.3 Análise de Dados

Nesta seção o modelo apresentado foi ajustado a dados sobre o número de acidentes em 647 operárias na fabricação de bombas durante cinco semanas, estudados por Greenwood & Yule (1920) (citado em Saito, 2005). Os dados são apresentados na Tabela 4.1.

Tabela 4.1. Número de acidentes em operárias

Nº de acidentes	0	1	2	3	4	5
Frequência	447	132	42	21	3	2

$$\bar{y} = 0,4652 \text{ e } s^2 = 0,6919$$

\bar{y} : média amostral s^2 : variância amostral

As estimativas de máxima verossimilhança dos parâmetros, apresentadas na Tabela 4.2, foram obtidas via maximização da função 4.8 e os intervalos de confiança foram obtidos assumindo-se uma distribuição normal assintótica para os estimadores de máxima verossimilhança. As rotinas utilizadas para os cálculos desta seção se encontram no Apêndice B e foram implementadas com o uso do software estatístico R.

Tabela 4.2. Estimativas de Máxima Verossimilhança e Intervalos de Confiança 95%

	EMV	IC 95%
l_0	-0,9962	[-1,1354; -0,8567]
l_1	-0,0422	[-0,2515; 0,1670]

Um resumo das frequências observadas e esperadas segundo a distribuição de Poisson ($l_0 = l_1$) e o modelo proposto PPE2 é apresentado na Tabela 4.3 e no histograma da Figura 4.1.

Tabela 4.3. Frequências esperadas e observadas

Nº de visitas	Frequência		
	Observada	Poisson	PPE2
0	447	406,31	447,20
1	132	189,03	124,80
2	42	43,97	53,97
3	21	6,82	16,37
4	3	0,79	3,80
5	2	0,07	0,71
$-2\ln(L)$		1234,369	1185,749
AIC		1236,369	1189,749
BIC		1247,313	1198,694

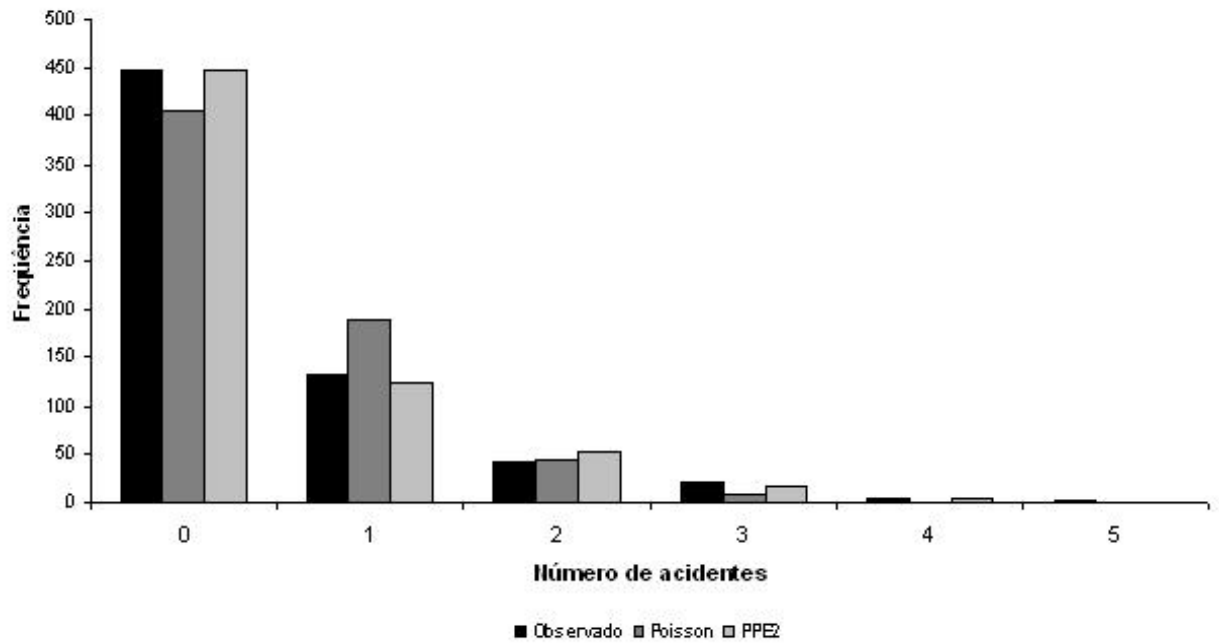


Figura 4.1. Histograma para os dados do número de acidentes em operárias

Podemos definir a função padronizada do logaritmo da função de verossimilhanças, l_p , como

$$l_p(\theta) = \frac{l(\theta)}{|l(\hat{\theta})|}, \quad (4.9)$$

onde $\hat{\theta}$ é a estimativa de máxima verossimilhança de θ . Os contornos para essa função são apresentados na Figura 4.2 abaixo.

Capítulo 5

Generalização da Distribuição

Binomial Negativa para Dados com Inflação ou Deflação de Zeros

5.1 Modelo Probabilístico

Sabemos do Teorema 2.4.5 que se $\{X(t); t \geq 0\}$ for um Processo de Poisson Estendido com taxas de transição $\lambda(n) = a(b+n)$ com $a > 0$ e $b > 0$, $n \in \mathbb{N}$, então a distribuição de $X(t)$ é Binomial Negativa com parâmetros b e e^{-at} , isto é,

$$P\{X(t) = n\} = f_{BN}(n; b, e^{-at}) = \binom{b+n-1}{n} e^{-abt} (1 - e^{-at})^n, \quad (5.1)$$

sendo $f_{BN}(\cdot; r, p)$ a função de probabilidades da distribuição Binomial Negativa com parâmetros r e p .

Uma generalização deste processo é obtida considerando um Processo de Poisson Estendido $\{Y(t); t \geq 0\}$ com taxas de transição dadas por

$$\lambda(n) = \begin{cases} \lambda_0, & \text{se } n = 0, \\ a(b+n), & \text{se } n \geq 1, \end{cases} \quad (5.2)$$

sendo $\lambda_0 > 0$, $a > 0$ e $b > -1$. Como no Capítulo 3, denotaremos este modelo por PPE4.

Utilizando a variável T_0 , que representa o tempo de permanência do processo no estado

0, isto é, o tempo até a primeira ocorrência do evento, podemos determinar a distribuição de probabilidades da variável $Y(t)$, $P\{Y(t) = n\}$, $n \in \mathbb{N}$. De fato, observe que

$$P\{Y(t) = n\} = \int_0^{\infty} P\{Y(t) = n | T_0 = \tau\} f_{T_0}(\tau) d\tau$$

e como T_0 tem distribuição exponencial com parâmetro $\lambda(0)$, segue que

$$P\{Y(t) = n\} = \int_0^{\infty} P\{Y(t) = n | T_0 = \tau\} \lambda(0) e^{-\lambda(0)\tau} d\tau. \quad (5.3)$$

Observemos que a probabilidade condicional de 5.3 para $n = 0$ é dada por

$$P\{Y(t) = 0 | T_0 = \tau\} = \begin{cases} 1, & \text{se } \tau > t, \\ 0, & \text{se } \tau \leq t, \end{cases}$$

o que substituído na equação 5.3 implica que

$$P\{Y(t) = 0\} = \int_t^{\infty} \lambda(0) e^{-\lambda(0)\tau} d\tau = e^{-\lambda(0)t}. \quad (5.4)$$

Para $n \geq 1$, a probabilidade condicional de 5.3 é dada por

$$P\{Y(t) = n | T_0 = \tau\} = \begin{cases} 0, & \text{se } \tau > t, \\ P\{Y(t) = n | [Y(s) = 0, s < \tau], Y(\tau) = 1\}, & \text{se } \tau \leq t, \end{cases}$$

o que pela propriedade Markoviana 1.1 implica que

$$P\{Y(t) = n | T_0 = \tau\} = \begin{cases} 0, & \text{se } \tau > t, \\ p_{1n}(t - \tau), & \text{se } \tau \leq t. \end{cases}$$

Substituindo este resultado em 5.3, temos para $n \geq 1$ que

$$P\{Y(t) = n\} = \int_0^t p_{1n}(t - \tau) \lambda(0) e^{-\lambda(0)\tau} d\tau.$$

Como as funções $p_{1n}(x)$ obedecem às equações diferenciais 2.1 e 2.3, segue que

$$\begin{aligned} p'_{11}(x) &= -\lambda(1)p_{11}(x), \\ p'_{1n}(x) &= -\lambda(n)p_{1n}(x) + \lambda(n-1)p_{1,n-1}(x), \quad n = 2, 3, \dots \end{aligned}$$

e considerando as transformações

$$\begin{aligned} g_n(x) &= p_{1,n+1}(x), \quad n = 0, 1, 2, \dots, \\ \xi(n) &= \lambda(n+1), \quad n = 0, 1, 2, \dots, \end{aligned}$$

obtem-se o seguinte sistema de equações diferenciais para as funções $\{g_n(x), n = 0, 1, \dots\}$:

$$\begin{aligned} g'_0(x) &= -\xi(0)g_0(x), \\ g'_n(x) &= -\xi(n)g_n(x) + \xi(n-1)g_{n-1}(x), \quad n = 1, 2, \dots, \end{aligned}$$

onde $\xi(n) = \lambda(n+1) = a(b+n+1)$ para $n = 0, 1, 2, \dots$. Este sistema pode ser resolvido de maneira análoga ao que foi feito no Teorema 2.4.5, obtendo-se as soluções

$$g_n(x) = f_{BN}(n; b+1, e^{-ax}), \quad n = 0, 1, 2, \dots,$$

o que implica

$$p_{1n}(x) = g_{n-1}(x) = f_{BN}(n-1; b+1, e^{-ax}), \quad n = 1, 2, 3, \dots,$$

isto é,

$$p_{1n}(x) = \binom{b+n-1}{n-1} (e^{-ax})^{b+1} (1 - e^{-ax})^{n-1}.$$

Substituindo este resultado em 5.3, temos para $n \geq 1$,

$$\begin{aligned} P\{Y(t) = n\} &= \int_0^t \binom{b+n-1}{n-1} [e^{-a(t-\tau)}]^{b+1} [1 - e^{-a(t-\tau)}]^{n-1} \lambda_0 e^{-\lambda_0 \tau} d\tau \\ &= \binom{b+n-1}{n-1} \lambda_0 \int_0^t e^{-a(b+1)(t-\tau)} e^{-\lambda_0 \tau} [1 - e^{-a(t-\tau)}]^{n-1} d\tau. \end{aligned}$$

Fazendo a substituição $x = t - \tau$ temos para $n \geq 1$,

$$\begin{aligned} P\{Y(t) = n\} &= \binom{b+n-1}{n-1} \lambda_0 e^{-\lambda_0 t} \int_0^t e^{[\lambda_0 - a(b+1)]x} [1 - e^{-ax}]^{n-1} dx \\ &= \binom{b+n-1}{n-1} \lambda_0 e^{-\lambda_0 t} \int_0^t (e^{-ax})^{b+1-\lambda_0/a} [1 - e^{-ax}]^{n-1} dx. \end{aligned}$$

Substituindo-se $y = e^{-ax}$ temos para $n \geq 1$,

$$P\{Y(t) = n\} = \binom{b+n-1}{n-1} \frac{\lambda_0 e^{-\lambda_0 t}}{a} \int_{e^{-at}}^1 y^{b-\lambda_0/a} [1-y]^{n-1} dy, \quad (5.5)$$

e usando 5.4 e 5.5, determinamos a função geradora de probabilidades de $Y(t)$, dada por

$$\begin{aligned} G(s) &= E\{s^{Y(t)}\} \\ &= \sum_{n=0}^{\infty} s^n P\{Y(t) = n\} \\ &= P\{Y(t) = 0\} + \sum_{n=1}^{\infty} s^n \binom{b+n-1}{n-1} \frac{\lambda_0 e^{-\lambda_0 t}}{a} \int_{e^{-at}}^1 y^{b-\lambda_0/a} [1-y]^{n-1} dy \\ &= e^{-\lambda_0 t} + \frac{\lambda_0 e^{-\lambda_0 t}}{a} s \int_{e^{-at}}^1 y^{b-\lambda_0/a} \left[\sum_{n=1}^{\infty} \binom{b+n-1}{n-1} [s(1-y)]^{n-1} \right] dy \\ &= e^{-\lambda_0 t} + \frac{\lambda_0 e^{-\lambda_0 t}}{a} s \int_{e^{-at}}^1 y^{b-\lambda_0/a} \left[\sum_{n=0}^{\infty} \binom{b+n}{n} [s(1-y)]^n \right] dy \\ &= e^{-\lambda_0 t} + \frac{\lambda_0 e^{-\lambda_0 t}}{a} s \int_{e^{-at}}^1 y^{b-\lambda_0/a} \left[\sum_{n=0}^{\infty} \binom{-b-1}{n} (-1)^n [s(1-y)]^n \right] dy \\ &= e^{-\lambda_0 t} + \frac{\lambda_0 e^{-\lambda_0 t}}{a} s \int_{e^{-at}}^1 y^{b-\lambda_0/a} \left[\sum_{n=0}^{\infty} \binom{-b-1}{n} [s(y-1)]^n \right] dy. \end{aligned}$$

Para que a série dada no integrando acima seja convergente, devemos ter s tal que $-1 < s(y-1) < 1$. Para que esta relação valha para todos os valores de y no intervalo de integração $[e^{-at}, 1]$, devemos ter s tal que $\frac{-1}{1-e^{-at}} < s < \frac{1}{1-e^{-at}}$. Então, a função $G(s)$

se expressa como

$$G(s) = e^{-\lambda_0 t} + \frac{\lambda_0 e^{-\lambda_0 t}}{a} s \int_{e^{-at}}^1 y^{b-\lambda_0/a} [1 + s(y-1)]^{-b-1} dy,$$

onde $s \in \left(\frac{-1}{1-e^{-at}}, \frac{1}{1-e^{-at}}\right)$.

A partir de $G(s)$ determina-se a média de $Y(t)$,

$$E\{Y(t)\} = G'(1) = \frac{\lambda_0(b+1)}{\lambda_0+a} (e^{at} - e^{-\lambda_0 t}) - b(1 - e^{-\lambda_0 t})$$

e a variância de $Y(t)$,

$$\begin{aligned} Var\{Y(t)\} &= G''(1) + G'(1) - [G'(1)]^2 = \\ &= \left[\frac{\lambda_0(b+1)}{\lambda_0+a} - \frac{\lambda_0(b+1)(b+2)}{\lambda_0+2a} + b^2 \right] e^{-\lambda_0 t} + \left[\frac{\lambda_0(b+1)(b+2)}{\lambda_0+2a} - \frac{\lambda_0^2(b+1)^2}{(\lambda_0+a)^2} \right] e^{2at} + \\ &= \left[\frac{2\lambda_0^2(b+1)^2}{(\lambda_0+a)^2} - \frac{2\lambda_0 b(b+1)}{\lambda_0+a} \right] e^{(a-\lambda_0)t} - \left[\frac{\lambda_0(b+1)}{\lambda_0+a} \right] e^{at} + \frac{2\lambda_0 b(b+1)}{\lambda_0+a} - \\ &= \frac{\lambda_0^2(b+1)^2}{(\lambda_0+a)^2} - b^2. \end{aligned}$$

5.2 Função de Verossimilhança

Considere Y_1, Y_2, \dots, Y_n uma amostra aleatória de tamanho n da distribuição de $Y(1)$ dada pelas expressões 5.4 e 5.5 com $t = 1$. Então, a função de verossimilhança é dada por

$$\begin{aligned}
 L(\lambda_0, a, b) &= \prod_{i=1}^n P\{Y_i = y_i\} \\
 &= \prod_{i:y_i=0} P\{Y_i = y_i\} \prod_{i:y_i>0} P\{Y_i = y_i\} \\
 &= \prod_{i:y_i=0} e^{-\lambda_0} \prod_{i:y_i>0} \binom{b+y_i-1}{y_i-1} \frac{\lambda_0 e^{-\lambda_0}}{a} \int_{e^{-a}}^1 x^{b-\lambda_0/a} [1-x]^{y_i-1} dx \\
 &= \frac{e^{-n_0 \lambda_0} \lambda_0^{(n-n_0)} e^{-(n-n_0)\lambda_0}}{a^{n-n_0}} \prod_{i:y_i>0} \binom{b+y_i-1}{y_i-1} \int_{e^{-a}}^1 x^{b-\lambda_0/a} [1-x]^{y_i-1} dx \\
 &= \frac{\lambda_0^{(n-n_0)} e^{-n\lambda_0}}{a^{n-n_0}} \prod_{i:y_i>0} \binom{b+y_i-1}{y_i-1} \int_{e^{-a}}^1 x^{b-\lambda_0/a} [1-x]^{y_i-1} dx,
 \end{aligned}$$

onde $y_i, i = 1, \dots, n$, são os valores observados e $n_0 = \sum_{j=1}^n I_{\{0\}}(y_j)$ o número de observações iguais a 0 na amostra.

O logaritmo da função de verossimilhança é dado por

$$l(\lambda_0, a, b) = -n\lambda_0 + (n-n_0) \log(\lambda_0) - (n-n_0) \log(a) + \sum_{i:y_i>0} \log \left\{ \binom{b+y_i-1}{y_i-1} \int_{e^{-a}}^1 x^{b-\lambda_0/a} [1-x]^{y_i-1} dy \right\}.$$

Com o objetivo de melhorar o desempenho dos algoritmos iterativos utilizados na obtenção das estimativas de máxima verossimilhança, tornaremos o espaço paramétrico irrestrito adotando a seguinte reparametrização:

$$\begin{aligned}
 l_o &= \log(\lambda_0), \\
 l_a &= \log(a), \\
 l_b &= \log(b+1).
 \end{aligned}$$

Com essa reparametrização, a função de verossimilhança é dada por

$$L(l_0, l_a, l_b) = \exp\{(n-n_0)(l_0-l_a)-ne^{l_0}\} \prod_{i:y_i>0} \binom{e^{l_b} + y_i - 2}{y_i - 1} \int_{\exp(-e^{l_a})}^1 x^{e^{l_b}-e^{(l_0-l_a)}-1} (1-x)^{y_i-1} dx \tag{5.6}$$

e conseqüentemente o logaritmo da função de verossimilhança será dado por

$$l(l_0, l_a, l_b) = (n-n_0)(l_0-l_a)-ne^{l_0} + \sum_{i:y_i>0} \log \left\{ \binom{e^{l_b} + y_i - 2}{y_i - 1} \int_{\exp(-e^{l_a})}^1 x^{e^{l_b}-e^{(l_0-l_a)}-1} (1-x)^{y_i-1} dx \right\}. \tag{5.7}$$

5.3 Análise de Dados

Nesta seção o modelo apresentado foi ajustado aos dados da seção 3.4 que se referem à contagem do número de visitas de indivíduos ao dentista no período de um ano. Os dados resumidos são apresentados na Tabela 5.1 e consistem de 766 observações em que o maior valor observado é 20.

Tabela 5.1. Número de visitas ao dentista em um ano

Nro de visitas	0	1	2	3	4	5	6	7	≥ 8
Frequência	134	314	149	69	32	26	14	6	22

$$\bar{y} = 1,93 \text{ e } s^2 = 5,87$$

\bar{y} : média amostral s^2 : variância amostral

As estimativas de máxima verossimilhança foram obtidas via maximização do logaritmo da função de verossimilhança 5.7. Os intervalos de confiança para os parâmetros são obtidos assumindo uma distribuição normal assintótica para os estimadores de máxima verossimilhança dos parâmetros. As rotinas utilizadas para estes cálculos se encontram

no Apêndice C e foram implementadas com o uso do software estatístico R.

Tabela 5.2. Estimativas de Máxima Verossimilhança e Intervalos de Confiança 95%

	EMV	IC 95%
l_0	0,5544	[0,4662; 0,6426]
l_a	0,3994	[0,2514; 0,5474]
l_b	-0,2854	[-0,5011; -0,0697]

Um resumo das frequências observadas e esperadas segundo a distribuição Binomial Negativa e o modelo proposto PPE4 é apresentado na Tabela 5.2 abaixo e no histograma da Figura 5.1.

Tabela 5.3. Frequências Esperadas e Observadas

Nº de visitas	Frequência		
	Observada	Binomial Negativa	PPE4
0	134	201,89	134,32
1	314	193,32	324,00
2	149	141,38	126,19
3	69	92,72	66,77
4	32	57,32	39,40
5	26	34,14	24,63
6	14	19,81	15,97
7	6	11,29	10,62
≥ 8	22	14,09	23,72
$-2 \ln(L)$		2834,030	2686,566
AIC		2838,030	2692,566
BIC		2847,313	2706,490

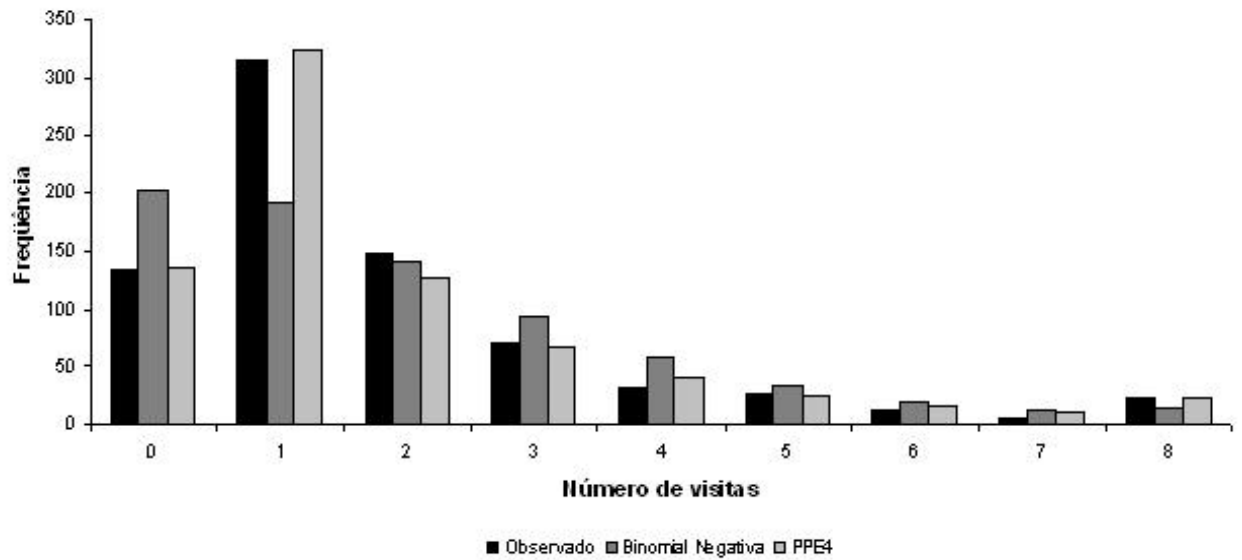


Figura 5.1. Histograma do número de visitas ao dentista

Os contornos para a função de log verossimilhança padronizada, como definido em 4.9, obtidos fixando-se \hat{l}_0 , \hat{l}_a e \hat{l}_b (EMV) são apresentados nas Figuras 5.2, 5.3 e 5.4, respectivamente.

Capítulo 6

Estudo de Simulação

Neste capítulo realizaremos um estudo de simulação com o intuito de verificar a adequação dos métodos assintóticos utilizados na construção de intervalos de confiança e teste de hipóteses dos modelos PPE2 e PPE4, apresentados nos Capítulos 4 e 5 respectivamente. Na seção 6.1 apresentaremos a metodologia que será adotada para este estudo. Nas seções 6.2 e 6.3 apresentaremos os resultados obtidos para as probabilidades de cobertura dos intervalos assintóticos, para o decaimento da variância dos EMV em função do tamanho amostral e para as taxas de rejeição dos testes de hipóteses para cada um dos modelos PPE2 e PPE4.

