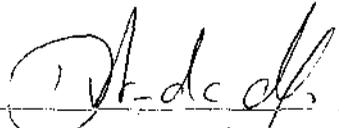


Análise de Dados Longitudinais com Respostas Categorizadas Através de Processos de Markov

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. Antônio de Queiroz Noletto e aprovada pela Comissão Julgadora.

Campinas, 10 de Maio de 1991.

Prof. Dr.


Dalton Francisco de Andrade †

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estatística.

22/9/90 511 X14

N717a

13984/BC

UNICAMP
BIBLIOTECA CENTRAL

Este trabalho é dedicado
a Antônio e Vanda, meus
pais.

Agradecimentos

Gostaria de agradecer a todos aqueles que de alguma forma colaboraram para a realização deste trabalho, em especial

aos professores e amigos Dalton Francisco de Andrade e Mauro Sérgio de Freitas Marques, cujas idéias e sugestões foram fundamentais no desenvolvimento deste trabalho;

ao amigo Sílvio Roberto Medeiros Evangelista, sem a ajuda do qual certamente não teria conseguido concluir a parte computacional deste trabalho;

ao amigo Ivo Dall’Agnol, que me introduziu na utilização do $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ usado para redigir o texto e cuja tese de mestrado foi de grande utilidade e

à Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), que com a política de capacitação profissional de seus técnicos, deu-me total tranquilidade e apoio para a realização deste projeto.

Conteúdo

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 2 | Resultados de Processos Estocásticos para Análise de Dados Longitudinais | 10 |
| 2.1 | Definições Preliminares | 10 |
| 2.2 | Cadeias de Markov | 12 |
| 2.3 | Processos de Markov ou de Saltos | 17 |
| 2.4 | Inferência Estatística Para Cadeias de Markov | 23 |
| 2.4.1 | Introdução | 23 |
| 2.4.2 | Estimação | 25 |
| 2.4.3 | Testes de Hipóteses | 31 |
| 3 | Análise Utilizando Modelos para Processos em Tempo Discreto | 39 |
| 3.1 | Introdução | 39 |
| 3.2 | Método de Stram, Wei & Ware para modelos marginais com respostas ordinais . | 41 |
| 3.3 | Adaptação do método de Stram, Wei & Ware para modelos de transição | 43 |
| 3.4 | Exemplo | 47 |
| 4 | Análise Utilizando Modelos para Processos em Tempo Contínuo | 52 |
| 4.1 | Introdução | 52 |
| 4.2 | Método de Kalbfleisch & Lawless | 53 |
| 4.2.1 | Estimação por máxima verossimilhança | 53 |
| 4.2.2 | Utilização do procedimento “Scoring” de Fisher | 56 |

| | | |
|----------|--|-----------|
| 4.2.3 | Algumas extensões para o modelo | 57 |
| 4.3 | A incorporação de variáveis exógenas | 59 |
| 4.4 | Exemplos | 61 |
| 5 | Apresentação dos Programas Computacionais Desenvolvidos | 67 |
| 5.1 | Introdução | 67 |
| 5.2 | Programas Computacionais Desenvolvidos | 68 |
| 5.3 | Conclusões Finais e Sugestões para Trabalhos Futuros | 78 |
| | Bibliografia | 80 |
| | Apêndice A | 84 |
| | Apêndice B | 87 |
| | Apêndice C | 92 |

Capítulo 1

Introdução

Estudos longitudinais são, de uma forma geral, estudos nos quais cada unidade experimental (indivíduo, planta, animal, etc.) é observada em mais de uma ocasião (tempo, distância, dosagem, etc.). Em geral, a ocasião tem o tempo como sendo sua forma mais utilizada e portanto, a menos que seja especificado, a ocasião que utilizamos é o tempo.

Exemplos de tais estudos são muito comuns na prática bem como na literatura. Plewis (1985), Andrade & Singer (1986), Landis *et al.* (1988), Dall'Agnol (1990) entre outros, descrevem várias situações que conduzem a estudos longitudinais nas mais diversas áreas. Vamos agora descrever um exemplo extraído de Koch *et al.* (1986) que será útil para uma melhor compreensão das questões que serão colocadas no decorrer dessa introdução.