6.1 Metodologia

A estimação intervalar para os parâmetros de interesse realizados nos capítulos 4 e 5 foi baseada na distribuição assintótica normal dos EMV, enquanto que os testes de hipóteses foram baseados na distribuição assintótica qui-quadrado da estatística da razão de verossimilhança. Para avaliar a eficiência destes métodos para diferentes tamanhos amostrais realizaremos um estudo de simulação com o intuito de estimar as probabilidades de cobertura dos intervalos de confiança assintóticos, assim como estimar as taxas de rejeição dos testes assintóticos da razão de verossimilhança sob a hipótese H_0 . Espera-se que as estimativas das probabilidades de cobertura e das taxas de rejeição sejam mais próximas dos valores nominais (0,95 e 0,05, respectivamente) quanto maior for o tamanho amostral.

Para a estimação das probabilidades de cobertura dos intervalos de confiança assintóticos e das taxas de rejeição dos testes de hipóteses, faremos um total de 4000 simulações para cada um dos modelos PPE2 e PPE4, sendo 1000 simulações para cada tamanho amostral $n = 30, 80, 150$ e 400 .

As probabilidades de cobertura são estimadas, em cada caso, como a proporção observada dos intervalos de confiança que contém o valor verdadeiro do parâmetro. Para analisar a proximidade entre a proporção observada de intervalos de confiança assintóticos que contém o valor verdadeiro do parâmetro e a probabilidade de cobertura nominal $(0, 95)$, vamos construir um intervalo (assintótico) que contenha esta proporção observada com probabilidade $0, 95$, supondo que a probabilidade de cobertura do intervalo de confiança assintótico seja igual a nominal $(0, 95)$. Para isto suponha $I_1, I_2, \dots, I_{1000}$, variáveis aleatórias tais que I_j assume o valor 1 se o j -ésimo intervalo de confiança assintótico tiver o valor verdadeiro do parâmetro e assume o valor 0 caso contrário, $j = 1, \dots, 1000$. Se a probabilidade de cobertura dos intervalos de confiança assintóticos for igual à probabilidade de cobertura nominal $(0, 95)$, então as variáveis aleatórias $I_j, j = 1, \dots, 1000$, são uma amostra aleatória da distribuição de Bernoulli com parâmetro $p = 0, 95$. Portanto, $\hat{p} = \sum_{j=1}^{1000} I_j/1000$ tem distribuição assintótica normal com média p e variância $p(1 - p)/1000$. Logo, \hat{p} está contido no intervalo $[0, 9365; 0, 9635]$ com probabilidade de 95%.

As estimativas das taxas de rejeição sob a hipótese H_0 são obtidas a partir da proporção de amostras em que se rejeita H_0 , sendo os valores verdadeiros dos parâmetros fixados sob H_0 . Para analisar se essa proporção de rejeição é próxima da taxa de rejeição nominal $(0, 05)$, construiremos um intervalo que contenha esta proporção com 95% de probabilidade, supondo que a taxa de rejeição seja igual à nominal $(0, 05)$. Analogamente ao exposto no parágrafo anterior, considerando agora $I_j, j = 1, \dots, 1000$, uma variável aleatória que vale 1 se H_0 foi rejeitada para a j -ésima amostra e 0 caso contrário, temos que $\hat{p} = \sum_{j=1}^{1000} I_j/1000$ está contido no intervalo $[0, 0365; 0, 0635]$ com 95% de probabilidade.

Para estudar o decaimento da variância dos EMV em função de n , faremos 1000 simulações para cada um dos tamanhos amostrais $n = 400, 300, 200, 150, 125, 100, 80, 60, 40, 30$, e estimaremos as variâncias destes estimadores a partir do inverso da matriz observada de Fisher. Para cada intervalo $[30, 80], [80, 150], [150, 400]$ faremos uma regressão linear dos

valores do logaritmo das variâncias estimadas em função do logaritmo de n . Para tamanhos amostrais grandes, espera-se que a inclinação dessa reta de regressão seja próxima de -1 .

Um aspecto fundamental para este estudo de simulação é a geração de amostras aleatórias a partir de determinados valores dos parâmetros. Apresentaremos a seguir um método de obtenção de amostras análogo ao método de geração da distribuição de Poisson baseado em um processo de Poisson Homogêneo (Ross, 1990).

Considere $\{Y(t); t \geq 0\}$ um PPE com seqüência de taxas de transição $\{\lambda(n); n \in \mathbb{N}\}$. Como visto na seção 2.2, a seqüência das variáveis $\{T_n; n \in \mathbb{N}\}$, onde T_n é o tempo de permanência do processo no estado n , é uma seqüência de variáveis aleatórias independentes com T_n exponencialmente distribuída com parâmetro $\lambda(n)$, $n \in \mathbb{N}$. Então, pelo método da função inversa (Ross, 1990), T_n tem a mesma distribuição da variável aleatória

$$\frac{-\log(U)}{\lambda(n)},$$

onde U tem distribuição uniforme no intervalo $(0, 1)$. Logo, como a variável $Y(1)$, que representa o número de ocorrências do evento até o tempo $t = 1$, tem a mesma distribuição da variável aleatória

$$\min_{n \in \mathbb{N}} \{n : T_0 + \dots + T_n > 1\}$$

segue que, se $\{U_n; n \in \mathbb{N}\}$ for uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com distribuição uniforme no intervalo $(0, 1)$, então $Y(1)$ tem a mesma distribuição da variável aleatória

$$\min_{n \in \mathbb{N}} \left\{ n : -\frac{\log(U_0)}{\lambda(0)} - \frac{\log(U_1)}{\lambda(1)} \dots - \frac{\log(U_n)}{\lambda(n)} > 1 \right\}.$$

A partir dessas idéias, um algoritmo para a geração de uma amostra aleatória de $Y(1)$ é dado por

- 1) inicie $soma = 0$ e $n = 0$;
- 2) gere um valor u com distribuição $U(0, 1)$;
- 3) faça $soma = soma - \ln(u)/\lambda(n)$;

- 4) faça $n = n + 1$;
- 5) repetir os passos 2 até 4 enquanto $soma \leq 1$;
- 6) o valor gerado para $Y(1)$ é $n - 1$;
- 7) repetir os passos de 1 a 6 até obter o tamanho amostral desejado.

6.2 Resultados das Simulações para o Modelo PPE2

6.2.1 Probabilidades de Cobertura

Para o modelo PPE2 apresentado no Capítulo 4, aplicamos a metodologia apresentada na seção 6.1 com os valores dos parâmetros fixados em $l_0 = -0,37$ e $l_1 = 1,5$, que foram escolhidos de forma a obter aproximadamente 50% de zeros nas amostras.

Os resultados das simulações para as estimativas das probabilidades de cobertura dos parâmetros l_0 e l_1 estão descritos na Tabela 6.1 abaixo. A partir desta tabela, observamos que os valores dessas estimativas são próximos de 0,95 para todos os tamanhos amostrais.

Tabela 6.1. Estimativas das probabilidades de cobertura do modelo PPE2

n	l_0	l_1
30	0,953	0,952
80	0,949	0,958
150	0,945	0,961
400	0,956	0,940

Conforme o exposto na seção 6.1, as estimativas das probabilidade de cobertura estão contidas no intervalo $[0,9365; 0,9635]$ com probabilidade de 95%, se a probabilidade de cobertura dos intervalos for igual à nominal (0,95). Os dados da Tabela 6.1 podem ser observados juntamente com esses limites na Figura 6.1.

Probabilidades de Cobertura para o modelo PPE2

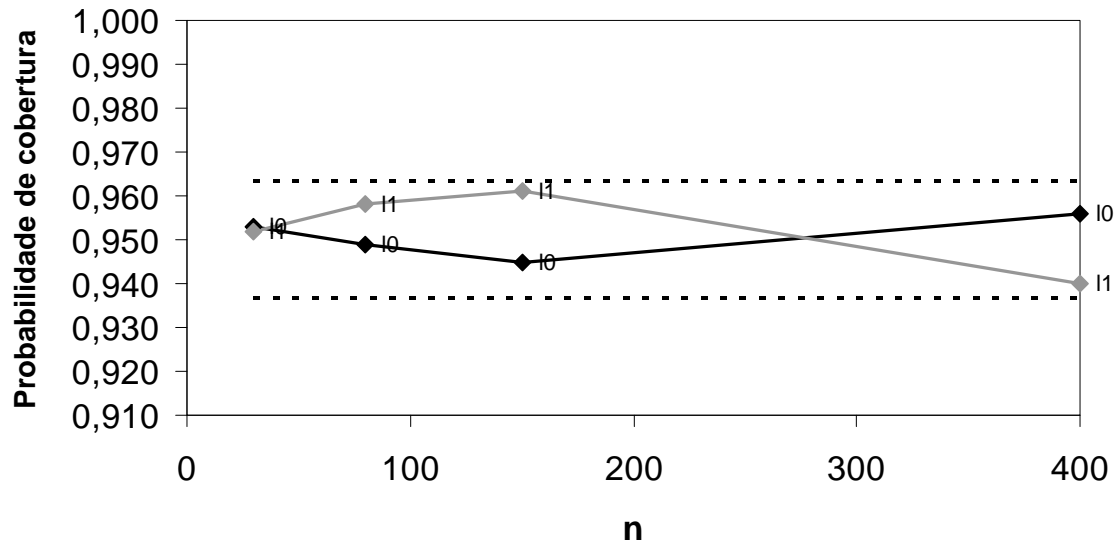


Figura 6.1. Probabilidades de cobertura

Na Tabela 6.2 encontram-se os comprimentos médios dos intervalos de confiança obtidos para cada um dos tamanhos amostrais. Nota-se um decaimento destes comprimentos à medida que o valor de n cresce, indicando uma melhora da precisão da estimação intervalar para amostras grandes.

Tabela 6.2. Comprimento médio dos IC

n	l_0	l_1
30	1,0451	0,8253
80	0,6343	0,4977
150	0,4616	0,3615
400	0,2826	0,2211

A Tabela 6.3 mostra as inclinações da reta de regressão dos valores de $\log(\text{var}(\cdot))$ em função de $\log(n)$. Pela análise dessa tabela concluímos, por exemplo, que para $30 \leq n \leq 80$, $\text{var}(\hat{l}_0) \propto n^{-0,9084}$, correspondendo a uma diferença de aproximadamente 9% em

comparação com as inclinações assintóticas esperadas que são iguais a -1 .

Tabela 6.3. Inclinação da reta de regressão $\log(\text{var}(\cdot))$ em função de $\log(n)$

Intervalos	l_0	l_1
[30; 80]	-0,9084	-0,9122
[80; 150]	-1,0740	-1,0811
[150; 400]	-0,7087	-0,7097

6.2.2 Taxas de Rejeição para o Teste de Hipóteses

Considerando o teste $H_0 : l_0 = l_1$ para os parâmetros do modelo PPE2, aplicamos a metodologia da seção 6.1 para estimar a taxa de rejeição deste teste com o valor dos parâmetros fixados sob H_0 como $l_0 = l_1 = -0,5$. De acordo com o exposto na seção 6.1, as taxas de rejeição observadas estão contidas no intervalo $[0,0365; 0,0635]$ com probabilidade de 95%, se a taxa de rejeição de H_0 for igual à nominal (0,05). Os resultados das simulações estão dispostos na Tabela 6.4, onde nota-se que as taxas de rejeição se encontram dentro deste intervalo para os tamanhos amostrais 30 e 400.

Tabela 6.4. Estimativas das taxas de rejeição para a hipótese H_0

n	30	80	150	400
taxa	0,053	0,031	0,033	0,042

6.3 Resultados das Simulações para o Modelo PPE4

6.3.1 Probabilidades de Cobertura

Para o modelo PPE4 apresentado no Capítulo 5 aplicamos a metodologia da seção 6.1, com 1000 simulações para cada um dos tamanhos amostrais $n = 30, 80, 150, 400$ e o valor dos parâmetros fixados em $l_0 = -0,37$, $l_a = -0,5$ e $l_b = 0,5$. Os resultados das simulações estão descritos na Tabela 6.5 abaixo.

Tabela 6.5. Estimativas das probabilidades de cobertura do modelo PPE4

n	l_0	l_a	l_b
30	0,921	0,979	0,978
80	0,947	0,962	0,965
150	0,945	0,966	0,970
400	0,944	0,967	0,971

Como citado na seção anterior, as estimativas das probabilidades de cobertura estão dentro do intervalo $[0,9365; 0,9635]$ com probabilidade de 95%, se a probabilidade de cobertura dos intervalos for igual à nominal (0,95). Os dados da Tabela 6.5 podem ser observados juntamente com os limites 0,9365 e 0,9635 na Figura 6.2. A partir desta, observamos uma sobre cobertura dos intervalos de confiança dos parâmetros l_a e l_b , no entanto esta sobre cobertura não ultrapassou 0,02. Para o parâmetro l_0 observa-se uma subcobertura dos intervalos de confiança apenas para $n = 30$.

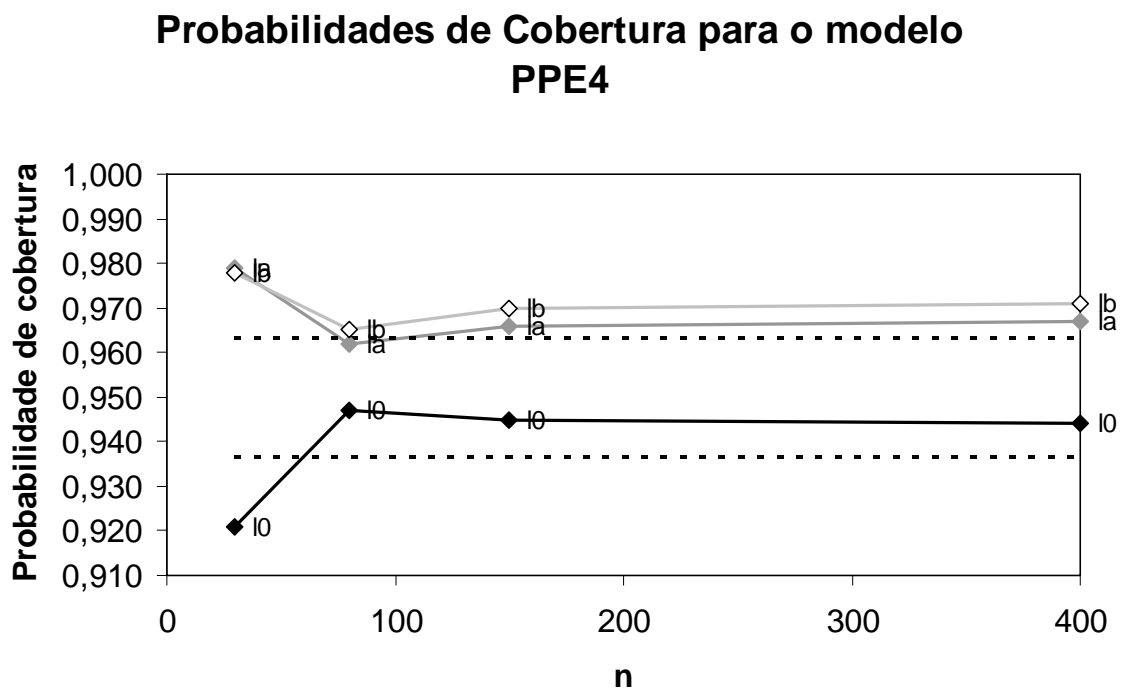


Figura 6.2. Probabilidades de cobertura

A Tabela 6.6 mostra os comprimentos médios dos intervalos de confiança para cada um dos tamanhos amostrais. Nota-se um decaimento destes comprimentos à medida que o valor de n cresce, mas observa-se comprimentos de intervalo muito grandes para os parâmetros l_a e l_b para tamanhos amostrais pequenos ou moderados (menor que 150), isto é, nestes casos os intervalos obtidos são não informativos.

Tabela 6.6. Comprimento médios dos IC

n	l_0	l_a	l_b
30	1,0483	223,0547	320,5744
80	0,6373	67,6354	96,7278
150	0,4640	24,7412	28,5622
400	0,2835	2,1481	3,4165

A Tabela 6.7 mostra as inclinações da reta de regressão dos valores de $\log(\text{var}(\cdot))$ em função de $\log(n)$. Pela análise dessa tabela concluímos, por exemplo, que para $30 \leq n \leq 80$, $\text{var}(\hat{l}_a) \propto n^{-2,1481}$. Notamos também que as inclinações das retas de regressão estão próximas de -1 para o parâmetro l_0 , enquanto que para os parâmetros l_a e l_b isso não ocorre, apesar dessas inclinações decaírem à medida que n cresce.

Tabela 6.7. Inclinação da reta de regressão $\log(\text{var}(\cdot))$ sobre $\log(n)$

Intervalos	l_0	l_a	l_b
[30; 80]	-1,0148	-2,1481	-2,1335
[80; 150]	-0,9992	-2,1162	-2,0962
[150; 400]	-1,0019	-1,5784	-1,5631

6.3.2 Taxas de Rejeição dos Testes de Hipóteses

Considerando o teste $H_0 : l_0 = l_a + \log(e^{l_b} - 1)$, $l_b > 0$ para os parâmetros do modelo PPE4, aplicamos a metodologia da seção 6.1 para estimar a taxa de rejeição sob a hipótese H_0 com os valores dos parâmetros fixados em $l_0 = -0,37$, $l_a = -0,5$, $l_b = 0,5$. De acordo com o exposto na seção 6.1, as taxas de rejeição observadas estão contidas no intervalo $[0,0365; 0,0635]$ com probabilidade de 95%, se a taxa de rejeição de H_0 for igual à nominal (0,05). Os resultados das simulações estão dispostos na Tabela 6.8, onde nota-

se que as taxas de rejeição se encontram dentro do intervalo $[0,0365; 0,0635]$ para todos os tamanhos amostrais exceto para $n = 30$, mostrando uma adequação desta metodologia para tamanhos amostrais moderados ou grandes.

Tabela 6.8. Estimativas das taxas de rejeição para a hipótese H_0

n	30	80	150	400
taxa	0,017	0,041	0,058	0,058

Capítulo 7

Inferência Bayesiana

Neste capítulo abordaremos, sob o enfoque bayesiano, os modelos discutidos nos capítulos 4 e 5. Adotaremos *prioris* normais independentes para os parâmetros l_0 e l_1 do modelo PPE2 e l_0 , l_a e l_b do modelo PPE4. Determinaremos a distribuição *a posteriori* para cada um dos modelos e as estimativas dos resumos (média, mediana, moda, 1º e 3º quartis e intervalo de credibilidade de 95%) da distribuição *a posteriori*, em cada caso, serão obtidas segundo um procedimento MCMC (Monte Carlo Markov Chain), denominado Passeio Aleatório (Gilks, 1995). As convergências das cadeias para a distribuição de equilíbrio são monitoradas pelo critério de Gelman-Rubin (Gelman, 1995).

7.1 Modelo PPE2

Nesta seção proporemos uma abordagem bayesiana para o modelo PPE2 com o uso de *prioris* normais independentes para os parâmetros l_0 e l_1 , isto é, suporemos que as funções densidade *a priori* para l_0 e l_1 são dadas por

$$\pi(l_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{\frac{-1}{2\sigma_0^2} (l_0 - \mu_0)^2\right\}, \quad (7.1)$$

$$\pi(l_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{\frac{-1}{2\sigma_1^2} (l_1 - \mu_1)^2\right\}, \quad (7.2)$$

onde μ_0, μ_1 são números reais e σ_0, σ_1 são números reais estritamente positivos. Os hiperparâmetros do modelo, μ_0, μ_1, σ_0 e σ_1 , são quantidades conhecidas.

Considere, como na seção 4.2, Y_1, Y_2, \dots, Y_n uma amostra aleatória de tamanho n de

$Y(1)$. Combinando a função de verossimilhança 4.7 com as densidades *a priori* 7.1 e 7.2, obtemos a função densidade *a posteriori* conjunta de l_0 e l_1 , dada por

$$\begin{aligned} \pi(l_0, l_1 | y_1, \dots, y_n) &\propto L(l_0, l_1) \pi(l_0) \pi(l_1) \\ &\propto \exp \left\{ -ne^{l_0} + (n - n_0)l_0 + [\sum_{i=1}^n y_i - (n - n_0)]l_1 - \frac{(l_0 - \mu_0)^2}{2\sigma_0^2} - \frac{(l_1 - \mu_1)^2}{2\sigma_1^2} \right\} \\ &\quad \prod_{i: y_i > 0} \int_0^1 x^{y_i - 1} \exp\{-(e^{l_1} - e^{l_0})x\} dx. \end{aligned} \quad (7.3)$$

Para gerar amostras da densidade *a posteriori* 7.3 utilizaremos um procedimento de simulação MCMC, denominado Passeio Aleatório, que é um caso particular do algoritmo de Metropolis. Considere N o número de simulações a serem realizadas, $\theta = (l_0, l_1)$ o vetor de parâmetros e $\theta^{(t)} = (l_0^{(t)}, l_1^{(t)})$ o valor de θ após t iterações da cadeia, $t = 1, \dots, N$. O estado inicial da cadeia será denotado por $\theta^{(0)} = (l_0^{(0)}, l_1^{(0)})$. Cada iteração do algoritmo consistirá em duas etapas: a primeira atualiza o valor de $l_0^{(t)}$ e a segunda atualiza o valor de $l_1^{(t)}$, $t = 0, 1, \dots, N - 1$. Na primeira etapa os candidatos a transição são gerados a partir de uma distribuição normal com média $l_0^{(t)}$ e desvio-padrão τ_0 , enquanto que na segunda etapa os candidatos são gerados a partir de uma distribuição normal com média $l_1^{(t)}$ e desvio-padrão τ_1 , $t = 0, 1, \dots, N - 1$. Os valores de τ_0 e τ_1 , fixados antes de iniciar o algoritmo, foram escolhidos empiricamente de forma que a convergência da cadeia para a distribuição de equilíbrio não fosse muita lenta. Um candidato θ' a transição é aceito com probabilidade $\alpha\{\theta^{(t)}, \theta'\}$ dada por

$$\alpha\{\theta^{(t)}, \theta'\} = \min\left\{1, \frac{\pi(\theta')}{\pi(\theta^{(t)})}\right\},$$

onde $\theta^{(t)}$ é o estado atual da cadeia. Os seguintes passos são realizados pelo algoritmo:

- 1) iniciar $t = 0$;
- 2) gerar um candidato a transição z_0 para l_0 a partir de uma distribuição $N(l_0^{(t)}, \tau_0^2)$;
- 3) gerar u de uma distribuição $U(0, 1)$;

4) se $u \leq \alpha\{(l_0^{(t)}, l_1^{(t)}), (z_0, l_1^{(t)})\}$ então aceitar $(z_0, l_1^{(t)})$, isto é, fazer $l_0^{(t+1)} = z_0$, onde

$$\alpha\{(l_0^{(t)}, l_1^{(t)}), (z_0, l_1^{(t)})\} = \min \left\{ 1, \frac{\pi(z_0, l_1^{(t)} | y_1, \dots, y_n)}{\pi(l_0^{(t)}, l_1^{(t)} | y_1, \dots, y_n)} \right\},$$

caso contrário, fazer $l_0^{(t+1)} = l_0^{(t)}$;

5) gerar um candidato a transição z_1 para l_1 a partir de uma distribuição $N(l_1^{(t)}, \tau_1^2)$;

6) gerar u de uma distribuição $U(0, 1)$;

7) se $u \leq \alpha\{(l_0^{(t+1)}, l_1^{(t)}), (l_0^{(t+1)}, z_1)\}$ então aceitar $(l_0^{(t+1)}, z_1)$, isto é, fazer $l_1^{(t+1)} = z_1$, onde

$$\alpha\{(l_0^{(t+1)}, l_1^{(t)}), (l_0^{(t+1)}, z_1)\} = \min \left\{ 1, \frac{\pi(l_0^{(t+1)}, z_1 | y_1, \dots, y_n)}{\pi(l_0^{(t+1)}, l_1^{(t)} | y_1, \dots, y_n)} \right\},$$

caso contrário, fazer $l_1^{(t+1)} = l_1^{(t)}$;

8) Repetir os passos 2 a 7 até $t = N$.

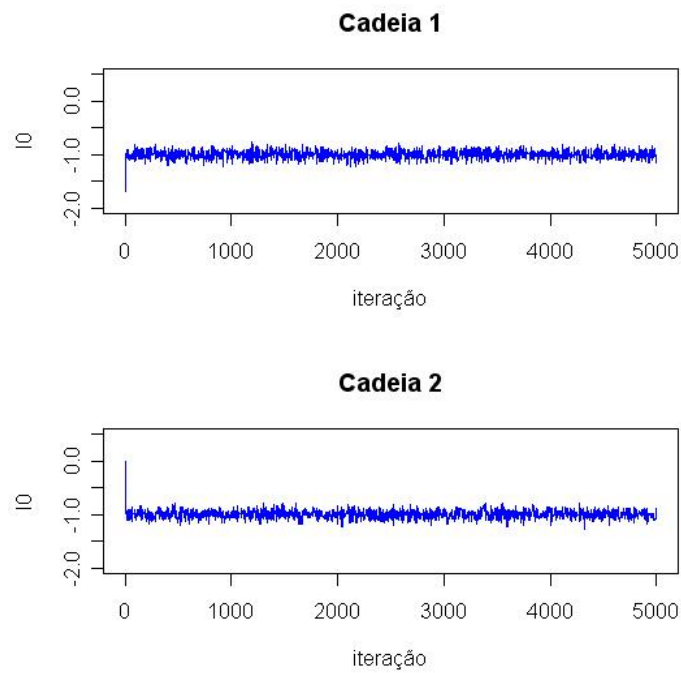
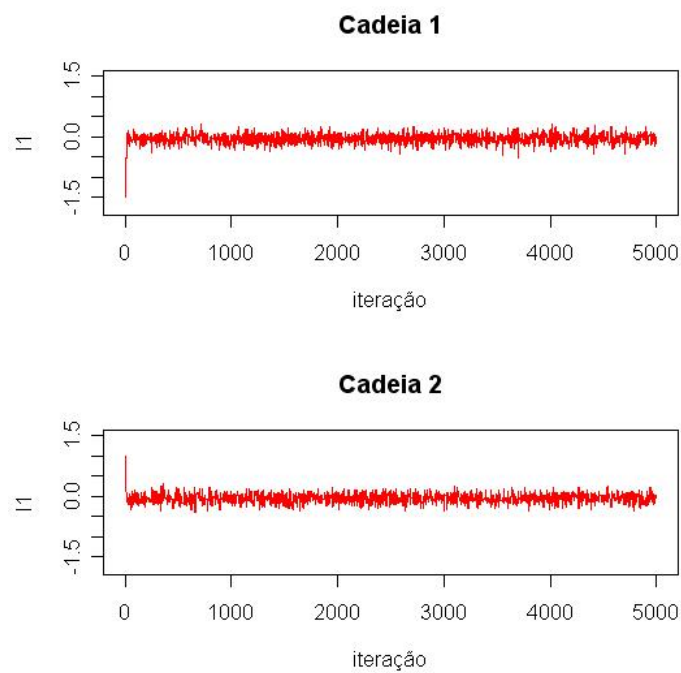
7.1.1 Prioris Não Informativas

Com o intuito de trabalharmos com *prioris* não informativas, consideraremos os seguintes valores para os hiperparâmetros

$$\mu_0 = \mu_1 = 0, \tag{7.4}$$

$$\sigma_0 = \sigma_1 = 20.$$

Usando o algoritmo descrito acima para os dados da Tabela 4.1 geramos duas cadeias com 5000 observações cada uma. Os gráficos de convergência para as duas cadeias são apresentados nas Figuras 7.1 e 7.2

Figura 7.1. Gráficos de convergência para l_0 .Figura 7.2. Gráficos de convergência para l_1 .

A convergência deste procedimento foi monitorada pelo critério de Gelman-Rubin (Gelman, 1995). Usando este critério obtivemos os gráficos sequenciais para os valores do

fator de redução $\sqrt{\widehat{R}}$ de l_0 e l_1 apresentados nas Figuras 7.3 e 7.4, respectivamente. Por estes gráficos notamos que o valor do fator $\sqrt{\widehat{R}}$ se aproxima de 1 à medida que a cadeia evolui, indicando a convergência da cadeia para a distribuição de equilíbrio.

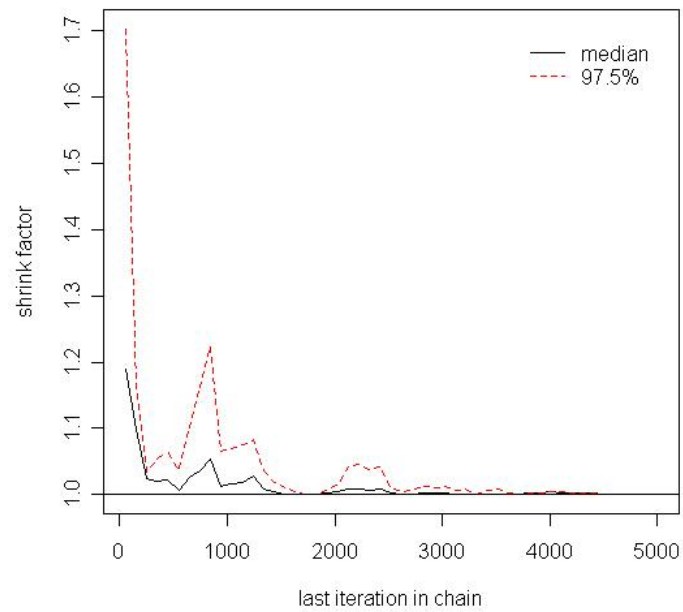


Figura 7.3. Gráfico do Fator de Redução $\sqrt{\widehat{R}}$ para l_0

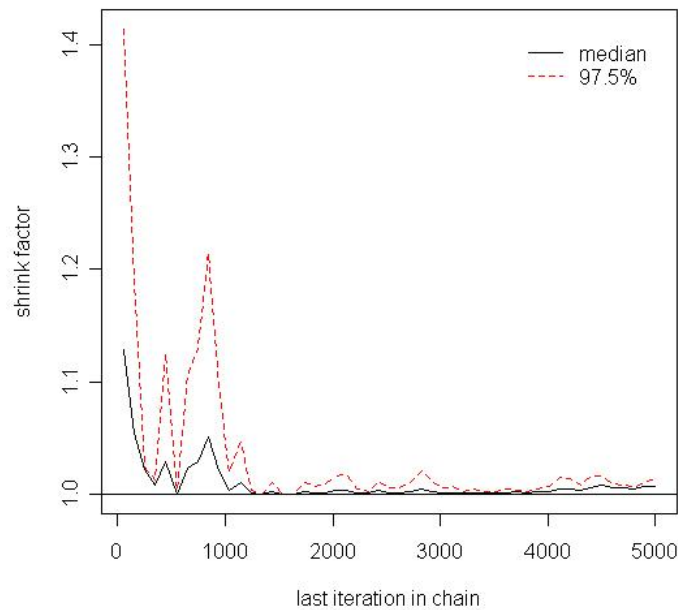


Figura 7.4. Gráfico do Fator de Redução $\sqrt{\widehat{R}}$ para l_1

Utilizando as informações obtidas a partir das duas cadeias e adotando um período de aquecimento de 300 iterações, salto de 20 iterações para as amostras de l_0 e 10 iterações para as amostras de l_1 , obtivemos estimativas dos resumos das funções densidades *a posteriori* de l_0 e l_1 , que são apresentadas na Tabela 7.1. Nota-se a partir desta tabela que as estimativas das médias *a posteriori* e dos intervalos de credibilidade de 95% são muito similares aos resultados obtidos via inferência clássica e apresentados na Tabela 4.2.

Tabela 7.1. Estimativas dos resumos das funções de densidade *a posteriori* de l_0 e l_1

	1º Quartil	Média	Mediana	Moda	3º Quartil	IC 95%
l_0	-1,0450	-0,9968	-0,9973	-0,9962	-0,9498	[-1,1450; -0,8586]
l_1	-0,1159	-0,0397	-0,0367	-0,0328	0,0362	[-0,2587; 0,1814]

A partir destas simulações também obtivemos estimativas das funções de densidades *a posteriori* de l_0 e l_1 , cujos gráficos são apresentados abaixo.

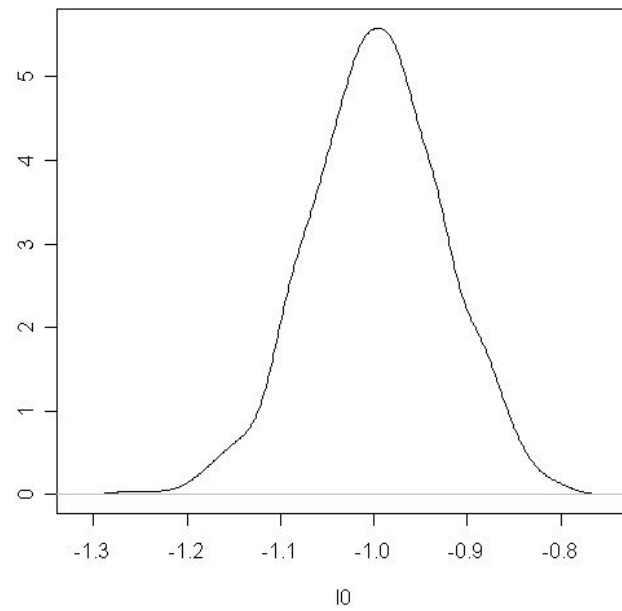


Figura 7.5 Gráfico da estimativa da função de densidade *a posteriori* de l_0

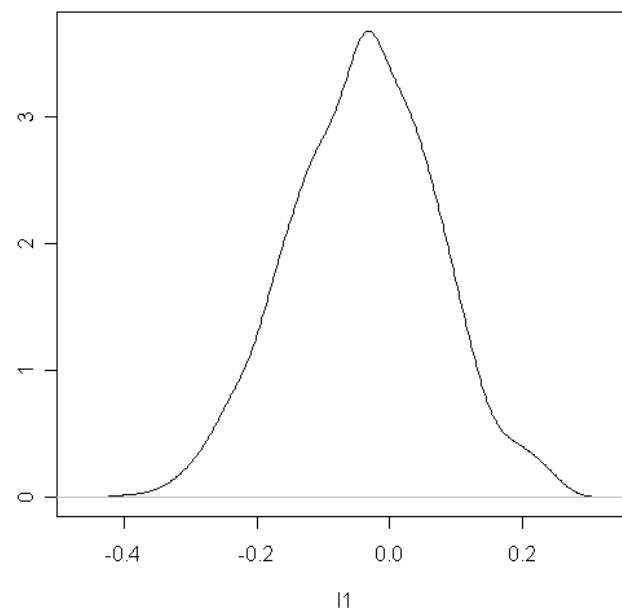


Figura 7.6. Gráfico da estimativa da função de densidade *a posteriori* de l_1

7.1.2 Prioris Informativas

Os resultados obtidos na seção anterior, com o uso de *prioris* não informativas, mostraram-se concordantes com os resultados obtidos via inferência clássica do Capítulo 4. Essa concordância significa que os procedimentos clássico e bayesiano com *prioris* não informativas são equivalentes. Para verificar a influência ou não das *prioris* em nossa análise variamos o valor dos hiperparâmetros σ_0 e σ_1 , de modo a tornar estas *prioris* mais informativas. Para a geração das amostras *a posteriori*, usamos os dados da Tabela 4.1 e a mesma metodologia da seção anterior, com duas cadeias de tamanho 10.000, um período de aquecimento de 300 iterações e salto de 20 iterações. A convergência para a distribuição de equilíbrio foi monitorada, como anteriormente, pelo critério de Gelman-Rubin. Os resultados são apresentados nas Tabelas 7.2 e 7.3.

Tabela 7.2. Estimativas dos resumos da função densidade *a posteriori* de l_0

σ_0	1º Quartil	Média	Mediana	Moda	3º Quartil	IC 95%
1	-1,0390	-0,9906	-0,9912	-1,0036	-0,9414	[-1,1274; -0,8547]
3	-1,0460	-0,9979	-0,9970	-1,0001	-0,9487	[-1,1367; -0,8692]
5	-1,0510	-0,9996	-1,0010	-1,0083	-0,9498	[-1,1445; -0,8638]
10	-1,0520	-1,0010	-1,0010	-0,9972	-0,9528	[-1,1363; -0,8711]

Tabela 7.3. Estimativas dos resumos da função densidade *a posteriori* de l_1

σ_1	1º Quartil	Média	Mediana	Moda	3º Quartil	IC 95%
1	-0,1212	-0,0457	-0,0450	-0,0541	0,0283	[-0,2512; 0,1683]
3	-0,1095	-0,0419	-0,0411	-0,0471	0,0291	[-0,2563; 0,1600]
5	-0,1209	-0,0512	-0,0457	-0,0250	0,0191	[-0,2649; 0,1521]
10	-0,1176	-0,0477	-0,0452	-0,0215	0,0278	[-0,2621; 0,1527]

Pela análise destes resultados, notamos que houve pouca influência dos valores dos hiperparâmetros σ_0 e σ_1 nas estimativas dos resumos da função densidade *a posteriori* dos parâmetros l_0 e l_1 .

7.2 Modelo PPE4

Nesta seção proporemos uma abordagem bayesiana para o modelo PPE4 estudado no Capítulo 5 supondo que l_0 , l_a e l_b são *a priori* independentes com funções densidade *a priori* dadas por

$$\pi(l_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{\frac{-1}{2\sigma_0^2} (l_0 - \mu_0)^2\right\}, \quad (7.5)$$

$$\pi(l_a) = \frac{1}{\sqrt{2\pi}\sigma_a} \exp\left\{\frac{-1}{2\sigma_a^2} (l_a - \mu_a)^2\right\}, \quad (7.6)$$

$$\pi(l_b) = \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left\{\frac{-1}{2\sigma_b^2} (l_b - \mu_b)^2\right\}, \quad (7.7)$$

onde μ_0, μ_a, μ_b são números reais e $\sigma_0, \sigma_a, \sigma_b$ são números reais estritamente positivos. Os hiperparâmetros do modelo, $\mu_0, \mu_a, \mu_b, \sigma_0, \sigma_a, \sigma_b$, são conhecidos.

Considere, como na seção 5.2, Y_1, Y_2, \dots, Y_n uma amostra aleatória de tamanho n de $Y(1)$. Combinando a função de verossimilhança dada em 5.6 com as densidade *a priori* dadas em 7.5, 7.6 e 7.7, obtemos a distribuição *a posteriori* dos parâmetros dada por

$$\begin{aligned} \pi(l_0, l_a, l_b | y_1, \dots, y_n) &\propto L(l_0, l_a, l_b) \pi(l_0) \pi(l_a) \pi(l_b) \\ &\propto \exp\left\{-ne^{l_0} + (n - n_0)(l_0 - l_a) - \frac{(l_0 - \mu_0)^2}{2\sigma_0^2} - \frac{(l_a - \mu_a)^2}{2\sigma_a^2} - \frac{(l_b - \mu_b)^2}{2\sigma_b^2}\right\} \times \\ &\quad \prod_{i: y_i > 0} \binom{e^{l_b} + y_i - 2}{y_i - 1} \int_{\exp(-e^{l_a})}^1 x^{e^{l_b} - e^{(l_0 - l_a)} - 1} (1 - x)^{y_i - 1} dx. \end{aligned} \quad (7.8)$$

Como na seção anterior utilizamos um Passeio Aleatório (Gilks, 1995) para gerar amostras da função densidade *a posteriori* 7.8. Considere N o número de simulações a serem realizadas, $\theta = (l_0, l_a, l_b)$ o vetor de parâmetros e $\theta^{(t)} = (l_0^{(t)}, l_a^{(t)}, l_b^{(t)})$ o valor de θ após t iterações, $t = 1, \dots, N$. O estado inicial da cadeia será denotado por $\theta^{(0)} = (l_0^{(0)}, l_a^{(0)}, l_b^{(0)})$. Cada iteração do algoritmo consistirá em três etapas: a primeira atualiza o valor de $l_0^{(t)}$, a segunda atualiza o valor de $l_a^{(t)}$ e a terceira atualiza o valor de $l_b^{(t)}$, $t = 0, 1, \dots, N - 1$. Na primeira etapa os candidatos a transição são gerados a partir de uma distribuição normal com média $l_0^{(t)}$ e desvio-padrão τ_0 , enquanto que na segunda etapa os candidatos são gerados a partir de uma distribuição normal com média $l_a^{(t)}$ e desvio-padrão τ_a e, finalmente, na terceira etapa os candidatos são gerados a partir de uma distribuição

normal com média $l_b^{(t)}$ e desvio-padrão τ_b , $t = 0, 1, \dots, N - 1$. Os valores de τ_0 , τ_a e τ_b , determinados antes de se iniciar o algoritmo, foram escolhidos empiricamente de forma que a convergência da cadeia para a distribuição de equilíbrio não fosse muita lenta. Um candidato θ' a transição é aceito com probabilidade $\alpha\{\theta^{(t)}, \theta'\}$ dada por

$$\alpha\{\theta^{(t)}, \theta'\} = \min\left\{1, \frac{\pi(\theta')}{\pi(\theta^{(t)})}\right\},$$

onde $\theta^{(t)}$ é o estado atual da cadeia. Os seguintes passos serão realizados pelo algoritmo:

- 1) iniciar $t = 0$;
- 2) gerar um candidato a transição z_0 para l_0 a partir de uma distribuição $N(l_0^{(t)}, \tau_0^2)$;
- 3) gerar u de uma distribuição $U(0, 1)$;
- 4) se $u \leq \alpha\{(l_0^{(t)}, l_a^{(t)}, l_b^{(t)}), (z_0, l_a^{(t)}, l_b^{(t)})\}$ então aceitar $(z_0, l_a^{(t)}, l_b^{(t)})$, isto é, fazer $l_0^{(t+1)} = z_0$, onde

$$\alpha\{(l_0^{(t)}, l_a^{(t)}, l_b^{(t)}), (z_0, l_a^{(t)}, l_b^{(t)})\} = \min\left\{1, \frac{\pi(z_0, l_a^{(t)}, l_b^{(t)} | y_1, \dots, y_n)}{\pi(l_0^{(t)}, l_a^{(t)}, l_b^{(t)} | y_1, \dots, y_n)}\right\},$$

caso contrário, fazer $l_0^{(t+1)} = l_0^{(t)}$;

- 5) gerar um candidato a transição z_a para l_a a partir de uma distribuição $N(l_a^{(t)}, \tau_a^2)$;
- 6) gerar u de uma distribuição $U(0, 1)$;
- 7) se $u \leq \alpha\{(l_0^{(t+1)}, l_a^{(t)}, l_b^{(t)}), (l_0^{(t+1)}, z_a, l_b^{(t)})\}$ então aceitar $(l_0^{(t+1)}, z_a, l_b^{(t)})$, isto é, fazer $l_a^{(t+1)} = z_a$, onde

$$\alpha\{(l_0^{(t+1)}, l_a^{(t)}, l_b^{(t)}), (l_0^{(t+1)}, z_a, l_b^{(t)})\} = \min\left\{1, \frac{\pi(l_0^{(t+1)}, z_a, l_b^{(t)} | y_1, \dots, y_n)}{\pi(l_0^{(t+1)}, l_a^{(t)}, l_b^{(t)} | y_1, \dots, y_n)}\right\},$$

caso contrário, fazer $l_a^{(t+1)} = l_a^{(t)}$;

- 8) gerar um candidato a transição z_b para l_b a partir de uma distribuição $N(l_b^{(t)}, \tau_b^2)$;
- 9) gerar u de uma distribuição $U(0, 1)$;

- 10) se $u \leq \alpha\{(l_0^{(t+1)}, l_a^{(t+1)}, l_b^{(t)}), (l_0^{(t+1)}, l_a^{(t+1)}, z_b)\}$ então aceitar $(l_0^{(t+1)}, l_a^{(t+1)}, z_b)$, isto é, fazer $l_b^{(t+1)} = z_b$, onde

$$\alpha\{(l_0^{(t+1)}, l_a^{(t+1)}, l_b^{(t)}), (l_0^{(t+1)}, l_a^{(t+1)}, z_b)\} = \min \left\{ 1, \frac{\pi(l_0^{(t+1)}, l_a^{(t+1)}, z_b | y_1, \dots, y_n)}{\pi(l_0^{(t+1)}, l_a^{(t+1)}, l_b^{(t)} | y_1, \dots, y_n)} \right\},$$

caso contrário, fazer $l_b^{(t+1)} = l_b^{(t)}$;

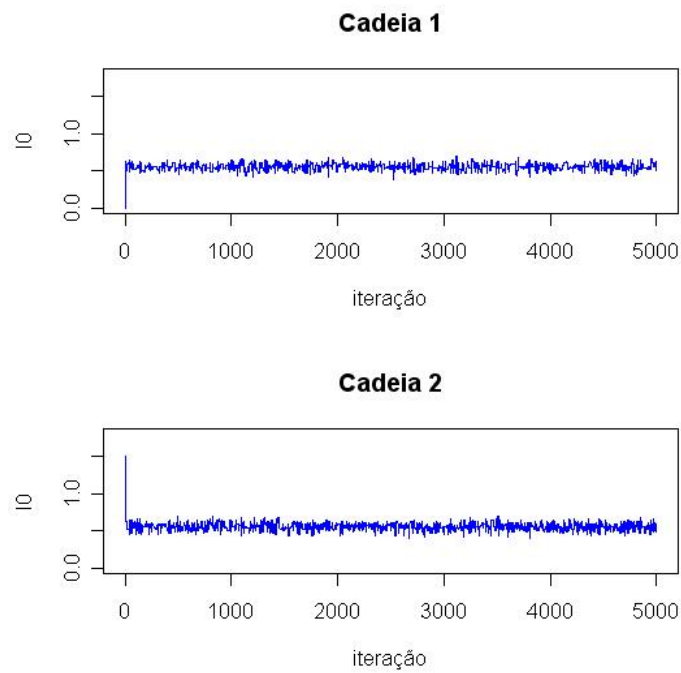
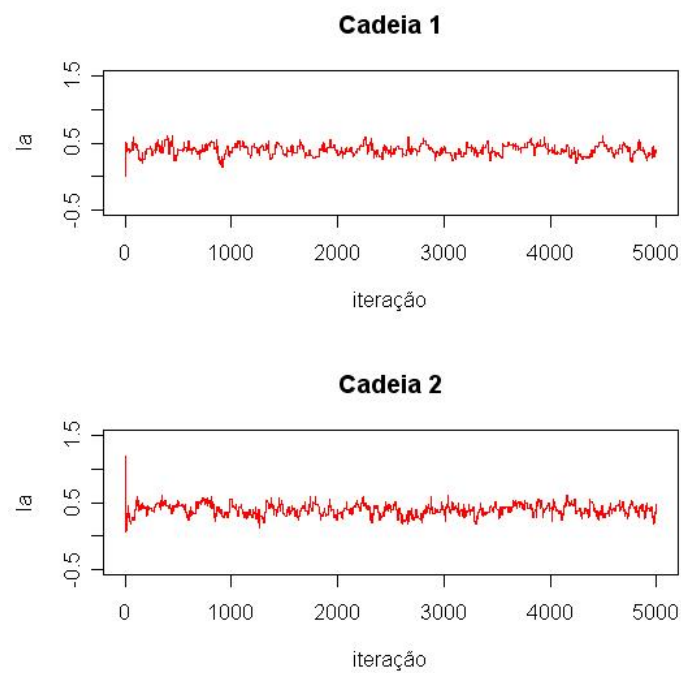
- 11) Repetir os passos 2 a 10 até $t = N$.