EXEMPLO 1.1: 72 pacientes foram submetidos a um exame clínico para avaliar o desempenho de um tratamento na cura de uma enfermidade de pele. Os 72 pacientes foram divididos em dois grupos de mesmo tamanho. O grupo que recebeu um determinado tratamento foi denominado ativo e o outro que não recebeu tratamento foi denominado placebo. Cada paciente foi avaliado aos 3, 7, 10 e 14 dias após a aplicação do tratamento. Em cada dia de avaliação o paciente foi classificado conforme o estado da pele em uma das quatro categorias: 0- excelente, 1- bom, 2- satisfatório ou 3- ruim.

Esse experimento conduz a um estudo longitudinal pelo fato de cada unidade experimental (paciente) ter sido observada em quatro diferentes ocasiões (3, 7, 10 e 14 dias após aplicação do tratamento).

Basicamente as respostas a serem analisadas são divididas em dois tipos: aquelas que representam classes ou categorias como por exemplo, ausência ou presença de sintomas de uma certa doença, ou ainda a classificação de um indivíduo de acordo com o escore obtido através de um teste psicológico, que são denominadas respostas categóricas ou categorizadas e aquelas representando medidas como peso, altura, pressão sanguínea, idade, denominadas respostas contínuas. No exemplo 1.1, temos um caso onde as respostas são do tipo categóricas.

Trabalhos como os de Kowalski & Guire (1974), Cook & Ware (1983), Ware (1985), Andrade & Singer (1986), Zeger & Liang (1986), entre outros, discutem métodos para análise e fornecem levantamentos bibliográficos para o caso de respostas do tipo contínuo. Um trabalho bastante interessante que discute uma série de questões técnicas e conceituais encontradas no desenvolvimento de métodos para análise de dados longitudinais com respostas categorizadas e revisa a metodologia existente sobre o assunto é o de Ware, Lipsitz & Speizer (1988). Outros trabalhos importantes na área são o de Koch *et al.* (1977) que descreve uma metodologia para estudos longitudinais com respostas categorizadas e o de Stram, Wei & Ware (1988) que descreve uma metodologia para comparar respostas em grupos de indivíduos observados no mesmo conjunto de ocasiões, quando a resposta é categorizada ordinal. No trabalho de Dall'Agnol (1990) esses dois métodos para análise de dados em estudos longitudinais com respostas categorizadas são enfocados e discutidos através de exemplos.

O que determina fundamentalmente os métodos para a análise de dados longitudinais é o tipo de resposta avaliada, e de acordo com o tipo, quais as questões de interesse a serem analisadas. Segundo Andrade & Singer (1986), Koch *et al.* (1986) e Dall'Agnol (1990), os seguintes objetivos podem ser de interesse em uma análise de dados longitudinais:

1. Comparação dos diferentes tratamentos quanto ao padrão de variação das respectivas distribuições de respostas ao longo das ocasiões, isto é, verificar a existência de efeito da

interação entre tratamento e ocasião;

2. Comparação dos diferentes tratamentos quanto às respectivas distribuições médias de respostas, isto é, verificar a existência de efeito de tratamento;
3. Comparação das diferentes ocasiões quanto às respectivas distribuições médias de respostas, isto é, verificar existência de efeito das ocasiões;
4. Descrição da distribuição das mudanças individuais de categoria nas ocasiões e os efeitos das características individuais ou fatores de risco nessas mudanças.

Questões relacionadas aos objetivos (1), (2) e (3) podem ser respondidas utilizando-se as metodologias descritas nos trabalhos citados anteriormente, por exemplo, no caso de respostas categorizadas, ver Dall'Agnol (1990). Os modelos utilizados com essa finalidade são denominados, segundo Ware, Lipsitz & Speizer, modelos marginais, isto é, modelos cujo objetivo é o estudo da distribuição marginal da resposta em cada ocasião. O objetivo (4) é estudado acompanhando-se o comportamento das mudanças de categoria da unidade experimental ao longo das ocasiões. Os modelos que dão respostas às questões relacionadas a esse objetivo são denominados modelos de transição. Esses modelos têm como finalidade o estudo das probabilidades de transição entre as categorias em ocasiões sucessivas. Eles podem ser representados como distribuições de probabilidade para a categoria no futuro dada a história passada da unidade experimental de interesse. Por essa razão, são também conhecidos como modelos condicionais.

Para o exemplo 1.1 descrito anteriormente, questões relacionadas aos objetivos (1), (2) e (3) e que podem ser respondidas através da utilização de modelos marginais são:

- (m1) a partir de que dia houve uma melhoria sensível no estado da pele dos pacientes do grupo ativo?
- (m2) os pacientes do grupo placebo apresentaram melhorias em relação ao estado da pele?