7.2.1 Prioris Não Informativas

Com o objetivo de trabalharmos com prioris não informativas, consideraremos os seguintes valores para os hiperparâmetros

$$\begin{aligned} \mu_0 &= \mu_a = \mu_b = 0, \\ \sigma_0 &= \sigma_a = \sigma_b = 20. \end{aligned}$$

Usando os dados da Tabela 5.1 e o algoritmo descrito acima com pontos iniciais superdispersos geramos duas cadeias com $N = 5000$ observações cada uma. Os gráficos de convergência para l_0 , l_a e l_b são apresentados nas Figuras 7.7, 7.8 e 7.9. A análise deste gráficos indica que as duas cadeias convergiram para uma mesma região do espaço paramétrico, apesar da superdispersão dos pontos iniciais.

Figura 7.7. Gráficos de convergência para l_0 .Figura 7.8. Gráficos de convergência para l_a .

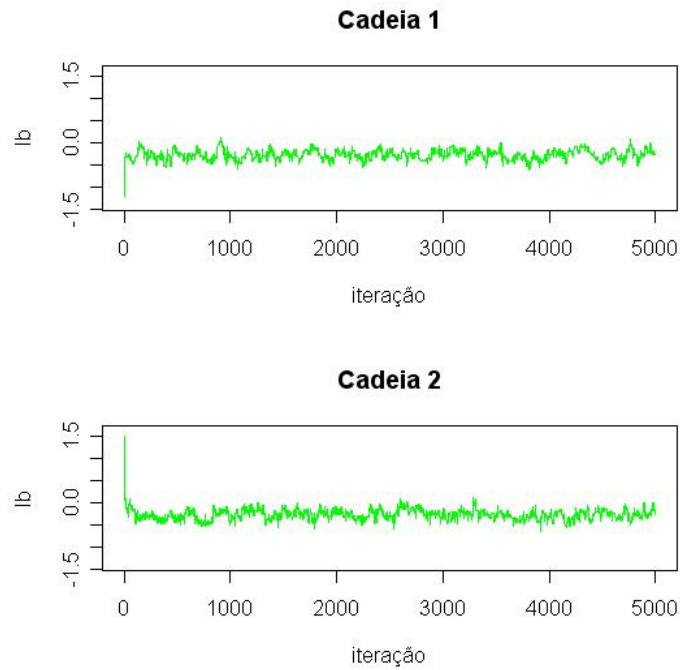
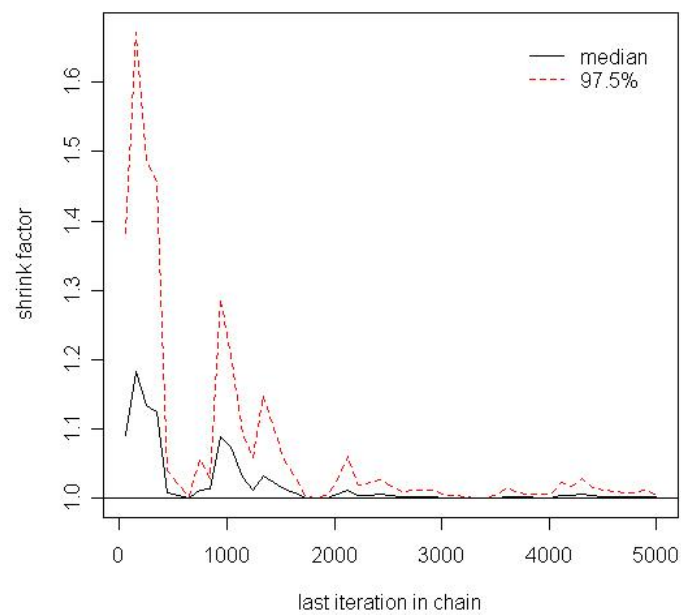
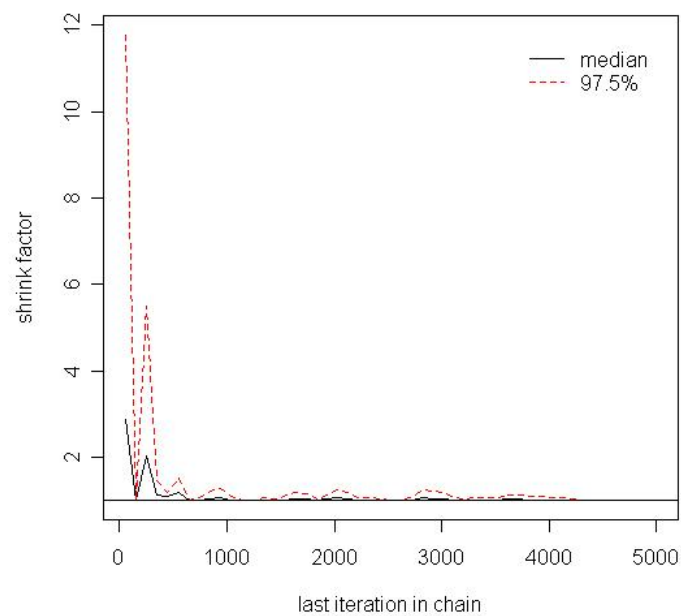


Figura 7.9. Gráficos de convergência para l_b

A convergência deste procedimento também foi monitorada pelo critério de convergência de Gelman-Rubin (Gelman, 1995). Usando este critério obtivemos os gráficos sequenciais para os valores do fator de redução $\sqrt{\widehat{R}}$ (Gelman, 1995) de l_0 , l_a e l_b apresentados nas Figuras 7.10, 7.11 e 7.12, respectivamente. A partir destes gráficos nota-se que o valor do fator $\sqrt{\widehat{R}}$ se aproxima de 1 à medida que a cadeia evolui, indicando a convergência da cadeia para a distribuição de equilíbrio.

Figura 7.10. Gráfico do Fator de Redução para l_0 Figura 7.11. Gráfico do Fator de Redução para l_a

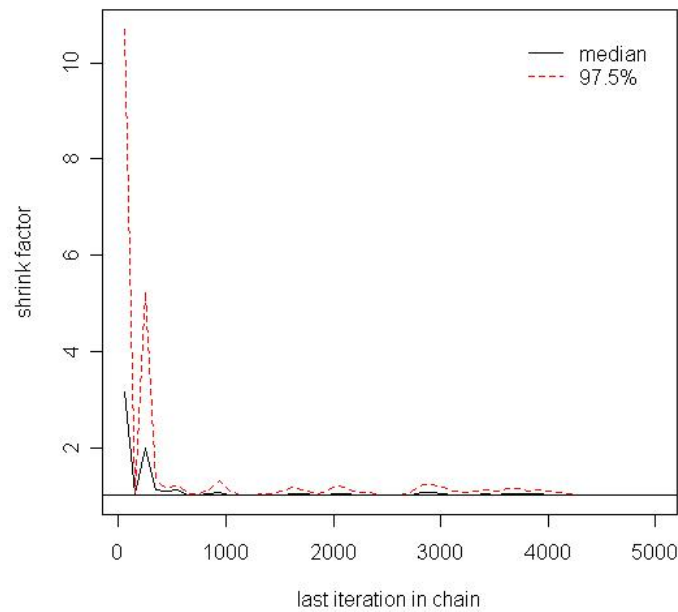


Figura 7.12. Gráfico do Fator de Redução para l_b

Utilizando as informações obtidas a partir das duas cadeias, adotando um período de aquecimento de 500 iterações, salto de 15 iterações para o parâmetro l_0 e de 40 iterações para os parâmetros l_a e l_b , obtivemos estimativas dos resumos das funções densidade *a posteriori* de l_0 , l_a e l_b , que são apresentadas na Tabela 7.4. Nota-se a partir desta tabela que as estimativas das médias *a posteriori* e os intervalos de credibilidade de 95% são muito similares às estimativas de máxima verossimilhança e intervalos de confiança obtidos via inferência clássica e apresentados na Tabela 5.2.

Tabela 7.4. Estimativas dos resumos das funções densidade *a posteriori* de l_0 , l_a e l_b

	1º Quartil	Média	Mediana	Moda	3º Quartil	IC 95%
l_0	0,5248	0,5554	0,5544	0,5516	0,5863	[0,4668; 0,6505]
l_a	0,3424	0,3920	0,3887	0,3740	0,4429	[0,2270; 0,5307]
l_b	-0,3495	-0,2782	-0,2777	-0,2724	-0,2055	[-0,5051; -0,0541]

A partir destas simulações também obtivemos estimativas das funções densidade *a posteriori* de l_0 , l_a e l_b , cujos gráficos são dados na seqüência.

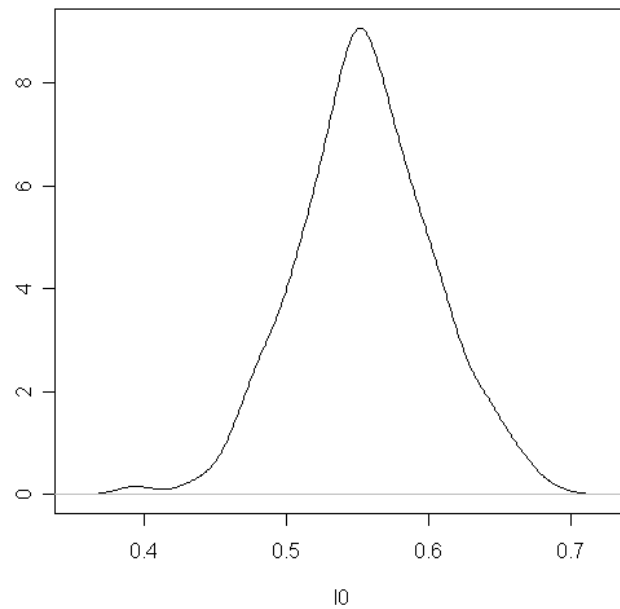


Figura 7.13. Gráfico da estimativa da função densidade *a posteriori* de l_0

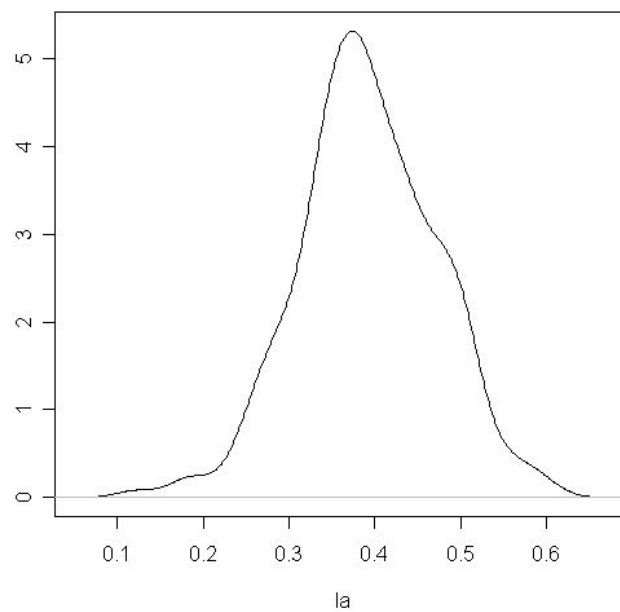


Figura 7.14. Gráfico da estimativa da função densidade *a posteriori* de l_a

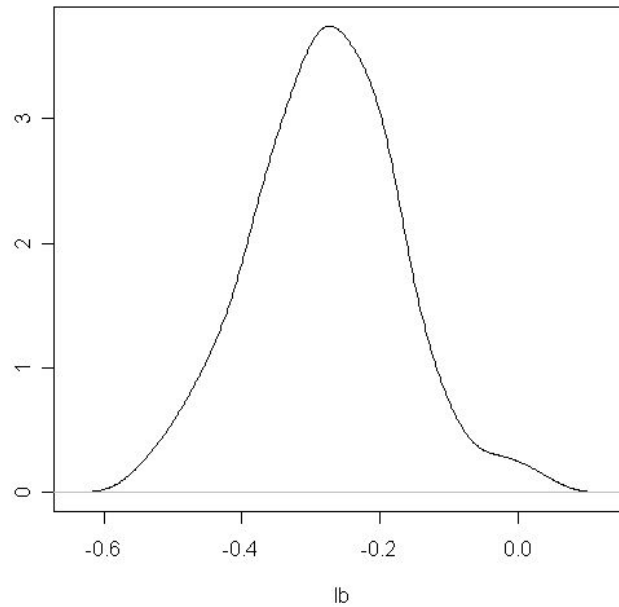


Figura 7.15. Gráfico da estimativa da função densidade *a posteriori* de l_b

7.2.2 Prioris Informativas

Os resultados obtidos na seção anterior, com o uso de *prioris* não informativas, são próximos dos resultados obtidos via inferência clássica do Capítulo 5. Essa proximidade implica que os procedimentos clássico e bayesiano com *prioris* não informativas são equivalentes. Para verificar a influência ou não de *prioris* em nossa análise variamos o valor dos hiperparâmetros σ_0 , σ_a e σ_b , de modo a tornar estas *prioris* mais informativas. Para a geração das amostras *a posteriori*, usamos os dados da Tabela 5.1 e a mesma metodologia aplicada na seção anterior, com duas cadeias de tamanho 10.000, um período de aquecimento de 500 iterações e salto de 40 iterações. Os resultados são apresentados nas Tabelas 7.5, 7.6 e 7.7.

Tabela 7.5. Estimativas dos resumos das funções densidade *a posteriori* de l_0

σ_0	1º Quartil	Média	Mediana	Moda	3º Quartil	IC 95%
1	0,5205	0,5510	0,5520	0,5511	0,5818	[0,4584; 0,6370]
3	0,5237	0,5547	0,5556	0,5578	0,5847	[0,4648; 0,6432]
5	0,5257	0,5552	0,5567	0,5672	0,5825	[0,4644; 0,6456]
10	0,5271	0,5561	0,5547	0,5425	0,5856	[0,4638; 0,6507]

Tabela 7.6. Estimativas dos resumos das funções densidade *a posteriori* de l_a

σ_a	1º Quartil	Média	Mediana	Moda	3º Quartil	IC 95%
1	0,3450	0,3917	0,3968	0,3911	0,4483	[0,2069; 0,5298]
3	0,3466	0,3934	0,3947	0,3738	0,4436	[0,2534; 0,5252]
5	0,3415	0,3971	0,4000	0,4039	0,4537	[0,2436; 0,5233]
10	0,3498	0,3981	0,3970	0,3981	0,4441	[0,2665; 0,5340]

Tabela 7.7. Estimativas dos resumos das funções densidade *a posteriori* de l_b

σ_b	1º Quartil	Média	Mediana	Moda	3º Quartil	IC 95%
1	-0,3505	-0,2744	-0,2829	-0,2966	-0,2027	[-0,5010; -0,0282]
3	-0,3372	-0,2780	-0,2750	-0,3021	-0,2058	[-0,5117; -0,0692]
5	-0,3427	-0,2767	-0,2750	-0,2611	-0,2031	[-0,5142; -0,0377]
10	-0,3611	-0,2901	-0,2820	-0,2600	-0,2243	[-0,5205; -0,0867]

A partir destes resultados, notamos que houve pouca influência dos valores dos hiperparâmetros σ_0 , σ_a e σ_b nas estimativas dos resumos das funções densidade *a posteriori* dos parâmetros l_0 , l_a e l_b .

Capítulo 8

Considerações Finais

O Processo de Poisson Estendido (PPE), uma generalização do Processo de Poisson Homogêneo apresentada nesta dissertação, mostra-se de grande utilidade na análise de dados discretos, pois qualquer distribuição discreta com suporte nos naturais possui uma única representação como PPE. Esta representação é obtida, como vimos, pela escolha adequada da seqüência das taxas de transição [Faddy, 1997]. Discutimos também como o comportamento das taxas de transição influi na variabilidade do processo e, desse modo, propusemos algumas parametrizações dessas taxas que originam modelos superdispersos, fato muito comum na prática, assim como parametrizações que acomodam dados com inflação ou deflação de zeros.

Os modelos propostos no Capítulo 3 mostraram-se bastante adequados à análise de dados discretos por generalizar os modelos de Poisson e Binomial Negativa, como também por acomodar de forma satisfatória diversos níveis de variabilidade dos dados. Alguns dos modelos propostos, PPE2, PPE4, PPE5 e PPE6, tiveram melhor ajuste para dados com inflação ou deflação de zeros e/ou uns, do que os modelos usuais PPE1 (Poisson) e PPE3 (Binomial Negativa). Além disso, os modelos PPE2, PPE7 e PPE8, quando comparados com modelos usuais (ZIP e ZOIP) para dados com essa característica, tiveram também um melhor ajuste segundo os critérios de seleção de modelos AIC e BIC.

No Capítulo 4 apresentamos expressões para a distribuição de probabilidades, média, variância, função geradora de probabilidades e função de verossimilhança para o modelo PPE2 sem envolver exponenciação de matrizes, que é um recurso computacionalmente custoso. Antes de proceder a análise de dados fizemos uma reparametrização com o intu-

ito de tornar o espaço paramétrico irrestrito e assim aumentar a eficiência dos algoritmos de otimização utilizados. Consideramos neste capítulo dados sobre o número de acidentes entre operárias estudados por Greenwood & Yule (1920). Para estes dados obtivemos estimativas de máxima verossimilhança e intervalos de confiança de 95% para os parâmetros. Como pudemos observar, o modelo PPE2 mostrou um melhor ajuste aos dados do que a distribuição de Poisson, notadamente para a quantidade de zeros. Ao testar a hipótese de que a distribuição de Poisson é adequada com o uso da distribuição assintótica da estatística da razão de verossimilhanças, obtivemos um p-valor que levou à rejeição da mesma para os níveis de significância usuais. Portanto, o modelo PPE2 apresentou-se como uma boa alternativa ao modelo de Poisson para dados com uma quantidade discrepante de zeros.

No Capítulo 5 apresentamos uma generalização da distribuição Binomial Negativa para dados com inflação ou deflação de zeros (PPE4). Foram discutidos os mesmos aspectos do capítulo anterior e também foi feita uma reparametrização do modelo, com o objetivo de tornar o espaço paramétrico irrestrito. Os dados sobre o número de visitas ao dentista em um ano considerados por Melkersson & Olsson (1999), citado em Saito (2005), foram considerados. A partir destes dados obtivemos estimativas de máxima verossimilhança e intervalos de confiança de 95% para os parâmetros do modelo. Conforme pudemos observar houve uma melhora significativa de ajuste aos dados quando optamos pelo modelo PPE4 ao invés da distribuição Binomial Negativa. A inadequação da distribuição Binomial Negativa também foi comprovada a partir de um teste de hipóteses baseado na estatística da razão de verossimilhanças.

No Capítulo 6 fizemos um estudo de simulação com o objetivo de verificar a adequação das estimações intervalares e dos testes de hipóteses realizados nos Capítulos 4 e 5 para os modelos PPE2 e PPE4, respectivamente. As estimativas das probabilidades de cobertura para a estimação intervalar dos parâmetros do modelo PPE2 mostraram-se bastante próximas das probabilidades de cobertura nominal (0,95), enquanto que, para o modelo PPE4, estas mesmas estimativas das probabilidades de cobertura mostraram-se com uma leve sobrecobertura (menor que 0,02) para dois dos parâmetros envolvidos (l_a e l_b). As estimativas das taxas de rejeição para o teste de hipóteses do modelo PPE2 foram aproximadamente iguais a taxa de rejeição nominal (0,05) para tamanhos amostrais

grandes, enquanto que, para o modelo PPE4, estas estimativas mostraram-se adequadas não só para tamanhos amostrais grandes com também para tamanhos amostrais moderados. Uma característica importante observada neste estudo de simulação foi o grande comprimento dos intervalos de confiança de 95% para dois dos parâmetros do modelo PPE4 (l_a e l_b), assim como um rápido decaimento da variância dos estimadores destes mesmos parâmetros em função do tamanho amostral.

No Capítulo 7 uma abordagem bayesiana para os modelos e dados apresentados nos capítulos 4 e 5 foi realizada com o uso de *prioris* normais e independentes para os parâmetros. Para cada modelo PPE2 e PPE4 foram analisados os casos de *prioris* não informativas, com variância 400, e o caso de *prioris* informativas com variâncias iguais a 1, 9, 25 e 100. A geração de amostras das distribuições *a posteriori* foram realizadas utilizando um Passeio Aleatório (Gilks, 1995), que é um caso particular do algoritmo de Metropolis. Pela análise destes resultados, concluiu-se que as estimativas *a posteriori*, para o caso não informativo, foram semelhantes àquelas obtidas via inferência clássica. Além disso, concluiu-se que as estimativas *a posteriori*, para o caso informativo, foram pouco influenciadas pela variância das *prioris* adotadas.

Os exemplos apresentados comprovaram a versatilidade dos modelos discretos baseados em Processos de Poisson Estendido, fornecendo bons ajustes a dados que apresentam superdispersão assim como àqueles que apresentam quantidades discrepantes de zeros e/ou uns. No entanto esses modelos, em geral, não apresentam formas fechadas para a distribuição de probabilidades e para a função de verossimilhança, envolvendo a exponenciação de matrizes ou integração numérica, onerando o procedimento de estimação sob o ponto de vista computacional. Este aspecto será evidenciado quando o maior valor amostral for grande ou houver muitas observações na amostra.

Referências Bibliográficas

- [1] Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. In: INTERNATIONAL SYMPOSIUM IN INFORMATION THEORY, 2., 1973, **Proceedings...** p. 267-281.
- [2] Ball, F. A Note on Variation in Birth Processes. **Mathematical Scientist**, v. 20, p. 50-55, 1995.
- [3] Çinlar, E. **Introduction to stochastic processes**. 1. ed. New Jersey: Englewood Cliffs, 1975.
- [4] Faddy, M. J. On Variation in Poisson Processes. **Mathematical Scientist**, v. 19, p. 47-51, 1994.
- [5] Faddy, M. J. Extended Poisson Process Modelling and Analysis of Count Data. **Biometrical Journal**, v. 39, p. 431-440, 1997.
- [6] Feller, W. **An Introduction to Probability Theory and Applications**. 3 ed. New York: Wiley, 1971. v. 1.
- [7] Gelman, A. Inference and monitoring convergence. In: Gilks, W. R.; Richardson, S.; Spiegelhalter, D. J.(Ed.) **Markov Chain Monte Carlo in Practice**. London: Chapman & Hall, 1995. p. 131-143.
- [8] Gilks, W. R., Richardson, S. e Spiegelhalter, D. J. Introducing Markov chain Monte Carlo. In: Gilks, W. R.; Richardson, S.; Spiegelhalter, D. J.(Ed.) **Markov Chain Monte Carlo in Practice**. London: Chapman & Hall, 1995. p. 1-19.