- (m3) como o estado da pele se altera com o passar dos dias?
- (m4) a maneira como o estado da pele se modifica é diferente nos dois grupos?
- (m5) a distribuição dos pacientes nas categorias, em cada dia de avaliação é a mesma para os dois grupos?

Ainda para o exemplo 1.1, questões relacionadas ao objetivo (4), que podem ser respondidas utilizando-se modelos de transição, são, por exemplo:

- (t1) qual é a probabilidade de pacientes que tinham uma pele com estado satisfatório no décimo dia passarem a ter um bom estado de pele no décimo-quarto dia?
- (t2) como essa probabilidade é influenciada pelo tratamento?
- (t3) quanto tempo, em média, os pacientes do grupo ativo vão permanecer com o estado da pele ruim?
- (t4) em que grupo a probabilidade de passar do estado de pele satisfatório para bom é maior?
- (t5) essa probabilidade se altera com o tempo?
- (t6) existe uma tendência de que a pele atinja um estado excelente num período razoável de tempo para os pacientes do grupo ativo?

A distinção entre modelos marginais e de transição é muito importante metodologicamente. A dicotomia conceitual e técnica entre os modelos tem produzido uma divisão na literatura no que diz respeito aos métodos de análise. Em função disso, é importante observar que a escolha do modelo a ser utilizado é feita basicamente de acordo com os objetivos que se têm quando do planejamento do experimento.

Os modelos marginais não utilizam toda a informação contida nos dados gerados por estudos longitudinais. Em particular, eles não estão aptos a modelar as mudanças individuais

no tempo ou os efeitos de variáveis exógenas nessas mudanças, por exemplo, como um tipo de tratamento ou a cidade onde reside, podem influenciar nas mudanças ocorridas para as unidades experimentais em estudo. Nesses casos, os modelos de transição, isto é, modelos para mudanças individuais, devem ser utilizados. O estudo desses modelos é baseado na teoria de processos estocásticos, uma vez que o interesse se concentra na estimação das probabilidades de transição, isto é, as probabilidades de passar de uma categoria para outra em ocasiões sucessivas, que são os parâmetros que caracterizam esse tipo de modelo.

De acordo com o tipo de processo estocástico envolvido na análise, podemos dividir os modelos de transição em dois grupos: aquele para processos ocorrendo em tempo discreto e outro para processos em tempo contínuo. A diferença entre esses dois grupos deve ficar bem definida. Os processos em tempo discreto são aqueles onde as respostas são obtidas somente na ocasião da observação, por exemplo, através de um questionário sem o qual não é possível a obtenção de resposta. Os processos em tempo contínuo são aqueles que ocorrem a qualquer instante. O exemplo 1.1 constitui um caso de processo com tempo contínuo, porém as observações foram feitas em ocasiões pré-determinadas, que na prática, é o caso mais comum. Um exemplo de processo ocorrendo em tempo discreto pode ser visto em Ware *et al.* (1988), onde as respostas são obtidas através de um exame anual, de modo que a observação referente a cada criança consiste de uma série de estados definidos em intervalos de um ano. Kalbfleisch & Lawless (1985) fornecem um exemplo de processo em tempo contínuo onde são discutidas mudanças com relação ao hábito de fumar entre crianças. O processo nesse caso é caracterizado como sendo de tempo contínuo porque uma mudança de estado pode ocorrer a qualquer momento dentro do período delimitado para o estudo.

Uma maneira de analisar tais dados é modelar as probabilidades de transição, ou seja, modelar as probabilidades de mudanças de uma categoria para outra em ocasiões sucessivas. Frequentemente a análise de transições começa com a suposição que observações da mesma unidade experimental formam uma cadeia de Markov, isto é, a probabilidade de estar em uma determinada categoria em um tempo específico depende somente da categoria no tempo imediatamente anterior. Esse tipo de modelo tem sido utilizado satisfatoriamente em muitas aplicações,

frequentemente como uma base a partir da qual desenvolvem-se modelos mais elaborados e adequados.