- [9] Melkersson, M.; Olsson, C. Is Visiting the Dentist a good Habit? Analyzing Count Data with Excess Zeros and Excess Ones. Disponível em: <<http://www.econ.umu.se>>, 1999.
- [10] Podlich, H.M. **Extended Poisson process models for count data**. 1999. 204 f. Tese. Departamento de Matemática, Universidade de Queensland, Queensland, 1999.
- [11] Ross, S. M. **Simulation**. 2 ed. San Diego: Academic Press, 1990.
- [12] Saito, M.Y. **Inferência Bayesiana para Dados Discretos com Excesso de Zeros e Uns**. 2005. 82 f. Dissertação (Mestrado em Estatística) - Departamento de Estatística, Universidade Federal de São Carlos, São Carlos, 2005.
- [13] Schwarz, G. Estimating the Dimension of a Model. **Annals of Statistics**, v. 6, p. 461-464, 1978.

Apêndice A

Rotinas do Capítulo 3

```
#MODELO PPE1

taxas_ppe1 <- function(n, lambda){

lambda}# end-function

# MODELO PPE2

taxas_ppe2 <- function(n, lambda0, lambda1){

if (n==0) return(lambda0) else return(lambda1);

}# end-function

# MODELO PPE3

taxas_ppe3 <- function(n, a, b){a*(b+n)}# end-function

# MODELO PPE4

taxas_ppe4 <- function(n, lambda0, a, b){

if (n==0) return(lambda0) else return(a*(b+n));

}# end-function

# MODELO PPE5

taxas_ppe5 <- function(n, a, b, c){a*(b+n)^c}# end-function

# MODELO PPE6

taxas_ppe6 <- function(n, lambda0, a, b, c){

if (n==0) return(lambda0) else return(a*(b+n)^c);

}# end-function

# MODELO PPE7

taxas_ppe7 <- function(n, lambda0, lambda1, lambda){
```

```
if (n==0) return(lambda0) else {
  if (n==1) return(lambda1) else return(lambda)}
}# end-function

# MODELO PPE8

taxas_ppe8 <- function(n, lambda0, lambda1, a, b){
  if (n==0) return(lambda0) else {
    if (n==1) return(lambda1) else return(a*(b+n))}
  }# end-function

# FUNÇÃO QUE RETORNA MATRIZ Q
require(Matrix)

Q_matriz <- function(taxas) {
  n <-length(taxas)
  Q <- matrix(0, nrow = n, ncol = n);
  for (i in 1:n) {
    for (j in 1:n) {
      if (j==i) Q[i,j] <- -taxas[i] else
      if (j==(i+1)) Q[i,j] <- taxas[i];
    }# end-for
  }# end-for
  Q <- Matrix(Q)
  return(Q)
}# end-function

# DADOS: NRO DE CONTAGENS NA LOCALIDADE A
dados <- scan("c:/dados_localidadeA.dat")
N <- length(dados) - 1 # MAIOR VALOR OBSERVADO
n <- sum(dados)
X <- rep(0, dados[1])
for (i in 2:length(dados)) {
  X <- c(X, rep(i-1, dados[i]))}
summary(X)
hist(X, nclass = 50)
```

```

# MODELO PPE1

# FUNÇÃO QUE RETORNA VETOR DE PROBABILIDADES (PPE1)

fp_ppe1 <- function(N, lambda) {

taxas <- numeric(N+1) # VETOR DAS TAXAS DE TRANSIÇÃO

for (i in 1:(N+1)) {taxas[i] <- taxas_ppe1(i-1, lambda)}

Q <- Q_matriz(taxas)

e1 <- numeric(N+1); e1[1] <- 1;

p <- as.vector(e1%*%expm(Q)) # VETOR DE PROBABILIDADES

return(p)

}# end-function

# FUNÇÃO LOG-VEROSSIMILHANÇA (PPE1)

require(stats4)

l_ppe1 <- function(freq, lambda) {

N <- length(freq) - 1 # MAIOR VALOR OBSERVADO NA AMOSTRA

p <- as.numeric(fp_ppe1(N, lambda)) # VETOR DE PROBABILIDADES

y <- 0

for (i in 1:(N+1)) {y <- y + freq[i]*log(p[i])}

return(y)

}# end-function

f <- function(lambda) {-l_ppe1(dados, lambda)} # (-1)*LOG-VEROSSIMILHANÇA

# ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA

emv_ppe1 <- mle(f, start = list(lambda = 1), method = "L-BFGS-B", lower=c(0))

summary(emv_ppe1)

lambda_max <- as.numeric(coef(emv_ppe1));

# FREQUENCIAS ESPERADAS PARA O MODELO AJUSTADO (PPE1)

fe_ppe1 <- n*fp(N, lambda_max)

# MEDIDA QUI-QUADRADO

q2_ppe1 <- sum((fe_ppe1 - dados)^2/fe_ppe1)

# AIC e BIC

AIC_ppe1 <- 2*f(lambda_max) + 2*1

BIC_ppe1 <- 2*f(lambda_max) + 1*log(n)

```

```

# MODELO PPE2

# FUNÇÃO QUE RETORNA VETOR DE PROBABILIDADES (PPE2)

fp_ppe2 <- function(N, lambda0, lambda1) {

taxas <- numeric(N+1) # VETOR DAS TAXAS DE TRANSIÇÃO

for (i in 1:(N+1)) {taxas[i] <- taxas_ppe2(i-1, lambda0, lambda1)}

Q <- Q_matriz(taxas)

e1 <- numeric(N+1); e1[1] <- 1;

p <- as.vector(e1%%expm(Q)) # VETOR DE PROBABILIDADES

return(p)

}# end-function

# FUNÇÃO LOG-VEROSSIMILHANÇA (PPE2)

l_ppe2 <- function(freq, lambda0, lambda1) {

N <- length(freq) - 1 # MAIOR VALOR OBSERVADO NA AMOSTRA

p <- fp_ppe2(N, lambda0, lambda1) # VETOR DE PROBABILIDADES

y <- 0

for (i in 1:(N+1)) {y <- y + freq[i]*log(p[i])}

return(y)

}# end-function

f <- function(lambda0, lambda1) {-l_ppe2(dados, lambda0, lambda1)} # (-1)*LOG-VEROSSIMILHANÇA

# ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA (PPE2)

emv_ppe2 <- mle(f, start = list(lambda0 = 3, lambda1 = 5), method = "L-BFGS-B", lower=c(0, 0))

summary(emv_ppe2)

lambda0_max <- as.numeric(coef(emv_ppe2)[1]);

lambda1_max <- as.numeric(coef(emv_ppe2)[2]);

# FREQUENCIAS ESPERADAS PARA O MODELO AJUSTADO (PPE2)

fe_ppe2 <- n*fp_ppe2(N, lambda0_max, lambda1_max)

# MEDIDA QUI-QUADRADO

q2_ppe2 <- sum((fe_ppe2 - dados)^2/fe_ppe2)

# AIC e BIC

AIC_ppe2 <- 2*f(lambda0_max, lambda1_max) + 2*2

BIC_ppe2 <- 2*f(lambda0_max, lambda1_max) + 2*log(n)

```

```

# MODELO PPE3

# FUNÇÃO QUE RETORNA VETOR DE PROBABILIDADES (PPE3)

fp_ppe3 <- function(N, a, b) {

taxas <- numeric(N+1) # VETOR DAS TAXAS DE TRANSIÇÃO

for (i in 1:(N+1)) {taxas[i] <- taxas_ppe3(i-1, a, b)}

Q <- Q_matriz(taxas)

e1 <- numeric(N+1); e1[1] <- 1;

p <- as.vector(e1%*%expm(Q)) # VETOR DE PROBABILIDADES

return(p)

}# end-function

# FUNÇÃO LOG-VEROSSIMILHANÇA (PPE3)

require(stats4)

l_ppe3 <- function(freq, a, b) {

N <- length(freq) - 1 # MAIOR VALOR OBSERVADO NA AMOSTRA

p <- fp_ppe3(N, a, b) # VETOR DE PROBABILIDADES

y <- 0

for (i in 1:(N+1)) {y <- y + freq[i]*log(p[i])}

return(y)

}# end-function

f <- function(a, b) {-l_ppe3(dados, a, b)} # (-1)*LOG-VEROSSIMILHANÇA

# ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA

emv_ppe3 <- mle(f, start = list(a = 1, b = 1), method = "L-BFGS-B", lower=c(0,0))

summary(emv_ppe3)

a_max <- as.numeric(coef(emv_ppe3)[1]);

b_max <- as.numeric(coef(emv_ppe3)[2]);

# FREQUENCIAS ESPERADAS PARA O MODELO AJUSTADO (PPE1)

fe_ppe3 <- n*fp_ppe3(N, a_max, b_max)

# MEDIDA QUI-QUADRADO

q2_ppe3 <- sum((fe_ppe3 - dados)^2/fe_ppe3)

# AIC e BIC

AIC_ppe3 <- 2*f(a_max, b_max) + 2*2

```

```

BIC_ppe3 <- 2*f(a_max, b_max) + 2*log(n)

# MODELO PPE4

# FUNÇÃO QUE RETORNA VETOR DE PROBABILIDADES (PPE4)

fp_ppe4 <- function(N, lambda0, a, b) {

taxas <- numeric(N+1) # VETOR DAS TAXAS DE TRANSIÇÃO

for (i in 1:(N+1)) {taxas[i] <- taxas_ppe4(i-1, lambda0, a, b)}

Q <- Q_matriz(taxas)

e1 <- numeric(N+1); e1[1] <- 1;

p <- as.vector(e1%%expm(Q)) # VETOR DE PROBABILIDADES

return(p)

}# end-function

# FUNÇÃO LOG-VEROSSIMILHANÇA (PPE4)

require(stats4)

l_ppe4 <- function(freq, lambda0, a, b) {

N <- length(freq) - 1 # MAIOR VALOR OBSERVADO NA AMOSTRA

p <- fp_ppe4(N, lambda0, a, b) # VETOR DE PROBABILIDADES

y <- 0

for (i in 1:(N+1)) {y <- y + freq[i]*log(p[i])}

return(y)

}# end-function

f <- function(lambda0, a, b) {-l_ppe4(dados, lambda0, a, b)} # (-1)*LOG-VEROSSIMILHANÇA

# ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA (PPE4)

emv_ppe4 <- mle(f, start = list(lambda0 = 0.5, a = 1, b = 0), method = "L-BFGS-B", lower=c(0,0, -1))

summary(emv_ppe4)

lambda0_max <- as.numeric(coef(emv_ppe4)[1]);

a_max <- as.numeric(coef(emv_ppe4)[2]);

b_max <- as.numeric(coef(emv_ppe4)[3]);

# FREQUENCIAS ESPERADAS PARA O MODELO AJUSTADO (PPE4)

fe_ppe4 <- n*fp_ppe4(N, lambda0_max, a_max, b_max)

# MEDIDA QUI-QUADRADO

q2_ppe4 <- sum((fe_ppe4 - dados)^2/fe_ppe4)

```

```

# AIC e BIC

AIC_ppe4 <- 2*f(lambda0_max, a_max, b_max) + 2*3
BIC_ppe4 <- 2*f(lambda0_max, a_max, b_max) + 3*log(n)

# MODELO PPE5

# FUNÇÃO QUE RETORNA VETOR DE PROBABILIDADES (PPE5)

fp_ppe5 <- function(N, a, b, C) {

taxas <- numeric(N+1) # VETOR DAS TAXAS DE TRANSIÇÃO

for (i in 1:(N+1)) {taxas[i] <- taxas_ppe5(i-1, a, b, C)}

Q <- Q_matriz(taxas)

e1 <- numeric(N+1); e1[1] <- 1;

p <- as.vector(e1%%expm(Q)) # VETOR DE PROBABILIDADES

return(p)

}# end-function

# FUNÇÃO LOG-VEROSSIMILHANÇA (PPE5)

require(stats4)

l_ppe5 <- function(freq, a, b, c) {

N <- length(freq) - 1 # MAIOR VALOR OBSERVADO NA AMOSTRA

p <- fp_ppe5(N, a, b, c) # VETOR DE PROBABILIDADES

y <- 0

for (i in 1:(N+1)) {y <- y + freq[i]*log(p[i])}

return(y)

}# end-function

f <- function(a, b, c) {-l_ppe5(dados, a, b, c)} # (-1)*LOG-VEROSSIMILHANÇA

# ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA

emv_ppe5 <- mle(f, start = list(a = 10, b = 0.5, c = 0.2), method = "L-BFGS-B", lower=c(0,0, -Inf), upper=c(Inf, Inf,

1))

summary(emv_ppe5)

a_max <- as.numeric(coef(emv_ppe5)[1]);

b_max <- as.numeric(coef(emv_ppe5)[2]);

c_max <- as.numeric(coef(emv_ppe5)[3]);

# FREQUENCIAS ESPERADAS PARA O MODELO AJUSTADO (PPE5)

```

```

fe_ppe5 <- n*fp_ppe5(N, a_max, b_max, c_max)

# MEDIDA QUI-QUADRADO

q2_ppe5 <- sum((fe_ppe5 - dados)^2/fe_ppe5)

# AIC e BIC

AIC_ppe5 <- 2*f(a_max, b_max, c_max) + 2*3

BIC_ppe5 <- 2*f(a_max, b_max, c_max) + 3*log(n)

# MODELO PPE6

# FUNÇÃO QUE RETORNA VETOR DE PROBABILIDADES (PPE6)

fp_ppe6 <- function(N, lambda0, a, b, C) {

taxas <- numeric(N+1) # VETOR DAS TAXAS DE TRANSIÇÃO

for (i in 1:(N+1)) {taxas[i] <- taxas_ppe6(i-1, lambda0, a, b, C)}

Q <- Q_matriz(taxas)

e1 <- numeric(N+1); e1[1] <- 1;

p <- as.vector(e1%%expm(Q)) # VETOR DE PROBABILIDADES

return(p)

}# end-function

# FUNÇÃO LOG-VEROSSIMILHANÇA (PPE6)

require(stats4)

l_ppe6 <- function(freq, lambda0, a, b, c) {

N <- length(freq) - 1 # MAIOR VALOR OBSERVADO NA AMOSTRA

p <- fp_ppe6(N, lambda0, a, b, c) # VETOR DE PROBABILIDADES

y <- 0

for (i in 1:(N+1)) {y <- y + freq[i]*log(p[i])}

return(y)

}# end-function

f <- function(lambda0, a, b, c) {-l_ppe6(dados, lambda0, a, b, c)} # (-1)*LOG-VEROSSIMILHANÇA

# ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA (PPE6)

emv_ppe6 <- mle(f, start = list(lambda0 = 1, a = 10, b = 0.5, c = 0.2), method = "L-BFGS-B", lower=c(0, 0, -1, -Inf),

upper=c(Inf, Inf, Inf, 1))

summary(emv_ppe6)

lambda0_max <- as.numeric(coef(emv_ppe6)[1])

```



```

a_max <- as.numeric(coef(emv_ppe6)[2]);
b_max <- as.numeric(coef(emv_ppe6)[3]);
c_max <- as.numeric(coef(emv_ppe6)[4]);

# FREQUENCIAS ESPERADAS PARA O MODELO AJUSTADO (PPE6)
fe_ppe6 <- n*fp_ppe6(N, lambda0_max, a_max, b_max, c_max)

# MEDIDA QUI-QUADRADO
q2_ppe6 <- sum((fe_ppe6 - dados)^2/fe_ppe6)

# AIC e BIC
AIC_ppe6 <- 2*f(lambda0_max, a_max, b_max, c_max) + 2*4
BIC_ppe6 <- 2*f(lambda0_max, a_max, b_max, c_max) + 4*log(n)

# DADOS: NRO DE VISITAS AO DENTISTA
dados <- scan("c:/dados_dentista.dat")

N <- length(dados) - 1 # MAIOR VALOR OBSERVADO
n <- sum(dados)

X <- rep(0, dados[1])

for (i in 2:length(dados)) {
  X <- c(X, rep(i-1, dados[i]))}

summary(X)

hist(X, nclass = 50)

# MODELO PPE7

# FUNÇÃO QUE RETORNA VETOR DE PROBABILIDADES (PPE7)
fp_ppe7 <- function(N, lambda0, lambda1, lambda) {

  taxas <- numeric(N+1) # VETOR DAS TAXAS DE TRANSIÇÃO

  for (i in 1:(N+1)) {taxas[i] <- taxas_ppe7(i-1, lambda0, lambda1, lambda)}

  Q <- Q_matriz(taxas)

  e1 <- numeric(N+1); e1[1] <- 1;

  p <- as.vector(e1%*%expm(Q)) # VETOR DE PROBABILIDADES

  return(p)

}# end-function

# FUNÇÃO LOG-VEROSSIMILHANÇA (PPE7)

require(stats4)

```

```

l_ppe7 <- function(freq, lambda0, lambda1, lambda) {
  N <- length(freq) - 1 # MAIOR VALOR OBSERVADO NA AMOSTRA
  p <- fp_ppe7(N, lambda0, lambda1, lambda) # VETOR DE PROBABILIDADES
  y <- 0
  for (i in 1:(N+1)) {y <- y + freq[i]*log(p[i])}
  return(y)
}# end-function

f <- function(lambda0, lambda1, lambda) {-l_ppe7(dados, lambda0, lambda1, lambda)} #
(-1)*LOG-VEROSSIMILHANÇA
# ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA (PPE7)
emv_ppe7 <- mle(f, start = list(lambda0 = 0.5, lambda1 = 1, lambda = 1), method = "L-BFGS-B", lower=c(0,0, 0))
summary(emv_ppe7)
lambda0_max <- as.numeric(coef(emv_ppe7)[1]);
lambda1_max <- as.numeric(coef(emv_ppe7)[2]);
lambda_max <- as.numeric(coef(emv_ppe7)[3]);
# FREQUENCIAS ESPERADAS PARA O MODELO AJUSTADO (PPE7)
fe_ppe7 <- n*fp_ppe7(N, lambda0_max, lambda1_max, lambda_max)
# MEDIDA QUI-QUADRADO
q2_ppe7 <- sum((fe_ppe7 - dados)^2/fe_ppe7)
# AIC e BIC
AIC_ppe7 <- 2*f(lambda0_max, lambda1_max, lambda_max) + 2*3
BIC_ppe7 <- 2*f(lambda0_max, lambda1_max, lambda_max) + 3*log(n)
# MODELO PPE8
# FUNÇÃO QUE RETORNA VETOR DE PROBABILIDADES (PPE8)
fp_ppe8 <- function(N, lambda0, lambda1, a, b) {
  taxas <- numeric(N+1) # VETOR DAS TAXAS DE TRANSIÇÃO
  for (i in 1:(N+1)) {taxas[i] <- taxas_ppe8(i-1, lambda0, lambda1, a, b)}
  Q <- Q_matriz(taxas)
  e1 <- numeric(N+1); e1[1] <- 1;
  p <- as.vector(e1%*%expm(Q)) # VETOR DE PROBABILIDADES
  return(p)
}

```

```

}# end-function

# FUNÇÃO LOG-VEROSSIMILHANÇA (PPE8)

require(stats4)

l_ppe8 <- function(freq, lambda0, lambda1, a, b) {

N <- length(freq) - 1 # MAIOR VALOR OBSERVADO NA AMOSTRA

p <- fp_ppe8(N, lambda0, lambda1, a, b) # VETOR DE PROBABILIDADES

y <- 0

for (i in 1:(N+1)) {y <- y + freq[i]*log(p[i])}

return(y)

}# end-function

f <- function(lambda0, lambda1, a, b) {-l_ppe8(dados, lambda0, lambda1, a, b)} # (-1)*LOG-VEROSSIMILHANÇA

# ESTIMAÇÃO DE MÁXIMA VEROSSIMILHANÇA (PPE8)

emv_ppe8 <- mle(f, start = list(lambda0 = 0.5, lambda1 = 1, a = 1, b = 0), method = "L-BFGS-B", lower=c(0,0, 0, -2))

summary(emv_ppe8)

lambda0_max <- as.numeric(coef(emv_ppe8)[1]);

lambda1_max <- as.numeric(coef(emv_ppe8)[2]);

a_max <- as.numeric(coef(emv_ppe8)[3]);

b_max <- as.numeric(coef(emv_ppe8)[4]);

# FREQUENCIAS ESPERADAS PARA O MODELO AJUSTADO (PPE7)

fe_ppe8 <- n*fp_ppe8(N, lambda0_max, lambda1_max, a_max, b_max)

# MEDIDA QUI-QUADRADO

q2_ppe8 <- sum((fe_ppe8 - dados)^2/fe_ppe8)

# AIC e BIC

AIC_ppe8 <- 2*f(lambda0_max, lambda1_max, a_max, b_max) + 2*4

BIC_ppe8 <- 2*f(lambda0_max, lambda1_max, a_max, b_max) + 4*log(n)

```

Apêndice B

Rotinas do Capítulo 4

```
# MODELO PPE2

dados <- scan("C:/dados_operarias.dat"); # VETOR DAS FREQUÊNCIAS OBSERVADAS

N <- length(dados) - 1; # MAIOR VALOR OBSERVADO

# ANALISE DESCRITIVA

X <- rep(0, dados[1]) # VETOR DE OBSERVAÇÕES

for (i in 1:(length(dados) - 1)) {

v <- rep(i, dados[i+1])

X <- c(X, v)}

n <- length(X); # TAMANHO AMOSTRAL

n0 <- sum(X==0); # QTDE DE ZEROS NA AMOSTRA

summary(X)

hist(X, nclass = 50, main=, xlab="Número de acidentes", ylab="Frequência");

# INTEGRANDO

integrando <- function(x, m, b) {

x^(m)*exp(-b*x)}

# FUNCAO DE PROBABILIDADES

p <- function(n, l0, l1) {

if (n == 0) y <- exp(-exp(l0))

else{

y <- exp(l0)*(exp(l1))^(n-1)*exp(-exp(l0))/factorial(n-1)*integrate(integrando, 0, 1, m = n-1, b = exp(l1)-exp(l0))$value};
```

```

return(y)}

# FUNÇÃO LOG-VEROSSIMILHANÇA
ll <- function(l0, l1) {
soma1 <- 0;
for (i in 1:n) {
if (X[i] > 0) soma1 <- soma1 + lfactorial(X[i] - 1)} # end-for

soma2 <- 0;
for (i in 1:n) {
if (X[i] > 0) soma2 <- soma2 + log(integrate(integrando, 0, 1, m = X[i] - 1, b = exp(l1) - exp(l0))$value)} # end-for

valor <- -n*exp(l0) + (n-n0)*l0 + (sum(X)-n+n0)*l1 - soma1 + soma2

return(valor)
} #end-function

ll1 <- function(l0, l1) {(-1)*ll(l0, l1)} # (-1)*log_verossimilhança

# EMV
require(stats4)

emv <- mle(ll1, start = list(l0 = -0.9, l1 = 0.01), method="Nelder-Mead")

l0_max <- coef(emv)[1];
l1_max <- coef(emv)[2];

summary(emv)

# AIC e BIC
AIC <- -2*ll(l0_max, l1_max) + 2*2
BIC <- -2*ll(l0_max, l1_max) + 2*log(n)

# INTERVALOS DE CONFIANÇA ASSINTÓTICOS USANDO A MATRIZ OBSERVADA DE FISHER
# l0
coef(emv)[1]- qnorm(0.975)*sqrt(vcov(emv)[1,1]) # limite inferior
coef(emv)[1]+ qnorm(0.975)*sqrt(vcov(emv)[1,1]) # limite superior
# l1
coef(emv)[2]- qnorm(0.975)*sqrt(vcov(emv)[2,2]) # limite inferior