A maioria dos métodos para análise de dados longitudinais baseados em cadeias de Markov, utiliza modelos saturados com a suposição que a amostra em estudo forma um grupo homogêneo, ou seja, todos os seus elementos têm as mesmas probabilidades de transição. Esse é o caso, por exemplo, de Goodman (1962) e Bishop *et al.* (1980) que mostram como conjuntos de dados representando mudanças de categoria de um período de tempo para outro podem ser analisados de forma bastante razoável através de técnicas estatísticas utilizadas em tabelas de contingência quadradas, testando as hipóteses de interesse através do ajuste de modelos log-lineares. Outro trabalho clássico sob o mesmo enfoque é o de Anderson & Goodman (1957) sobre inferência em cadeias de Markov.

Geralmente a amostra não pode ser tratada como sendo um grupo homogêneo, pelo fato de haver variáveis exógenas atuando de forma diferenciada dentro da mesma. Por exemplo, podemos considerar uma amostra na qual um grupo é composto por fumantes e outro por não fumantes e a partir dela estudar a evolução de uma enfermidade pulmonar. O que pode ser feito é uma estratificação da amostra em subgrupos que são homogêneos com relação aos valores assumidos pelas variáveis exógenas presentes no estudo e utilizá-la para investigar se as probabilidades de transição da cadeia de Markov diferem entre os subgrupos. Pode-se ainda verificar se a subdivisão é necessária, testando-se a hipótese de que as probabilidades de transição para os diferentes subgrupos são as mesmas. No caso de respostas binárias, Cox (1970) propôs o uso da regressão logística para quantificar os efeitos das variáveis exógenas nas probabilidades de transição de uma cadeia de Markov com dois estados. Korn & Whittemore (1979) e Muenz & Rubinstein (1985) utilizam esse enfoque na área biomédica. Esses trabalhos são voltados para respostas binárias, apesar do segundo permitir a generalização para mais de duas categorias com a desvantagem de aumentar muito o número de parâmetros. No trabalho de Korn & Whittemore, o estado ocupado pelo indivíduo no tempo imediatamente anterior é utilizado como sendo uma covariável. A mesma idéia foi utilizada por Ware, Lipsitz & Speizer (1988) no desenvolvimento de uma metodologia para modelar as probabilidades de transição utilizando-se um modelo de

transição para o caso de respostas ordinais.

Para muitos dos processos estocásticos analisados como estudos longitudinais, as transições ocorrem em tempo contínuo e não em tempo discreto. Os modelos de Markov em tempo contínuo são conhecidos como processos de Markov e têm tido uma grande aplicabilidade na área de ciências sociais, epidemiologia entre outras. Um caso especial é quando se tem disponível o número, tempo e sequência de todas as mudanças ocorridas em uma variável categorizada, para toda a amostra em estudo. Dados gerados dessa forma são conhecidos na literatura como “história do evento”. Para esse tipo de dados, métodos para análise de processos estocásticos em tempo contínuo são diretamente aplicáveis. O trabalho de Tuma *et al.* (1979), por exemplo, discute o uso de modelos de Markov em tempo contínuo tanto para processos que dependem do tempo de ocorrência como para processos independentes desse tempo, isto é, onde só interessa o tamanho do intervalo, descreve métodos para inclusão de variáveis exógenas nos modelos e ainda compara a análise da história do evento com outros tipos de estudos longitudinais existentes, concluindo que ela tem vantagens substanciais sobre os outros enfoques. Entretanto na prática, a obtenção dos dados dessa forma é muito difícil, já que nem sempre é possível um acompanhamento detalhado de cada unidade experimental em estudo. O que se faz para contornar esse problema é coletar os dados periodicamente e aproximar o processo real, contínuo, por um modelo discreto no qual as mudanças são assumidas como ocorridas na ocasião da observação. Essa cadeia em tempo discreto utilizada para explicar o processo em tempo contínuo é denominada a “cadeia embebida”.

Nos últimos anos tem sido crescente a utilização desses modelos, e isso fez com que se criasse uma necessidade de desenvolvimento de métodos de estimação dos parâmetros dos modelos contínuos através de observações discretas. Enquanto para os processos em tempo discreto, os parâmetros de interesse são as probabilidades de transição, nos processos em tempo contínuo o interesse se concentra nas taxas ou intensidades de transição, também conhecidos como parâmetros infinitesimais, uma vez que esses parâmetros caracterizam um processo de Markov. A utilização de cadeias embebidas acarreta alguns problemas relacionados à estimação das intensidades de transição. Algumas vezes, dados longitudinais categorizados não podem ser

embebidos no modelo, ou seja, descritos por um processo de Markov. Além disso, a variabilidade amostral e erros de medida podem tornar os dados não embebíveis mesmo que eles tenham sido originalmente gerados por um processo de Markov. Outro problema é que mesmo que os dados possam ser descritos por um processo de Markov, é possível que existam mais de uma matriz de intensidades de transição que descrevem o mesmo processo, e para cada matriz diferentes conclusões podem ser tiradas. Temos ainda, que pequenas mudanças em uma matriz de probabilidades de transição observada podem provocar mudanças radicais nas intensidades de transição estimadas. A estimação das intensidades de transição via dados longitudinais é sensível ao tamanho do intervalo de tempo entre observações. De acordo com Plewis (1985), observações que não se ajustam ao modelo de Markov em tempo discreto não se ajustarão ao modelo de Markov em tempo contínuo, porém se os dados são consistentes com uma cadeia de Markov, não há nada que garanta a consistência deles com um processo de Markov. Esses problemas foram resumidos por Singer & Spilerman (1976) em duas questões básicas:

- (a) “embeddability” ou como determinar se observações oriundas de um processo empírico podem ter sido geradas via a evolução de uma estrutura de Markov em tempo contínuo;
- (b) identificação ou o que fazer se as observações são consistentes com mais de uma estrutura de Markov em tempo contínuo e como selecionar a estrutura específica de uma lista de alternativas que estariam associadas com o processo empírico.

Entretanto, segundo Kalbfleisch & Lawless (1985), apesar da caracterização desses problemas ser matematicamente muito interessante e intrigante, do ponto de vista estatístico é irrelevante. Nesse trabalho, os autores explicam porque estimativas podem ser obtidas mesmo que os dados sejam não embebíveis e a matriz de covariância estimada dá a precisão dessas estimativas. O interesse então é obter estimativas para os parâmetros do modelo, ou seja, as intensidades de transição, bem como a estimativa da matriz de covariância dos estimadores.

Métodos para análise de modelos em tempo contínuo são discutidos por Albert (1962), Singer & Spilerman (1976), Singer & Cohen (1980), Wasserman (1980), Bartholomew (1983)

entre outros. Porém um problema existente é que grande parte da metodologia proposta tem sérias limitações tais como:

1. os tempos de observação têm que ser igualmente espaçados;
2. não se tem como fazer inferências sobre todas as características de interesse para o processo em questão;
3. não permitem a inclusão de variáveis exógenas no modelo.

Kalbfleisch & Lawless (1985) descrevem um método para estimação das intensidades de transição do processo em tempo contínuo a partir de dados obtidos através de observações periódicas, que não tem as limitações citadas acima. Para a estimação dos parâmetros, eles utilizam máxima verossimilhança e consideram o processo como sendo de tempo homogêneo, isto é, onde as probabilidades de transição não se alteram com o tempo. No trabalho de Phelan (1988), é descrita uma metodologia não paramétrica que permite a não homogeneidade do processo bem como a introdução de alguns esquemas de censura (saída ou entrada de unidades experimentais no decorrer do estudo).

Neste trabalho temos como objetivo a descrição de técnicas para análise de dados longitudinais através de modelos de transição quando a resposta é categorizada. No capítulo 2 fazemos um apanhado geral dos resultados de processos estocásticos que são necessários para o desenvolvimento dos capítulos posteriores. No capítulo 3 descrevemos o método de Ware, Lipsitz & Speizer para modelos de transição quando temos respostas ordinais obtidas através de um processo em tempo discreto. No capítulo 4 é discutido o método de Kalbfleisch & Lawless para análise de dados longitudinais, utilizando processos de Markov em tempo contínuo com observações em tempo discreto. No capítulo 5 são apresentados, através de um exemplo, os programas computacionais desenvolvidos.

O leitor interessado somente no que diz respeito à metodologia utilizada para análise de modelos de transição, ou aquele que está familiarizado com a teoria de processos estocásticos, pode omitir a leitura do capítulo 2 sem perda da sequência.

Capítulo 2

Resultados de Processos Estocásticos para Análise de Dados Longitudinais

2.1 Definições Preliminares

Vamos neste capítulo apresentar alguns conceitos e resultados básicos da teoria de processos estocásticos que são úteis no desenvolvimento do nosso trabalho. Deve-se observar que os resultados aqui expostos são apenas aqueles que de alguma forma serão utilizados no decorrer do trabalho. Um enfoque mais abrangente da teoria de processos estocásticos pode ser encontrado, por exemplo, em Hoel, Port & Stone (1972), Karlin & Taylor (1975), Cox & Miller (1977) e outros textos básicos.