```

```

coef(emv)[2]+ qnorm(0.975)*sqrt(vcov(emv)[2,2]) # limite superior

# FREQUÊNCIAS ESPERADAS PPE2

fe <- numeric(length(dados))

for (i in 1:length(dados)) {

fe[i] <- n*p(i - 1, l0_max, l1_max)}

# Cálculo dos EMV SOB H0:

# FUNÇÃO DE (-1)*LOG-VEROSSIMILHANÇA PARA O MODELO RESTRITO (SOB H0)

ll_H0 <- function(l) {(-1)*ll(l, l)}

emv_H0 <- mle(ll_H0, start = list(l = log(mean(X))), method="BFGS")

l_maxH0 <- coef(emv_H0)[1]

summary(emv_H0)

# AIC e BIC

AIC_H0 <- -2*ll(l_maxH0, l_maxH0) + 2*1

BIC_H0 <- -2*ll(l_maxH0, l_maxH0) + 2*log(n)

# TESTE DE HIPÓTESE - H0:  $\theta_0 = \theta_1$  (POISSON)

# ESTATÍSTICA DA RAZÃO DE VEROSSIMILHANÇA

ERV <- -2*ll(l_maxH0, l_maxH0) + 2*ll(l0_max, l1_max)

# P-VALOR

valor_p <- 1 - pchisq(ERV, 1)

# FREQUÊNCIAS ESPERADAS SOB H0

fe0 <- numeric(length(dados))

for (i in 1:length(dados)) {

fe0[i] <- n*p(i - 1, l_maxH0, l_maxH0)}

# FUNÇÃO LOG-VEROSSIMILHANÇA PADRONIZADA

maior <- ll(l0_max, l1_max);

l_max <- function(l0, l1) { ll(l0, l1)/maior };

# Contornos

```

```
x <- seq(-1.3, -0.7, 0.01);
y <- seq(-0.3, 0.2, 0.01);
z <- matrix(0, nrow = length(x), ncol = length(y));
for (i in 1:length(x)) {
  for (j in 1:length(y)) {
    z[i,j] <- 1 - max(x[i], y[j])
  }
}
contour(x,y,z, nlevels = 40,labcex = 0.5, xlab = 'l0', ylab = 'l1')
title("Contornos da Log_verossimilhança padronizada")
```

Apêndice C

Rotinas do Capítulo 5

```
# FUNÇÃO QUE RETORNA A COMBINAÇÃO DE x e k, x REAL e k INTEIRO NÃO-NEGATIVO
```

```
combina <- function(x, k) {
```

```
  if ( k == 0) y <- 1
```

```
  else{
```

```
    y <- prod((x - k + 1):x)/factorial(k);
```

```
  return(y)}
```

```
# FUNÇÃO DE PROBABILIDADES
```

```
integrando <- function(x, m, v, c) {
```

```
  exp(m*x)*(1 - exp(v*x))^c}
```

```
p <- function(n, l0, la, lb) {
```

```
  if (n == 0) y <- exp(-exp(l0))
```

```
  else{
```

```
    y <- exp(l0-exp(l0))*combina(exp(lb)+n-2, n-1)*integrate(integrando, 0, 1, m = exp(l0)-exp(la+lb), v = -exp(la), c = n - 1)$value};
```

```
  return(y)}
```

```
# DADOS: NRO DE VISITAS AO DENTISTA
```

```
dados <- scan("c:/dados_dentista.dat");
```

```
N <- length(dados) - 1 # Maior valor observado na amostra
```

```
# ANÁLISE DESCRITIVA
```



```

X <- rep(0, dados[1])
for (i in 1:(length(dados) - 1)) {
v <- rep(i, dados[i+1])
X <- c(X, v)}
n <- length(X) # Quantidade de observações
n0 <- sum(X == 0)
summary(X)
hist(X, nclass = 50, xlab = "Número de visitas", ylab = "Frequência", main=)

# FUNÇÃO LOG-VEROSSIMILHANÇA MODELO PPE4 (IRRESTRITO)
ll <- function(l0, la, lb){
soma <- 0;
for (i in 1:n) {
if (X[i] > 0) soma <- soma + log(combina(exp(lb)+X[i]-2,X[i]-1)) +
log(integrate(integrando, 0, 1, m = exp(l0)-exp(la+lb), v = -exp(la), c = X[i]-1)$value)
} # end-for
y <- soma + (n - n0)*l0 - n*exp(l0)
return(y);
}
ll1 <- function(l0, la, lb) {(-1)*ll(l0, la, lb)} # (-1)*log_verossimilhança

# FUNÇÃO (-1)*LOG-VEROSSIMILHANÇA MODELO B. NEGATIVA (RESTRITO)
ll1_bn <- function(la, lb) {ll1(la + log(exp(lb)-1), la, lb)}

# AJUSTE AO MODELO B. NEGATIVA (RESTRITO)

require(stats4) # Carrega pacote stats4
emv_bn <- mle(ll1_bn, start = list(la = 0.1, lb = 0.5), method="L-BFGS-B", lower=c(-Inf, 0))
summary(emv_bn)

la.bn_max <- as.numeric(coef(emv_bn)[1])
lb.bn_max <- as.numeric(coef(emv_bn)[2])

# INTERVALOS DE CONFIANÇA ASSINTÓTICOS USANDO A MATRIZ OBSERVADA DE FISHER

```

```

la.bn_inf <- as.numeric(coef(emv_bn)[1] - qnorm(0.975)*sqrt(vcov(emv_bn)[1,1])) # limite inferior
la.bn_sup <- as.numeric(coef(emv_bn)[1] + qnorm(0.975)*sqrt(vcov(emv_bn)[1,1])) # limite superior
lb.bn_inf <- as.numeric(coef(emv_bn)[2] - qnorm(0.975)*sqrt(vcov(emv_bn)[2,2])) # limite inferior
lb.bn_sup <- as.numeric(coef(emv_bn)[2] + qnorm(0.975)*sqrt(vcov(emv_bn)[2,2])) # limite superior

# FREQUENCIAS ESPERADAS PARA O MODELO AJUSTADO

fe_bn <- numeric(max(X)+1)

for (i in 1:(max(X)+1)) { fe_bn[i] <- n*p(i-1,la.bn_max + log(exp(lb.bn_max)-1), la.bn_max, lb.bn_max)}

# AIC e BIC

AIC_bn <- 2*lll_bn(la.bn_max, lb.bn_max) + 2*2
BIC_bn <- 2*lll_bn(la.bn_max, lb.bn_max) + 2*log(n)

# AJUSTE AO MODELO PPE4 (IRRESTRITO)

require(stats4)

emv <- mle(lll, start = list(l0 = 0.1, la = 0.6, lb = -0.5), method="Nelder-Mead")

summary(emv)

l0_max <- as.numeric(coef(emv)[1]);
la_max <- as.numeric(coef(emv)[2]);
lb_max <- as.numeric(coef(emv)[3]);

# INTERVALOS DE CONFIANÇA ASSINTÓTICOS USANDO A MATRIZ OBSERVADA DE FISHER

l0_inf <- as.numeric(coef(emv)[1] - qnorm(0.975)*sqrt(vcov(emv)[1,1])) # limite inferior
l0_sup <- as.numeric(coef(emv)[1] + qnorm(0.975)*sqrt(vcov(emv)[1,1])) # limite superior
la_inf <- as.numeric(coef(emv)[2] - qnorm(0.975)*sqrt(vcov(emv)[2,2])) # limite inferior
la_sup <- as.numeric(coef(emv)[2] + qnorm(0.975)*sqrt(vcov(emv)[2,2])) # limite superior
lb_inf <- as.numeric(coef(emv)[3] - qnorm(0.975)*sqrt(vcov(emv)[3,3])) # limite inferior
lb_sup <- as.numeric(coef(emv)[3] + qnorm(0.975)*sqrt(vcov(emv)[3,3])) # limite superior

# FREQUENCIAS ESPERADAS PARA O MODELO AJUSTADO

fe <- numeric(length(dados))

for (i in 1:length(dados)) {

fe[i] <- p(i - 1, l0_max, la_max, lb_max)*n}

```

```
# AIC e BIC

AIC <- 2*ll(10_max, la_max, lb_max) + 2*3
BIC <- 2*ll(10_max, la_max, lb_max) + 3*log(n)

# TESTE DE HIPÓTESE H0: lambda0 = ab (Binomial Negativa)

# ESTATÍSTICA DA RAZÃO DE VEROSSIMILHANÇA

ERV <- 2*ll(la.bn_max + log(exp(lb.bn_max)+1), la.bn_max, lb.bn_max) + 2*ll(10_max, la_max, lb_max)

# P_VALOR

valor_p <- 1 - pchisq(ERV, 1)

# CONTORNOS PARA A FUNÇÃO LOG-VEROSSIMILHANÇA PADRONIZADA

maior <- ll(10_max, la_max, lb_max)

# FUNÇÃO LOG-VEROSSIMILHANÇA PADRONIZADA

l_max <- function(10, la, lb) {ll(10, la, lb)/maior};

# Fixando 10_max

x <- seq(0.2, 0.6, 0.01);
y <- seq(-0.5, 0, 0.01);

z <- matrix(0, nrow = length(x), ncol = length(y));

for (i in 1:length(x)) {
  for (j in 1:length(y)) {
    z[i,j] <- l_max(10_max, x[i], y[j])}
}

contour(x,y,z, nlevels = 40,labcex = 0.5, xlab = 'la', ylab = 'lb')

title("Contornos para Log_verossimilhança padronizada")

# Fixando la_max

x <- seq(0.4, 0.7, 0.01);
y <- seq(-0.5, 0, 0.01);

z <- matrix(0, nrow = length(x), ncol = length(y));

for (i in 1:length(x)) {
  for (j in 1:length(y)) {
```

```
z[i,j] <- 1_max(x[i], la_max, y[j]) }}  
  
contour(x,y,z, nlevels = 40,labcex = 0.5, xlab = 'l0', ylab = 'lb')  
  
title("Contornos para Log_verossimilhança padronizada")  
  
# Fixando lb_max  
  
x <- seq(0.4, 0.7, 0.01);  
  
y <- seq(0.2, 0.6, 0.01);  
  
z <- matrix(0, nrow = length(x), ncol = length(y));  
  
for (i in 1:length(x)) {  
  for (j in 1:length(y)) {  
    z[i,j] <- 1_max(x[i], y[j], lb_max)}}  
  
contour(x,y,z, nlevels = 40,labcex = 0.5, xlab = 'l0', ylab = 'la')  
  
title("Contornos para Log_verossimilhança padronizada")
```

Apêndice D

Rotinas do Capítulo 6

D.1 Modelo PPE2 - Probabilidade de Cobertura

```
# TAXAS DE TRANSIÇÃO

taxas_ppe2 <- function(n, l0, l1){
  if (n==0) return(exp(l0)) else return(exp(l1));
}# end-function

# GERADOR DE AMOSTRAS

rnd_ppe2 <- function(n, l0, l1) {
  vetor <- numeric(n)

  for (j in 1:n) {
    i <- 0; p <- 1;
    while (p <= exp(l1)) {
      u <- runif(1)
      p <- p*(1/u)^(1/taxas_ppe2(i, l0, l1))
      i <- i + 1
    } # end-while
    vetor[j] <- i - 1;
  } # end-for

  return(vetor)
} # end-function
```

```

# INTEGRANDO

integrando <- function(x, m, b) {
  x^(m)*exp(-b*x)}

# FUNÇÃO LOG-VEROSSIMILHANÇA

ll <- function(l0, l1, X) {
  soma1 <- 0; n <- length(X); n0 <- sum(X==0)

  for (i in 1:n) {

  if (X[i] > 0) soma1 <- soma1 + lfactorial(X[i] - 1)} # end-for

  soma2 <- 0;

  for (i in 1:n) {

  if (X[i] > 0) soma2 <- soma2 + log(integrate(integrando, 0, 1, m = X[i] - 1, b = exp(l1) - exp(l0))$value)} # end-for

  valor <- -n*exp(l0) + (n-n0)*l0 + (sum(X)-n+n0)*l1 - soma1 + soma2

  return(valor)

} #end-function

ll1 <- function(l0, l1, X) {(-1)*ll(l0, l1, X)} # (-1)*log_verossimilhança

# FUNÇÃO QUE CALCULA IC

require(stats4)

ic_ppe2 <- function(amostra, conf = 0.95) {

  emv <- mle(ll1, start = list(l0 = 0.5, l1 = 0.7), fixed = list(X = amostra), method="Nelder-Mead")

  # INTERVALOS DE CONFIANÇA ASSINTÓTICOS USANDO A MATRIZ OBSERVADA DE FISHER

  l0_inf <- as.numeric(coef(emv)[1] - qnorm((1 + conf)/2)*sqrt(vcov(emv)[1,1])) # limite inferior
  l0_sup <- as.numeric(coef(emv)[1] + qnorm((1 + conf)/2)*sqrt(vcov(emv)[1,1])) # limite superior
  l1_inf <- as.numeric(coef(emv)[2] - qnorm((1 + conf)/2)*sqrt(vcov(emv)[2,2])) # limite inferior
  l1_sup <- as.numeric(coef(emv)[2] + qnorm((1 + conf)/2)*sqrt(vcov(emv)[2,2])) # limite superior

  return(list(l0 = c(l0_inf, l0_sup), l1 = c(l1_inf, l1_sup)))

} # end-function

# SIMULAÇÃO - PROBABILIDADES DE COBERTURA

# VALORES VERDADEIROS DOS PARAMETROS

l0 <- -0.37

l1 <- 1.5

```

```
# QTDE DE SIMULAÇÕES
S <- 1000;

# VETOR DOS TAMANHOS AMOSTRAIS
n <- c(400, 150, 80, 30)

# MATRIZ DAS PROBABILIDADES DE COBERTURA
prob_cob <- matrix(0, nrow = length(n), ncol = 3)

prob_cob[, 1] <- n;

# MATRIZ DAS COMPRIMENTOS MÉDIOS DOS ICS
compr_ic <- matrix(0, nrow = length(n), ncol = 3);

compr_ic[,1] <- n;

# CALCULO DAS PROBABILIDADES DE COBERTURA PARA CADA n (tamanho amostral)
cont.falhas <- numeric(length(n));

set.seed(1)

for (i in 1:length(n)) {

  if (i == 1) {

    soma_compr.l0 <- 0; soma_compr.l1 <- 0;

    total_l0 <- 0; total_l1 <- 0;

    amostras <- matrix(0, nrow = S, ncol = n[i]);

    for (j in 1:S) {

      continua <- TRUE;

      while (continua) {

        amostras[i,j] <- rnd_pp2(n[i], l0, l1);

        ic <- ic_pp2(amostras[i,j]);

        if (any(is.nan(c(ic$l0, ic$l1))))

          cont.falhas[i] <- cont.falhas[i] + 1 else continua <- FALSE;

      } # end-while

      if ((ic$l0[1] < l0) & (ic$l0[2] > l0)) total_l0 <- total_l0 + 1;

      if ((ic$l1[1] < l1) & (ic$l1[2] > l1)) total_l1 <- total_l1 + 1;

      soma_compr.l0 <- soma_compr.l0 + ic$l0[2] - ic$l0[1];

      soma_compr.l1 <- soma_compr.l1 + ic$l1[2] - ic$l1[1];

      print(paste("iteração",j, "---->n =", n[i]));

    }

  }

}
```

```

} # end-for

prob_cob[i,2:3] <- c(total_10/S, total_11/S);

compr_ic[i,2:3] <- c(soma_compr.10/S, soma_compr.11/S);

} else {

amostras.new <- matrix(0, nrow = S, ncol = n[i]);

soma_compr.10 <- 0; soma_compr.11 <- 0;

total_10 <- 0; total_11 <- 0;

for (j in 1:S) {

continua <- TRUE;

while (continua) {

amostras.new[i,] <- amostras[i, sample(ncol(amostras), n[i])];

ic <- ic_ppe2(amostras.new[i,]);

if (any(is.nan(c(ic$10, ic$11))))

cont.falhas[i] <- cont.falhas[i] + 1 else continua <- FALSE;

} # end-while

if ((ic$10[1] < 10) & (ic$10[2] > 10)) total_10 <- total_10 + 1;

if ((ic$11[1] < 11) & (ic$11[2] > 11)) total_11 <- total_11 + 1;

soma_compr.10 <- soma_compr.10 + ic$10[2] - ic$10[1];

soma_compr.11 <- soma_compr.11 + ic$11[2] - ic$11[1];

print(paste("iteração",j, "-->n =", n[i]));

} # end-for

prob_cob[i, 2:3] <- c(total_10/S, total_11/S);

compr_ic[i, 2:3] <- c(soma_compr.10/S, soma_compr.11/S);

amostras <- amostras.new;

} # end-if-else

} # end-for

```

D.2 Modelo PPE2 - Teste de Hipóteses

```

# FUNÇÃO (-1)*LOG-VEROSSIMILHANÇA MODELO POISSON (SOB H0)

ll1_pois <- function(l1, X) {(-1)*ll(l1, l1, X)};

```



```

# FUNÇÃO QUE CALCULA ERV

require(stats4)

erv_ppe2 <- function(amostra) {

  emv <- try(mle(l11, start = list(l0 = -0.5, l1 = -0.5), fixed = list(X = amostra), method="BFGS"), silent = TRUE);

  emv_pois <- try(mle(l11_pois, start = list(l1 = -0.5), fixed = list(X = amostra), method="BFGS"), silent = TRUE);

  if ((class(emv)== "try-error") || (class(emv_pois)== "try-error")) return(NaN) else {

    l0_max <- as.numeric(coef(emv)[1]);

    l1_max <- as.numeric(coef(emv)[2]);

    l1.pois_max <- as.numeric(coef(emv_pois)[1]);

    return(2*l1_pois(l1.pois_max, amostra) - 2*l1(l0_max, l1_max, amostra));} # end-if

  } # end-function

# SIMULAÇÃO - TESTE DE HIPOTÉSES

# VALORES VERDADEIROS PARA OS PARÂMETROS (SOB H0)

l0 <- -0.5;

l1 <- -0.5;

# QTDE DE SIMULAÇÕES

S <- 1000;

# DETERMINAÇÃO DO VALOR CRÍTICO DO TESTE

alpha <- 0.05;

valor.critico <- qchisq(1 - alpha, 1)

# VETOR DOS TAMANHOS AMOSTRAIS

n <- c(400, 150, 80, 30)

# MATRIZ DOS NÍVEIS DE SIGNIFICÂNCIA OBSERVADOS PARA CADA n (tamanho amostral)

nivel.obs <- matrix(0, nrow = length(n), ncol = 2);

nivel.obs[,1] <- n;

# CÁLCULO DOS NÍVEIS DE SIGNIFICÂNCIA EMPÍRICOS PARA CADA n (tamanho amostral)

cont.falhas <- numeric(length(n));

set.seed(0);

for (i in 1:length(n)) {# Calcula ERV e conta quantas vezes H0 foi rejeitada

  if (i == 1) {

    total_rejeita <- 0;

```

```
amostras <- matrix(0, nrow = S, ncol = n[i]);

for (j in 1:S) {

continua <- TRUE;

while (continua) {

amostras[j,] <- rnd_ppe2(n[i], 10, 11);

erv <- erv_ppe2(amostras[j,]);

if (!is.nan(erv)){

continua <- FALSE} else cont.falhas[i] <- cont.falhas[i] + 1;

} # end-while

if (erv > valor.critico) total_rejeita <- total_rejeita + 1;

print(paste("iteracao",j,"-> n=",n[i]))

} # end-for

nivel.obs[i, 2] <- total_rejeita/S;

} else {

total_rejeita <- 0;

amostras.new <- matrix(0, nrow = S, ncol = n[i]);

for (j in 1:S) {# Calcula ERV e conta quantas vezes H0 foi rejeitada

continua <- TRUE;

while (continua) {

amostras.new[j,] <- amostras[j, sample(ncol(amostras), n[i])];

erv <- erv_ppe2(amostras.new[j,]);

if (!is.nan(erv)){

continua <- FALSE} else cont.falhas[i] <- cont.falhas[i] + 1;

} # end-while

if (erv > valor.critico) total_rejeita <- total_rejeita + 1;

print(paste("iteracao",j,"-> n=",n[i]));

} # end-for

nivel.obs[i, 2] <- total_rejeita/S;

amostras <- amostras.new;

} # end-if-else

} # end-for
```

D.3 Modelo PPE2 - Decaimento da Variância

```
# FUNÇÃO QUE CALCULA AS VARIÂNCIAS

require(stats4)

var_ppe2 <- function(amostra) {

  emv <- try(mle(l11, start = list(l0 = -0.37, l1 = 1.5), fixed = list(X = amostra), method="Nelder-Mead"), silent = TRUE)

  if (class(emv) == "try-error") return(list(l0 = NaN, l1 = NaN)) else {

    # VARIÂNCIAS DOS ESTIMADORES USANDO A MATRIZ OBSERVADA DE FISHER

    var_l0 <- as.numeric(vcov(emv)[1,1])

    var_l1 <- as.numeric(vcov(emv)[2,2])

    return(list(l0 = var_l0, l1 = var_l1))} # end-if

  } # end-function

# SIMULAÇÃO - DECAIMENTO DA VARIÂNCIA

# VALORES VERDADEIROS PARA OS PARAMETROS

l0 <- -0.37

l1 <- 1.5

# VETOR DOS TAMANHOS AMOSTRAIS

n <- c(400, 300, 200, 150, 125, 100, 80, 60, 40, 30)

# QTDE DE SIMULAÇÕES PARA CADA n

S <- 1000;

# MATRIZ DAS VARIÂNCIAS

variancia <- list();

variancia$10 <- matrix(0, nrow = S, ncol = length(n));

variancia$11 <- matrix(0, nrow = S, ncol = length(n));

colnames(variancia$10) <- as.character(n);

colnames(variancia$11) <- as.character(n);

# CALCULO DAS VARIÂNCIAS DOS ESTIMADORES PARA CADA n (tamanho amostral)

cont.falhas <- numeric(length(n));

set.seed(1234)

for (i in 1:length(n)) {

  if (i == 1) {
```

```
amostras <- matrix(0, nrow = S, ncol = n[i]);

for (j in 1:S) {

continua <- TRUE;

while (continua) {

amostras[j,] <- rnd_ppe2(n[i], 10, 11);

var <- var_ppe2(amostras[j,]);

if (any(is.nan(c(var$10, var$11))) || any(c(var$10, var$11) <= 0))

cont.falhas[i] <- cont.falhas[i] + 1 else

continua <- FALSE;

} # end-while

variancia$10[j, as.character(n[i])] <- var$10;

variancia$11[j, as.character(n[i])] <- var$11;

print(paste("iteração",j, "-->n =", n[i]));

} # end-for

} else {

amostras.new <- matrix(0, nrow = S, ncol = n[i]);

for (j in 1:S) {

continua <- TRUE;

while (continua) {

amostras.new[j,] <- amostras[j, sample(ncol(amostras), n[i])];

var <- var_ppe2(amostras[j,]);

if (any(is.nan(c(var$10, var$11))) || any(c(var$10, var$11) <= 0))

cont.falhas[i] <- cont.falhas[i] + 1 else continua <- FALSE;

} # end-while

variancia$10[j, as.character(n[i])] <- var$10

variancia$11[j, as.character(n[i])] <- var$11

print(paste("iteração",j, "-->n =", n[i]));

} # end-for

amostras <- amostras.new;

} # end-if

} # end-for
```

```

# ANÁLISE DO DECAIMENTO DA VARIÂNCIA EM FUNÇÃO DE n

# GRÁFICO DE DISPERSÃO

# 10
y <- c(variancia$10[, "30"], variancia$10[, "40"], variancia$10[, "60"], variancia$10[, "80"], variancia$10[, "100"],
      variancia$10[, "125"], variancia$10[, "150"], variancia$10[, "200"], variancia$10[, "300"], variancia$10[, "400"]);
x <- c(rep(30, S), rep(40, S), rep(60, S), rep(80, S), rep(100, S), rep(125, S), rep(150, S), rep(200, S),
      rep(300, S), rep(400, S));
y <- log(y); x <- log(x); plot(x,y, main = "Relação log log da variância de 10 sobre n", xlab = "log(n)", ylab = "log(var)");
abline(lsfit(x,y)); x11()

# 11
y <- c(variancia$11[, "30"], variancia$11[, "40"], variancia$11[, "60"], variancia$11[, "80"], variancia$11[, "100"],
      variancia$11[, "125"], variancia$11[, "150"], variancia$11[, "200"], variancia$11[, "300"], variancia$11[, "400"]);
x <- c(rep(30, S), rep(40, S), rep(60, S), rep(80, S), rep(100, S), rep(125, S), rep(150, S), rep(200, S),
      rep(300, S), rep(400, S));
y <- log(y); x <- log(x); plot(x,y, main = "Relação log log da variância de 11 sobre n", xlab = "log(n)", ylab = "log(var)");
abline(lsfit(x,y))

lsfit(x,y)$coefficients

# INTERVALO 30-80

# 10
y <- c(variancia$10[, "30"], variancia$10[, "40"], variancia$10[, "60"], variancia$10[, "80"]);
x <- c(rep(30, S), rep(40, S), rep(60, S), rep(80, S))
y <- log(y); x <- log(x); plot(x,y, main = "Relação log log da variância de 10 sobre n", xlab = "log(n)", ylab = "log(var)");
ajuste1.10 <- lm(y ~ x); abline(ajuste1.10$coefficients); ajuste1.10$coefficients

# 11
y <- c(variancia$11[, "30"], variancia$11[, "40"], variancia$11[, "60"], variancia$11[, "80"]);
x <- c(rep(30, S), rep(40, S), rep(60, S), rep(80, S))
y <- log(y); x <- log(x); plot(x,y, main = "Relação log log da variância de 11 sobre n", xlab = "log(n)", ylab = "log(var)");
ajuste1.11 <- lm(y ~ x); abline(ajuste1.11$coefficients); ajuste1.11$coefficients

# INTERVALO 80-150

# 10