A teoria de processos estocásticos lida com sistemas aleatórios ocorrendo no tempo ou espaço de acordo com certa distribuição de probabilidade. Aplicações da teoria podem ser utilizadas em uma grande variedade de fenômenos em muitos ramos da ciência e tecnologia.

Um *processo estocástico* pode ser definido de uma forma bem geral, como uma coleção de variáveis aleatórias $\{X(t); t \in \mathcal{T}\}$ definidas no mesmo espaço de probabilidades, indexadas por um conjunto arbitrário não vazio \mathcal{T} , onde \mathcal{T} é frequentemente um subconjunto dos números

reais. A menos que seja especificado, esses números representam os tempos de ocorrência do processo que se está estudando. Se \mathcal{T} é um conjunto de pontos finito ou infinito enumerável $\{t_1, t_2, \dots, t_T\}$ ou $\{t_1, t_2, \dots\}$ respectivamente, dizemos que o processo é de *tempo discreto*, se por outro lado, \mathcal{T} é composto por elementos t formando um intervalo contínuo não enumerável por exemplo, $0 \leq t \leq 4$ ou $t \geq 0$, temos um processo em *tempo contínuo*. No nosso estudo levamos em consideração dois casos específicos: processos em tempo discreto e processos em tempo contínuo, porém observados em apenas alguns pontos pré-determinados, os quais são discutidos com detalhes nos capítulos 3 e 4 respectivamente.

O valor assumido por $X(t)$ é denominado o *estado* do processo e o conjunto de todos os possíveis valores que as variáveis aleatórias $\{X(t); t \in \mathcal{T}\}$ podem assumir é chamado o *espaço de estados* do processo e é denotado por S . Se $S = \{1, 2, \dots\}$ ou $S = \{1, 2, \dots, k\}$ referenciamos-nos ao espaço de estados como sendo discreto, caso contrário, se S é um conjunto não enumerável contido no conjunto dos reais ($S \subset \mathfrak{R}$), o espaço de estados é contínuo. Neste trabalho abordamos somente problemas onde o espaço de estados S é discreto e finito representando as possíveis categorias de resposta.

Vamos considerar processos satisfazendo a seguinte propriedade:

$$\begin{aligned} \Pr\{a < X(s) \leq b \mid X(t_1) = x_1, \dots, X(t_n) = x_n, X(u) = x\} \\ = \Pr\{a < X(s) \leq b \mid X(u) = x\} \end{aligned} \tag{2.1}$$

onde $t_1 < \dots < t_n < u < s$. De acordo com essa propriedade, temos que $\{X(t); t \in \mathcal{T}\}$ é um processo estocástico com a propriedade que conhecido o estado ocupado em um específico tempo u , $X(s)$ onde $s > u$, não depende dos valores assumidos por $X(t_1), \dots, X(t_n)$, $t_1 < \dots < t_n < u$, isto é, a probabilidade de qualquer particular comportamento futuro do processo, quando seu presente estado é conhecido exatamente, não é alterada pelo conhecimento adicional relacionado ao seu comportamento no passado. Deve ser observado contudo, que se o conhecimento do estado presente do processo é impreciso, então a probabilidade de algum comportamento futuro será, em geral, alterada pela informação adicional relacionando o comportamento passado do sistema. A propriedade (2.1) é conhecida como *propriedade de Markov*, e os processos estocásticos que possuem essa propriedade são denominados *processos de Markov*.

Os processos de Markov em tempo discreto e espaço de estados discreto são denominados *cadeias de Markov*. No caso de processos de Markov em tempo contínuo e espaço de estados discreto temos o que se chama na literatura um *processo de Saltos*. Os processos de Markov com espaço de estados e tempo contínuos são denominados *processos de difusão*. Não vamos lidar com processos de difusão. Exposições detalhadas sobre o assunto são encontradas por exemplo, em Karlin & Taylor (1975) e Basawa & Prakasa Rao (1980) entre outros.