```

```

y <- c(variancia$10[, "80"], variancia$10[, "100"], variancia$10[, "125"], variancia$10[, "150"]);
x <- c(rep(80, S), rep(100, S), rep(125, S), rep(150, S))
y <- log(y); x <- log(x); plot(x,y, main = "Relação log log da variância de 10 sobre n", xlab = "log(n)", ylab = "log(var)");
ajuste2.10 <- lm(y ~ x); abline(ajuste2.10$coefficients); ajuste2.10$coefficients
x11()

# 11
y <- c(variancia$11[, "80"], variancia$11[, "100"], variancia$11[, "125"], variancia$11[, "150"]);
x <- c(rep(80, S), rep(100, S), rep(125, S), rep(150, S))
y <- log(y); x <- log(x); plot(x,y, main = "Relação log log da variância de 11 sobre n", xlab = "log(n)", ylab = "log(var)");
ajuste2.11 <- lm(y ~ x); abline(ajuste2.11$coefficients); ajuste2.11$coefficients

# INTERVALO 150-400
# 10
y <- c(variancia$10[, "150"], variancia$10[, "200"], variancia$10[, "300"], variancia$10[, "400"]);
x <- c(rep(150, S), rep(200, S), rep(300, S), rep(400, S))
y <- log(y); x <- log(x); plot(x,y, main = "Relação log log da variância de 10 sobre n", xlab = "log(n)", ylab = "log(var)");
ajuste3.10 <- lm(y ~ x); abline(ajuste3.10$coefficients); ajuste3.10$coefficients
x11()
# 11
y <- c(variancia$11[, "150"], variancia$11[, "200"], variancia$11[, "300"], variancia$11[, "400"]);
x <- c(rep(150, S), rep(200, S), rep(300, S), rep(400, S))
y <- log(y); x <- log(x); plot(x,y, main = "Relação log log da variância de 11 sobre n", xlab = "log(n)", ylab = "log(var)");
ajuste3.11 <- lm(y ~ x); abline(ajuste3.11$coefficients); ajuste3.11$coefficients

```

D.4 Modelo PPE4 - Probabilidade de Cobertura

```

# TAXAS DE TRANSIÇÃO
taxas_ppe4 <- function(n, l0, la, lb){
if (n==0) return(exp(l0)) else return(exp(la)*(exp(lb)+n-1));
}# end-function

# GERADOR DE AMOSTRAS

```

```

rnd_ppe4 <- function(n, l0, la, lb) {
vetor <- numeric(n)
for (j in 1:n) {
i <- 0; p <- 1;
while (p <= exp(1)) {
u <- runif(1)
p <- p*(1/u)^(1/taxas_ppe4(i, l0, la, lb))
i <- i + 1
} # end-while
vetor[j] <- i - 1;
} # end-for
return(vetor)
} # end-function

# FUNÇÃO QUE RETORNA A COMBINAÇÃO DE x e k, x REAL e k INTEIRO NÃO-NEGATIVO
combina <- function(x, k) {
if ( k == 0) y <- 1
else{
y <- prod((x - k + 1):x)/factorial(k);
return(y)}

# FUNÇÃO LOG-VEROSSIMILHANÇA MODELO PPE4
integrando <- function(x, m, v, c) {
exp(m*x)*(1 - exp(v*x))^c}
ll <- function(l0, la, lb, X){
soma <- 0; n <- length(X); n0 <- sum(X==0);
for (i in 1:n) {
if (X[i] > 0) soma <- soma + log(combina(exp(lb)+X[i]-2,X[i]-1)) +
log(integrate(integrando, 0, 1, m = exp(l0)-exp(la+lb), v = -exp(la), c = X[i]-1)$value)
} # end-for
y <- soma + (n - n0)*l0 - n*exp(l0)
return(y);

```

```

}

ll1 <- function(l0, la, lb, X) {(-1)*ll(l0, la, lb, X)} # (-1)*log_verossimilhança

# FUNÇÃO QUE CALCULA IC

require(stats4)

ic_ppe4 <- function(amostra, conf = 0.95) {# conf: nivel de confiança desejado

emv <- try(mle(ll1, start = list(l0 = -0.3, la = -0.5, lb = 0.5), fixed = list(X = amostra), method="Nelder-Mead"), silent
= TRUE)

if (class(emv)=="try-error") return(list(l0 = c(NaN, NaN), la = c(NaN, NaN), lb = c(NaN, NaN))) else {

# CALCULA INTERVALOS DE CONFIANÇA ASSINTÓTICOS USANDO A MATRIZ OBSERVADA DE FISHER

l0_inf <- as.numeric(coef(emv)[1] - qnorm((1 + conf)/2)*sqrt(vcov(emv)[1,1])) # limite inferior p/ l0

l0_sup <- as.numeric(coef(emv)[1] + qnorm((1 + conf)/2)*sqrt(vcov(emv)[1,1])) # limite superior p/ l0

la_inf <- as.numeric(coef(emv)[2] - qnorm((1 + conf)/2)*sqrt(vcov(emv)[2,2])) # limite inferior p/ la

la_sup <- as.numeric(coef(emv)[2] + qnorm((1 + conf)/2)*sqrt(vcov(emv)[2,2])) # limite superior p/ la

lb_inf <- as.numeric(coef(emv)[3] - qnorm((1 + conf)/2)*sqrt(vcov(emv)[3,3])) # limite inferior p/ lb

lb_sup <- as.numeric(coef(emv)[3] + qnorm((1 + conf)/2)*sqrt(vcov(emv)[3,3])) # limite superior p/ lb

return(list(l0 = c(l0_inf, l0_sup), la = c(la_inf, la_sup), lb = c(lb_inf, lb_sup)))} # end-if

} # end-function

# SIMULAÇÃO - PROBABILIDADES DE COBERTURA

# VALORES VERDADEIROS PARA OS PARAMETROS

l0 <- -0.37

la <- -0.5

lb <- 0.5

# QTDE DE AMOSTRAS

S <- 1000;

# VETOR DOS TAMANHOS AMOSTRAIS

n <- c(400, 150, 80, 30)

# MATRIZ DAS PROBABILIDADES DE COBERTURA

prob_cob <- matrix(0, nrow = length(n), ncol = 4);

prob_cob[,1] <- n;

# MATRIZ DAS COMPRIMENTOS MÉDIOS DOS ICS

```



```

compr_ic <- matrix(0, nrow = length(n), ncol = 4);

compr_ic[,1] <- n;

# CALCULO DAS PROBABILIDADES DE COBERTURA PARA CADA n (tamanho amostral)

cont.falhas <- numeric(length(n));

set.seed(12);

for (i in 1:length(n)) {

  if (i == 1) {

    soma_compr.l0 <- 0; soma_compr.la <- 0; soma_compr.lb <- 0;

    total_l0 <- 0; total_la <- 0; total_lb <- 0;

    amostras <- matrix(0, nrow = S, ncol = n[i]);

    for (j in 1:S) { # Calcula IC's e conta o nro de intervalos que obedeceram ao critério

      continua <- TRUE;

      while (continua) {

        amostras[i,j] <- rnd_ppe4(n[i], l0, la, lb);

        ic <- ic_ppe4(amostras[i,]);

        if (any(is.nan(c(ic$l0, ic$la, ic$lb))))

          cont.falhas[i] <- cont.falhas[i] + 1 else continua <- FALSE;

      } # end-while

      if ((ic$l0[1] < l0) & (ic$l0[2] > l0)) total_l0 <- total_l0 + 1;

      if ((ic$la[1] < la) & (ic$la[2] > la)) total_la <- total_la + 1;

      if ((ic$lb[1] < lb) & (ic$lb[2] > lb)) total_lb <- total_lb + 1;

      soma_compr.l0 <- soma_compr.l0 + ic$l0[2] - ic$l0[1];

      soma_compr.la <- soma_compr.la + ic$la[2] - ic$la[1];

      soma_compr.lb <- soma_compr.lb + ic$lb[2] - ic$lb[1]

      print(paste("iteracao",j,"--> n=",n[i]))

    } # end-for

    prob_cob[i, 2:4] <- c(total_l0/S, total_la/S, total_lb/S);

    compr_ic[i, 2:4] <- c(soma_compr.l0/S, soma_compr.la/S, soma_compr.lb/S);

  } else {

    amostras.new <- matrix(0, nrow = S, ncol = n[i]);

    soma_compr.l0 <- 0; soma_compr.la <- 0; soma_compr.l0 <- 0;

```

```

total_l0 <- 0; total_la <- 0; total_lb <- 0;

for (j in 1:S) { # Calcula IC's e conta o nro de intervalos que obedeceram ao critério

continua <- TRUE;

while (continua) {

amostras.new[i,] <- amostras[i, sample(ncol(amostras), n[i])];

ic <- ic_ppe4(amostras.new[i,]);

if (any(is.nan(c(ic$l0, ic$la, ic$lb))))

cont.falhas[j] <- cont.falhas[j] + 1 else continua <- FALSE;

} # end-while

if ((ic$l0[1] < l0) & (ic$l0[2] > l0)) total_l0 <- total_l0 + 1;

if ((ic$la[1] < la) & (ic$la[2] > la)) total_la <- total_la + 1;

if ((ic$lb[1] < lb) & (ic$lb[2] > lb)) total_lb <- total_lb + 1;

soma_compr.l0 <- soma_compr.l0 + ic$l0[2] - ic$l0[1];

soma_compr.la <- soma_compr.la + ic$la[2] - ic$la[1];

soma_compr.lb <- soma_compr.lb + ic$lb[2] - ic$lb[1];

print(paste("iteracao",j,"--> n=",n[i]));

} # end-for

prob_cob[i, 2:4] <- c(total_l0/S, total_la/S, total_lb/S);

compr_ic[i, 2:4] <- c(soma_compr.l0/S, soma_compr.la/S, soma_compr.lb/S);

amostras <- amostras.new;

} # end-if

} # end-for

prob_cob; compr_ic;

# IC para a probabilidade de cobertura nominal de 95%

linf <- 0.95 - 1.96*sqrt(0.95*0.05/1000)

lsup <- 0.95 + 1.96*sqrt(0.95*0.05/1000)

# Gráficos das probabilidades de cobertura em função de n

plot(prob_cob[,1], prob_cob[,2], type = "l", lty = 2, col = "red", xlim = c(0, 1000),

ylim = c(0.85,1), xlab = "n", ylab = "Probabilidade", main = "Probabilidades de cobertura")

lines(prob_cob[,1], prob_cob[,3], lty = 5, col = "blue", xlim = c(0,1000),

ylim = c(0.85,1))

```

```
lines(prob_cob[,1], prob_cob[,4], lty = 6, col = "green", xlim = c(0,1000),
ylim = c(0.85,1))
abline(h = lmf, col = "gray", lty = 2)
abline(h = lsup, col = "gray", lty = 2)
```

D.5 Modelo PPE4 - Teste de Hipóteses

```
# FUNÇÃO (-1)*LOG-VEROSSIMILHANÇA MODELO B. NEGATIVA (SOB H0)
ll1_bn <- function(la, lb, X) {(-1)*ll(la + log(exp(lb)-1), la, lb, X)}

# FUNÇÃO QUE CALCULA ERV
require(stats4)
erv_ppe4 <- function(amostra) {
  emv <- try(mle(ll1, start = list(l0 = -0.3, la = -0.5, lb = 0.5), fixed = list(X = amostra), method="Nelder-Mead"), silent
= TRUE)
  emv_bn <- try(mle(ll1_bn, start = list(la = -0.5, lb = 0.5), fixed = list(X = amostra), method="L-BFGS-B",
lower=c(-Inf, 0)), silent = TRUE);
  if ((class(emv)== "try-error") || (class(emv_bn)== "try-error")) return(NaN) else {
    l0_max <- as.numeric(coef(emv)[1]);
    la_max <- as.numeric(coef(emv)[2]);
    lb_max <- as.numeric(coef(emv)[3]);
    la.bn_max <- as.numeric(coef(emv_bn)[1]);
    lb.bn_max <- as.numeric(coef(emv_bn)[2]);
    return(2*ll1_bn(la.bn_max, lb.bn_max, amostra) - 2*ll1(l0_max, la_max, lb_max, amostra));} # end-if
} # end-function

# SIMULAÇÃO - TESTE DE HIPOTETES
# VALORES VERDADEIROS PARA OS PARAMETROS (SOB H0)
la <- -0.5
lb <- 0.5
l0 <- la + log(exp(lb)-1)
# QTDE DE SIMULAÇÕES
```

```
S <- 1000;

# DETERMINAÇÃO DO VALOR CRÍTICO DO TESTE

alpha <- 0.05;

valor.critico <- qchisq(1 - alpha, 1)

# VETOR DOS TAMANHOS AMOSTRAIS

n <- c(400, 150, 80, 30)

# MATRIZ DOS NÍVEIS DE SIGNIFICANCIA OBSERVADOS PARA CADA n (tamanho amostral)

nivel.obs <- matrix(0, nrow = length(n), ncol = 2);

nivel.obs[,1] <- n;

# CALCULO DOS NÍVEIS DE SIGNIFICÂNCIA EMPÍRICOS PARA CADA n (tamanho amostral)

cont.falhas <- numeric(length(n));

set.seed(1);

for (i in 1:length(n)) {# Calcula ERV e conta quantas vezes H0 foi rejeitada

  if (i == 1) {

    total_rejeita <- 0;

    amostras <- matrix(0, nrow = S, ncol = n[i]);

    for (j in 1:S) {

      continua <- TRUE;

      while (continua) {

        amostras[j,] <- rnd_ppe4(n[i], 10, la, lb);

        erv <- erv_ppe4(amostras[j,]);

        if (!is.nan(erv)){

          continua <- FALSE} else cont.falhas[i] <- cont.falhas[i] + 1;

        } # end-while

        if (erv > valor.critico) total_rejeita <- total_rejeita + 1;

        print(paste("iteracao",j,"--> n=",n[i]))

      } # end-for

      nivel.obs[i, 2] <- total_rejeita/S;

    } else {

      total_rejeita <- 0;

      amostras.new <- matrix(0, nrow = S, ncol = n[i]);
```

```

for (j in 1:S) {# Calcula ERV e conta quantas vezes H0 foi rejeitada

continua <- TRUE;

while (continua) {

amostras.new[j,] <- amostras[j, sample(ncol(amostras), n[i])];

erv <- erv_ppe4(amostras.new[j,]);

if (lis.nan(erv)){

continua <- FALSE} else cont.falhas[i] <- cont.falhas[i] + 1;

} # end-while

if (erv > valor.critico) total_rejeita <- total_rejeita + 1;

print(paste("iteracao",j,"----> n=", n[i]));

} # end-for

nivel.obs[i, 2] <- total_rejeita/S;

amostras <- amostras.new;

} # end-if

} # end-for

```

D.6 Modelo PPE4 - Decaimento da Variância

```

# FUNÇÃO QUE CALCULA AS VARIÂNCIAS

require(stats4)

var_ppe4 <- function(amostra) {

emv <- try(mle(l11, start = list(l0 = -0.37, la = -0.5, lb = 0.5), fixed = list(X = amostra), method="Nelder-Mead"),

silent = TRUE)

if (class(emv)=="try-error") return(list(l0 = NaN, la = NaN, lb = NaN)) else {

# VARIÂNCIAS DOS ESTIMADORES USANDO A MATRIZ OBSERVADA DE FISHER

var_l0 <- as.numeric(vcov(emv)[1,1])

var_la <- as.numeric(vcov(emv)[2,2])

var_lb <- as.numeric(vcov(emv)[3,3])

return(list(l0 = var_l0, la = var_la, lb = var_lb))} # end-if

} # end-function

# SIMULAÇÃO - VARIÂNCIAS

```

```
# VALORES VERDADEIROS PARA OS PARAMETROS

l0 <- -0.37

la <- -0.5

lb <- 0.5

# VETOR DOS TAMANHOS AMOSTRAIS

n <- c(400, 300, 200, 150, 125, 100, 80, 60, 40, 30)

# QTDE DE AMOSTRAS PARA CADA n

S <- 1000;

# MATRIZ DAS VARIÂNCIAS

variancia <- list();

variancia$l0 <- matrix(0, nrow = S, ncol = length(n));

variancia$la <- matrix(0, nrow = S, ncol = length(n));

variancia$lb <- matrix(0, nrow = S, ncol = length(n));

colnames(variancia$l0) <- as.character(n);

colnames(variancia$la) <- as.character(n);

colnames(variancia$lb) <- as.character(n);

# CALCULO DAS VARIÂNCIAS DOS ESTIMADORES PARA CADA n (Sanho amostral)

cont.falhas <- numeric(length(n));

set.seed(1)

for (i in 1:length(n)) {

  for (j in 1:S) {

    continua <- TRUE;

    while (continua) {

      amostra <- rnd_ppe4(n[i], l0, la, lb);

      var <- var_ppe4(amostra);

      if (any(is.nan(c(var$l0, var$la, var$lb))) || any(c(var$l0, var$la, var$lb) <= 0))

        continua <- FALSE;

    } # end-while

    variancia$l0[j, as.character(n[i])] <- var$l0;

    variancia$la[j, as.character(n[i])] <- var$la;

    variancia$lb[j, as.character(n[i])] <- var$lb;

  }

}
```

```

print(paste("iteração",j, "---->n =", n[i]));

} # end-for

} # end-for

# ANÁLISE DO DECAIMENTO DA VARIÂNCIA EM FUNÇÃO DE n

# GRÁFICO DE DISPERSÃO

# 10

y <- c(variancia$10[, "30"], variancia$10[, "40"], variancia$10[, "60"], variancia$10[, "80"], variancia$10[, "100"],
variancia$10[, "125"], variancia$10[, "150"], variancia$10[, "200"], variancia$10[, "300"], variancia$10[, "400"]);

x <- c(rep(30, S), rep(40, S), rep(60, S), rep(80, S), rep(100, S), rep(125, S), rep(150, S), rep(200, S),
rep(300, S), rep(400, S));

y <- log(y); x <- log(x); plot(x,y, main = "Relação log log da variância de 10 sobre n", xlab = "log(n)", ylab = "log(var)");
abline(lsfrit(x,y)); x11()

# la

y <- c(variancia$la[, "30"], variancia$la[, "40"], variancia$la[, "60"], variancia$la[, "80"], variancia$la[, "100"],
variancia$la[, "125"], variancia$la[, "150"], variancia$la[, "200"], variancia$la[, "300"], variancia$la[, "400"]);

x <- c(rep(30, S), rep(40, S), rep(60, S), rep(80, S), rep(100, S), rep(125, S), rep(150, S), rep(200, S),
rep(300, S), rep(400, S));

y <- log(y); x <- log(x); plot(x,y, main = "Relação log log da variância de la sobre n", xlab = "log(n)", ylab = "log(var)");
abline(lsfrit(x,y)); x11()

# lb

y <- c(variancia$lb[, "30"], variancia$lb[, "40"], variancia$lb[, "60"], variancia$lb[, "80"], variancia$lb[, "100"],
variancia$lb[, "125"], variancia$lb[, "150"], variancia$lb[, "200"], variancia$lb[, "300"], variancia$lb[, "400"]);

x <- c(rep(30, S), rep(40, S), rep(60, S), rep(80, S), rep(100, S), rep(125, S), rep(150, S), rep(200, S),
rep(300, S), rep(400, S));

y <- log(y); x <- log(x); plot(x,y, main = "Relação log log da variância de lb sobre n", xlab = "log(n)", ylab = "log(var)");
abline(lsfrit(x,y)); x11()

# INTERVALO 30-80

# 10

y <- c(variancia$10[, "30"],variancia$10[, "40"], variancia$10[, "60"], variancia$10[, "80"]);

x <- c(rep(30, S), rep(40, S), rep(60, S), rep(80, S))

```

```
y <- log(y); x <- log(x);

ajuste1.l0 <- lm(y ~x); ajuste1.l0$coefficients

# la

y <- c(variancia$la[, "30"],variancia$la[, "40"], variancia$la[, "60"], variancia$la[, "80"]);

x <- c(rep(30, S), rep(40, S), rep(60, S), rep(80, S))

y <- log(y); x <- log(x);

ajuste1.la <- lm(y ~x); ajuste1.la$coefficients

# lb

y <- c(variancia$lb[, "30"],variancia$lb[, "40"], variancia$lb[, "60"], variancia$lb[, "80"]);

x <- c(rep(30, S), rep(40, S), rep(60, S), rep(80, S))

y <- log(y); x <- log(x);

ajuste1.lb <- lm(y ~x); ajuste1.lb$coefficients

# INTERVALO 80-150

# l0

y <- c(variancia$l0[, "80"],variancia$l0[, "100"], variancia$l0[, "125"], variancia$l0[, "150"]);

x <- c(rep(80, S), rep(100, S), rep(125, S), rep(150, S))

y <- log(y); x <- log(x);

ajuste2.l0 <- lm(y ~x); ajuste2.l0$coefficients

# la

y <- c(variancia$la[, "80"],variancia$la[, "100"], variancia$la[, "125"], variancia$la[, "150"]);

x <- c(rep(80, S), rep(100, S), rep(125, S), rep(150, S))

y <- log(y); x <- log(x);

ajuste2.la <- lm(y ~x); ajuste2.la$coefficients

# lb

y <- c(variancia$lb[, "80"],variancia$lb[, "100"], variancia$lb[, "125"], variancia$lb[, "150"]);

x <- c(rep(80, S), rep(100, S), rep(125, S), rep(150, S))

y <- log(y); x <- log(x);

ajuste2.lb <- lm(y ~x); ajuste2.lb$coefficients

# INTERVALO 150-400

# l0
```



```
y <- c(variancia$10[, "150"], variancia$10[, "200"], variancia$10[, "300"], variancia$10[, "400"]);
x <- c(rep(150, S), rep(200, S), rep(300, S), rep(400, S))
y <- log(y); x <- log(x);
ajuste3.10 <- lm(y ~ x); ajuste3.10$coefficients
# la
y <- c(variancia$1a[, "150"], variancia$1a[, "200"], variancia$1a[, "300"], variancia$1a[, "400"]);
x <- c(rep(150, S), rep(200, S), rep(300, S), rep(400, S))
y <- log(y); x <- log(x);
ajuste3.1a <- lm(y ~ x); ajuste3.1a$coefficients
# lb
y <- c(variancia$1b[, "150"], variancia$1b[, "200"], variancia$1b[, "300"], variancia$1b[, "400"]);
x <- c(rep(150, S), rep(200, S), rep(300, S), rep(400, S))
y <- log(y); x <- log(x);
ajuste3.1b <- lm(y ~ x); ajuste3.1b$coefficients
```

Apêndice E

Rotinas do Capítulo 7

E.1 PPE2 - Prioris Não Informativas

```
# DADOS: NRO DE ACIDENTES EM OPERARIAS

dados <- scan("C:/dados_operarias.dat"); # VETOR DAS FREQUÊNCIAS OBSERVADAS

N <- length(dados) - 1; # MAIOR VALOR OBSERVADO

X <- rep(0, dados[1])    # VETOR DE OBSERVAÇÕES

for (i in 1:(length(dados) - 1)) {

v <- rep(i, dados[i+1])

X <- c(X, v)}

n <- length(X);        # TAMANHO AMOSTRAL

n0 <- sum(X==0);      # NRO DE ZEROS NA AMOSTRA

integrando <- function(x, m, b) {

x^(m)*exp(-b*x)}

# l0 ~ normal(mu.l0, sigma.l0), l1 ~ normal(mu.l1, sigma.l1)

mu.l0 <- 0; sigma.l0 <- 20;

mu.l1 <- 0; sigma.l1 <- 20;

# logaritmo da função densidade a posteriori

ln.post <- function(l0, l1) {

soma <- 0;

#Cálculo do somatório de integrais via integração numérica

for (i in 1:n) {
```

```

if (X[i] > 0) soma <- soma + log(integrate(integrando, 0, 1, m = X[i] - 1, b = exp(l1) - exp(l0))$value);}# end-for
valor <- -n*exp(l0) + (n-n0)*l0 + (sum(X)-n+n0)*l1 - (l0-mu.l0)^2/(2*sigma.l0^2) - (l1-mu.l1)^2/(2*sigma.l1^2) +
soma;
} # end-function

# MÉTODO DE SIMULAÇÃO MCMC - PASSEIO ALEATÓRIO

# CADEIA 1

x.l0 <- -1.7; x.l1 <- -1.5; # chutes iniciais para l0 e l1, respectivamente.

tau.l0 <- 0.2 # desvio padrão para v.a x1
tau.l1 <- 0.3 # desvio padrão para v.a x2

seq.l0 <- x.l0;
seq.l1 <- x.l1;

nac.l0 <- 0;
nac.l1 <- 0;

set.seed(1)

# SIMULAÇÃO

S <- 5000

for (t in 1:S) {

  # Gera candidato z.l0 para l0

  z.l0 <- rnorm(1, x.l0, tau.l0);

  alpha <- min(1, exp(ln.post(z.l0,x.l1)-ln.post(x.l0,x.l1)));

  u <- runif(1);

  if (u < alpha) {x.l0 <- z.l0; nac.l0 <- nac.l0 + 1}

  seq.l0 <- c(seq.l0, x.l0)

  # Gera candidato z.l1 para l1

  z.l1 <- rnorm(1, x.l1, tau.l1);

  alpha <- min(1, exp(ln.post(x.l0,z.l1)-ln.post(x.l0,x.l1)));

  u <- runif(1);

  if (u < alpha) {x.l1 <- z.l1; nac.l1 <- nac.l1 + 1}

  seq.l1 <- c(seq.l1, x.l1)

```

```
# Imprime estado da cadeia a cada 100 iteracoes
if (t %% 100 == 0) {
  print(paste("iteração ", t));
  print(paste("l0 ", x.l0));
  print(paste("l1 ", x.l1));
} # end-if
} # end-for
l0.c1 <- seq.l0
l1.c1 <- seq.l1

# TAXA DE ACEITAÇÃO PARA CADA PARAMETRO
txl0.c1 <- nac.l0/S
txl1.c1 <- nac.l1/S

# CADEIA 2
x.l0 <- 0; x.l1 <- 1; # chutes iniciais para l0 e l1, respectivamente.
tau.l0 <- 0.2 # desvio padrão para v.a x.l0
tau.l1 <- 0.3 # desvio padrão para v.a x.l1
seq.l0 <- x.l0; # sequencia da cadeia para l0
seq.l1 <- x.l1; # sequencia da cadeia para l1
nac.l0 <- 0;
nac.l1 <- 0;

set.seed(2)

# SIMULAÇÃO
S <- 5000
for (t in 1:S) {
  # Gera candidato z0 para x0
  z.l0 <- rnorm(1, x.l0, tau.l0);
  alpha <- min(1, exp(ln.post(z.l0,x.l1)-ln.post(x.l0,x.l1)));
  u <- runif(1);
  if (u < alpha) {x.l0 <- z.l0; nac.l0 <- nac.l0 + 1}
  seq.l0 <- c(seq.l0, x.l0)
```

```

# Gera candidato z1 para x1

z.l1 <- rnorm(1, x.l1, tau.l1);

alpha <- min(1, exp(ln.post(x.l0,z.l1)-ln.post(x.l0,x.l1)));

u <- runif(1);

if (u < alpha) {x.l1 <- z.l1; nac.l1 <- nac.l1 + 1}

seq.l1 <- c(seq.l1, x.l1)

# Imprime estado da cadeia a cada 100 iteracoes

if (t %% 100 == 0) {

print(paste("iteração ", t));

print(paste("l0 ", x.l0));

print(paste("l1 ", x.l1));

} # end-if

} # end-for

l0.c2 <- seq.l0

l1.c2 <- seq.l1

```

E.2 PPE2 - Prioris Informativas

```

# logaritmo da função densidade a posteriori

ln.post <- function(l0, l1, sigma) {

# Depende de sigma =c(sigma.l0, sigma.l1)

soma <- 0;

#Cálculo do somatório de integrais via integração numérica

for (i in 1:n) {

if (X[i] > 0) soma <- soma + log(integrate(integrando, 0, 1, m = X[i] - 1, b = exp(l1) - exp(l0))$value);}# end-for

valor <- -n*exp(l0) + (n-n0)*l0 + (sum(X)-n+n0)*l1 - (l0-mu.l0)^2/(2*sigma[1]^2) - (l1-mu.l1)^2/(2*sigma[2]^2) +

soma;

} # end-function

# Valores dos hiperparametros

mu.l0 <- 0; mu.l1 <- 0;

```

```
sigma <- cbind(c(1, 3, 5, 10), c(1, 3, 5, 10));

# Número de Simulações

S <- 10000

# Matrizes com os valores da simulação para cada sigma

l0.c1 <- matrix(0, nrow = nrow(sigma), ncol = S+1);

l1.c1 <- matrix(0, nrow = nrow(sigma), ncol = S+1);

l0.c2 <- matrix(0, nrow = nrow(sigma), ncol = S+1);

l1.c2 <- matrix(0, nrow = nrow(sigma), ncol = S+1);

# SIMULAÇÃO - CADEIA 1

set.seed(1)

for (i in 1:nrow(sigma)) {

  # Parâmetros iniciais da cadeia

  x.l0 <- -1.7; x.l1 <- -1.5; # chutes iniciais para l0 e l1, respectivamente.

  tau.l0 <- 0.5; # desvio padrão para v.a x.l0

  tau.l1 <- 0.5; # desvio padrão para v.a x.l1

  seq.l0 <- x.l0;

  seq.l1 <- x.l1;

  for (t in 1:S) {

    # Gera candidato z.l0 para l0

    z.l0 <- rnorm(1, x.l0, tau.l0);

    alpha <- min(1, exp(ln.post(z.l0,x.l1, sigma[i,])-ln.post(x.l0, x.l1, sigma[i,])));

    u <- runif(1);

    if (u < alpha) x.l0 <- z.l0;

    seq.l0 <- c(seq.l0, x.l0)

    # Gera candidato z.l1 para l1

    z.l1 <- rnorm(1, x.l1, tau.l1);

    alpha <- min(1, exp(ln.post(x.l0,z.l1, sigma[i,])-ln.post(x.l0,x.l1, sigma[i,])));

    u <- runif(1);

    if (u < alpha) x.l1 <- z.l1;

    seq.l1 <- c(seq.l1, x.l1)
```

```
# Imprime estado da cadeia a cada 100 iteracoes

if (t %% 100 == 0) {
print(paste(i,"º sigma"));
  print(paste("iteração ", t));
print(paste("l0 ", x.l0));
print(paste("l1 ", x.l1));
} # end-if
} # end-for

l0.c1[i,] <- seq.l0
l1.c1[i,] <- seq.l1
} # end-for

# SIMULAÇÃO - CADEIA 2
set.seed(2)

for (i in 1:nrow(sigma)) {
  # Parâmetros iniciais da cadeia
  x.l0 <- 0; x.l1 <- 1; # chutes iniciais para l0 e l1, respectivamente.
  tau.l0 <- 0.5; # desvio padrão para v.a x.l0
  tau.l1 <- 0.5; # desvio padrão para v.a x.l1
  seq.l0 <- x.l0;
  seq.l1 <- x.l1;
  for (t in 1:S) {
    # Gera candidato z.l0 para l0
    z.l0 <- rnorm(1, x.l0, tau.l0);
    alpha <- min(1, exp(ln.post(z.l0,x.l1, sigma[i,])-ln.post(x.l0, x.l1, sigma[i,])));
    u <- runif(1);
    if (u < alpha) x.l0 <- z.l0;
    seq.l0 <- c(seq.l0, x.l0)
    # Gera candidato z.l1 para l1
    z.l1 <- rnorm(1, x.l1, tau.l1);
    alpha <- min(1, exp(ln.post(x.l0,z.l1, sigma[i,])-ln.post(x.l0,x.l1, sigma[i,])));
    u <- runif(1);
```

```

if (u < alpha) x.l1 <- z.l1;

seq.l1 <- c(seq.l1, x.l1)

# Imprime estado da cadeia a cada 100 iteracoes

if (t %% 100 == 0) {
print(paste(i,"º sigma"));

  print(paste("iteração ", t));

print(paste("l0 ", x.l0));

print(paste("l1 ", x.l1));

} # end-if

} # end-for

l0.c2[i,] <- seq.l0

l1.c2[i,] <- seq.l1

} # end-for

```

E.3 PPE4 - Prioris Não Informativas

```

# FUNÇÃO QUE RETORNA A COMBINAÇÃO DE x e k, x REAL e k INTEIRO NÃO-NEGATIVO

combina <- function(x, k) {

if ( k == 0) y <- 1

else{

y <- prod((x - k + 1):x)/factorial(k);

return(y)}

integrando <- function(x, m, v, c) {

exp(m*x)*(1 - exp(v*x))^c}

# DADOS: NRO DE VISITAS AO DENTISTA

dados <- scan("c:/dados_dentista.dat");

N <- length(dados) - 1 # Maior valor observado na amostra

# ANÁLISE DESCRITIVA

X <- rep(0, dados[1])

for (i in 1:(length(dados) - 1)) {

```



```

v <- rep(i, dados[i+1])

X <- c(X, v)}

n <- length(X) # Quantidade de observações

n0 <- sum(X == 0)

# função logaritmo da densidade a posteriori

ln.post <- function(l0, la, lb) {

soma <- 0;

n <- length(X);

n0 <- sum(X == 0);

for (i in 1:n) {

if (X[i]>0) soma <- soma + log(combina(exp(lb)+X[i]-2,X[i]-1)) +

log(integrate(integrando, 0, 1, m = exp(l0)-exp(la+lb), v = -exp(la), c = X[i]-1)$value) # end-if

} # end-for

valor <- soma + (n-n0)*l0 - n*exp(l0) - (l0-mu.l0)^2/(2*sigma.l0^2) - (la-mu.la)^2/(2*sigma.la^2) -

(lb-mu.lb)^2/(2*sigma.lb^2);

return(valor)

}# end-function

# MÉTODO DE SIMULAÇÃO MCMC - PASSEIO ALEATÓRIO

# PRIORI NÃO-INFORMATIVA

mu.l0 <- 0; sigma.l0 <- 20;

mu.la <- 0; sigma.la <- 20;

mu.lb <- 0; sigma.lb <- 20;

# CADEIA 1

x.l0 <- 0; x.la <- 0; x.lb <- -1.2 # chutes iniciais para l0, la e lb, respectivamente.

tau.l0 <- 0.3 # desvio padrão para v.a x1

tau.la <- 0.3 # desvio padrão para v.a x2

tau.lb <- 0.3 # desvio padrão para v.a x2

seq.l0 <- x.l0;

seq.la <- x.la;

seq.lb <- x.lb;

```

```
nac.l0 <- 0;

nac.la <- 0;

nac.lb <- 0;

set.seed(0)

# SIMULAÇÃO

S <- 5000

for (t in 1:S) {

  # Gera candidato z.l0 para l0

  z.l0 <- rnorm(1, x.l0, tau.l0);

  alpha <- min(1, exp(ln.post(z.l0,x.la, x.lb) - ln.post(x.l0, x.la, x.lb)));

  u <- runif(1);

  if (u < alpha) {x.l0 <- z.l0; nac.l0 <- nac.l0 + 1} # end-if

  seq.l0 <- c(seq.l0, x.l0)

  # Gera candidato z.la para la

  z.la <- rnorm(1, x.la, tau.la);

  alpha <- min(1, exp(ln.post(x.l0,z.la,x.lb) - ln.post(x.l0, x.la, x.lb)));

  u <- runif(1);

  if (u < alpha) {x.la <- z.la; nac.la <- nac.la + 1}

  seq.la <- c(seq.la, x.la)

  # Gera candidato z.lb para lb

  z.lb <- rnorm(1, x.lb, tau.lb);

  alpha <- min(1, exp(ln.post(x.l0,x.la,z.lb) - ln.post(x.l0, x.la, x.lb)));

  u <- runif(1);

  if (u < alpha) {x.lb <- z.lb; nac.lb <- nac.lb + 1}

  seq.lb <- c(seq.lb, x.lb)

  if (t %% 100 == 0) {

    print(paste("iteração ", t));

    print(paste("l0 ", x.l0));

    print(paste("la ", x.la));

    print(paste("lb ", x.lb));
```

```
} # end-if

} # end-for

l0.c1 <- seq.l0

la.c1 <- seq.la

lb.c1 <- seq.lb

# TAXA DE ACEITAÇÃO PARA CADA PARAMETRO

txl0.c1 <- nac.l0/S

txla.c1 <- nac.la/S

txlb.c1 <- nac.lb/S

# CADEIA 2

x.l0 <- 1.5; x.la <- 1.2; x.lb <- 1.5 # chutes iniciais para l0 e l1, respectivamente.

tau.l0 <- 0.2 # desvio padrão para v.a x.l0

tau.la <- 0.2 # desvio padrão para v.a x.la

tau.lb <- 0.2 # desvio padrão para v.a x.lb

seq.l0 <- x.l0;

seq.la <- x.la;

seq.lb <- x.lb;

nac.l0 <- 0;

nac.la <- 0;

nac.lb <- 0;

set.seed(2)

# SIMULAÇÃO

S <- 5000

for (t in 1:S) {

  # Gera candidato z.l0 para l0

  z.l0 <- rnorm(1, x.l0, tau.l0);

  alpha <- min(1, exp(ln.post(z.l0,x.la, x.lb) - ln.post(x.l0, x.la, x.lb)));

  u <- runif(1);

  if (u < alpha) {x.l0 <- z.l0; nac.l0 <- nac.l0 + 1} # end-if

  seq.l0 <- c(seq.l0, x.l0)
```

```

# Gera candidato z.la para la
z.la <- rnorm(1, x.la, tau.la);
alpha <- min(1, exp(ln.post(x.l0,z.la,x.lb) - ln.post(x.l0, x.la, x.lb)));
u <- runif(1);
if (u < alpha) {x.la <- z.la; nac.la <- nac.la + 1}
seq.la <- c(seq.la, x.la)

# Gera candidato z.lb para lb
z.lb <- rnorm(1, x.lb, tau.lb);
alpha <- min(1, exp(ln.post(x.l0,x.la,z.lb) - ln.post(x.l0, x.la, x.lb)));
u <- runif(1);
if (u < alpha) {x.lb <- z.lb; nac.lb <- nac.lb + 1}
seq.lb <- c(seq.lb, x.lb)

if (t %% 100 == 0) {
print(paste("iteração ", t));
print(paste("l0 ", x.l0));
print(paste("la ", x.la));
print(paste("lb ", x.lb));
} # end-if
} # end-for

l0.c2 <- seq.l0
la.c2 <- seq.la
lb.c2 <- seq.lb

# TAXA DE ACEITAÇÃO PARA CADA PARAMETRO
txl0.c2 <- nac.l0/S
txla.c2 <- nac.la/S
txlb.c2 <- nac.lb/S

```

E.4 PPE4 - Prioris Informativas

```

# função logaritmo da densidade a posteriori
ln.post <- function(l0, la, lb, sigma) {

```

```

# Depende de sigma =c(sigma.l0, sigma.la, sigma.lb)

soma <- 0;

n <- length(X);

n0 <- sum(X == 0);

for (i in 1:n) {

  if (X[i]>0) soma <- soma + log(combina(exp(lb)+X[i]-2,X[i]-1)) +

  log(integrate(integrando, 0, 1, m = exp(l0)-exp(la+lb), v = -exp(la), c = X[i]-1)$value) # end-if

} # end-for

valor <- soma + (n-n0)*l0 - n*exp(l0) - (l0-mu.l0)^2/(2*sigma[1]^2) - (la-mu.la)^2/(2*sigma[2]^2) -

(lb-mu.lb)^2/(2*sigma[3]^2);

return(valor)

}# end-function

# Valor dos hiperparametros

mu.l0 <- 0; mu.la <- 0; mu.lb <- 0;

sigma <- cbind(c(1, 3, 5, 10), c(1, 3, 5, 10), c(1, 3, 5, 10));

# Número de Simulações por cadeia

S <- 10000

# Matrizes com os valores da simulação para cada sigma

l0.c1 <- matrix(0, nrow = nrow(sigma), ncol = S+1);

la.c1 <- matrix(0, nrow = nrow(sigma), ncol = S+1);

lb.c1 <- matrix(0, nrow = nrow(sigma), ncol = S+1);

l0.c2 <- matrix(0, nrow = nrow(sigma), ncol = S+1);

la.c2 <- matrix(0, nrow = nrow(sigma), ncol = S+1);

lb.c2 <- matrix(0, nrow = nrow(sigma), ncol = S+1);

set.seed(1)

# SIMULAÇÃO - CADEIA 1

for (i in 1:nrow(sigma)) {

  # Parametros iniciais da cadeia

  x.l0 <- 0.5; x.la <- 0.1; x.lb <- -0.5 # chutes iniciais para l0, la e lb, respectivamente.

  tau.l0 <- 0.4 # desvio padrão para v.a x.l0

```

```
tau.la <- 0.4 # desvio padrão para v.a x.la
tau.lb <- 0.4 # desvio padrão para v.a x.lb
seq.l0 <- x.l0;
seq.la <- x.la;
seq.lb <- x.lb;
for (t in 1:S) {
  # Gera candidato z.l0 para l0
  z.l0 <- rnorm(1, x.l0, tau.l0);
  alpha <- min(1, exp(ln.post(z.l0,x.la, x.lb, sigma[i,]) - ln.post(x.l0, x.la, x.lb, sigma[i,])));
  u <- runif(1);
  if (u < alpha) {x.l0 <- z.l0} # end-if
  seq.l0 <- c(seq.l0, x.l0)
  # Gera candidato z.la para la
  z.la <- rnorm(1, x.la, tau.la);
  alpha <- min(1, exp(ln.post(x.l0,z.la,x.lb, sigma[i,]) - ln.post(x.l0, x.la, x.lb, sigma[i,])));
  u <- runif(1);
  if (u < alpha) {x.la <- z.la}
  seq.la <- c(seq.la, x.la)

  # Gera candidato z.lb para lb
  z.lb <- rnorm(1, x.lb, tau.lb);
  alpha <- min(1, exp(ln.post(x.l0,x.la,z.lb, sigma[i,]) - ln.post(x.l0, x.la, x.lb, sigma[i,])));
  u <- runif(1);
  if (u < alpha) {x.lb <- z.lb}
  seq.lb <- c(seq.lb, x.lb)

  # Imprime estado da cadeia a cada 100 iteracoes
  if (t %% 100 == 0) {
print(paste(i,"o sigma"));
print(paste("iteração ", t, ", i, ", "o sigma"));
print(paste("l0 ", x.l0));
print(paste("la ", x.la));
```

```
print(paste("lb ", x.lb));
  } # end-if
} # end-for
l0.c1[i,] <- seq.l0
l1.c1[i,] <- seq.l1
l2.c1[i,] <- seq.l2
} # end-for

set.seed(2)

# SIMULAÇÃO - CADEIA 2

for (i in 1:nrow(sigma)) {

  # Parametros iniciais da cadeia

  x.l0 <- 0; x.l1 <- 0; x.l2 <- -1.2 # chutes iniciais para l0, l1 e l2, respectivamente.

  tau.l0 <- 0.4 # desvio padrão para v.a x.l0
  tau.l1 <- 0.4 # desvio padrão para v.a x.l1
  tau.l2 <- 0.4 # desvio padrão para v.a x.l2

  seq.l0 <- x.l0;
  seq.l1 <- x.l1;
  seq.l2 <- x.l2;

  for (t in 1:S) {

    # Gera candidato z.l0 para l0

    z.l0 <- rnorm(1, x.l0, tau.l0);

    alpha <- min(1, exp(ln.post(z.l0,x.l1, x.l2, sigma[i,]) - ln.post(x.l0, x.l1, x.l2, sigma[i,])));

    u <- runif(1);

    if (u < alpha) {x.l0 <- z.l0} # end-if

    seq.l0 <- c(seq.l0, x.l0)

    # Gera candidato z.l1 para l1

    z.l1 <- rnorm(1, x.l1, tau.l1);

    alpha <- min(1, exp(ln.post(x.l0,z.l1,x.l2, sigma[i,]) - ln.post(x.l0, x.l1, x.l2, sigma[i,])));

    u <- runif(1);

    if (u < alpha) {x.l1 <- z.l1}

    seq.l1 <- c(seq.l1, x.l1)
```

```
# Gera candidato z.lb para lb
z.lb <- rnorm(1, x.lb, tau.lb);
alpha <- min(1, exp(ln.post(x.l0,x.la,z.lb, sigma[i,]) - ln.post(x.l0, x.la, x.lb, sigma[i,])));
u <- runif(1);
if (u < alpha) {x.lb <- z.lb}
seq.lb <- c(seq.lb, x.lb)

# Imprime estado da cadeia a cada 100 iteracoes
if (t %% 100 == 0) {
print(paste(i, "o sigma"));
print(paste("iteração ", t, ", i, "o sigma"));
print(paste("l0 ", x.l0));
print(paste("la ", x.la));
print(paste("lb ", x.lb));
} # end-if
} # end-for
l0.c2[i,] <- seq.l0
la.c2[i,] <- seq.la
lb.c2[i,] <- seq.lb
} # end-for
```