No âmbito deste trabalho o termo cadeias de Markov é utilizado para denominar os processos de Markov em tempo e espaço de estados discretos, enquanto que para os processos de Markov em tempo contínuo e espaço de estados discreto usamos a denominação geral, processos de Markov. Dessa forma, os termos processos de Markov, processos de Markov em tempo contínuo e processos de Saltos são utilizados indistintamente. Na seção 2.2 são discutidas as cadeias de Markov, através de definições de alguns conceitos básicos. O mesmo é feito na seção 2.3, para os processos de Saltos. A seção 2.4 é dedicada à descrição de alguns resultados relacionados à inferência estatística em cadeias de Markov.

2.2 Cadeias de Markov

Vamos considerar um sistema $\{X(t); t \in \mathcal{T}\}$ onde \mathcal{S} é um subconjunto finito dos inteiros e $\mathcal{T} = \{0, 1, \dots, T\}$. Nesse caso, temos um processo estocástico com tempo e espaço de estados discretos, onde a variável aleatória $X(t)$ denota o estado desse sistema, no tempo t . Além disso, é imposta a esse sistema uma estrutura de forma que os estados presente e passado tenham influência na distribuição de probabilidades do estado futuro. Conforme definido na seção anterior, os sistemas com essas características são conhecidos como cadeias de Markov. A propriedade de Markov é definida precisamente, nesse caso, como:

$$\begin{aligned} \Pr\{X(t+1) = j \mid X(0) = x_0, X(1) = x_1, \dots, X(t-1) = x_{t-1}, X(t) = i\} \\ = \Pr\{X(t+1) = j \mid X(t) = i\} \end{aligned} \tag{2.2}$$

para todo $t \in \mathcal{T}$ e valores $x_0, x_1, \dots, x_{t-1}, i$ e $j \in \mathcal{S}$. A interpretação de (2.2) é que a distribuição de probabilidade do futuro $X(t+1)$, condicionada ao passado e presente $X(0), \dots, X(t)$, só

depende do presente $X(t)$.

As probabilidades condicionais $\Pr\{X(t+1) = j \mid X(t) = i\}$ são denominadas as *probabilidades de transição* da cadeia de Markov e são denotadas por $p_{ij}(t, t+1)$. Essas probabilidades têm que satisfazer as seguintes condições:

$$1) \quad p_{ij}(t, t+1) \geq 0 \quad \forall i, j \in S \text{ e } \forall t \in T$$

$$2) \quad \sum_{j=1}^k p_{ij}(t, t+1) = 1 \quad \forall i \in S \text{ e } \forall t \in T.$$

É mais elegante matematicamente arranjar as probabilidades de transição na forma matricial:

$$P(t, t+1) = \begin{pmatrix} p_{11}(t, t+1) & \cdots & p_{1k}(t, t+1) \\ p_{21}(t, t+1) & \cdots & p_{2k}(t, t+1) \\ \vdots & \ddots & \vdots \\ p_{k1}(t, t+1) & \cdots & p_{kk}(t, t+1) \end{pmatrix}$$

e essa matriz é denominada a *matriz de probabilidades de transição* da cadeia de Markov.

Se $p_{ii}(t, t+1) = 1$ temos que o estado i é um estado absorvente, isto é, uma vez alcançado, é impossível mover-se para um estado diferente. Caso contrário, ele é um estado não absorvente.

Quando se trabalha com cadeias de Markov, o que se está fazendo é descrevendo em termos de probabilidades condicionais, as mudanças ou transições observadas entre os tempos.

É importante fazer uma observação relacionada ao termo “transição”. No contexto que estamos estudando, uma transição não necessariamente implica que os estados ocupados em ocasiões sucessivas sejam distintos, ou seja, é possível que exista uma probabilidade positiva do indivíduo que estava em um determinado estado no tempo t , permanecer no mesmo estado no tempo $t+1$ e ainda assim consideramos que houve uma transição, apesar de não ter ocorrido uma mudança de estado. Portanto, uma transição está relacionada com o tempo, ou seja, o que ocorre de uma ocasião para a outra.

Vamos agora descrever um exemplo com a finalidade de ilustrar a utilização de uma cadeia de Markov em um caso prático.

EXEMPLO 2.1: Suponhamos um componente eletrônico sujeito a falhas. Vamos considerar que o componente seja inspecionado diariamente e classificado em uma das três categorias a seguir: 0 - satisfatório, 1 - insatisfatório ou 2 - inutilizado. Temos portanto, um processo estocástico $\{X(t); t \in \mathcal{T}\}$ onde \mathcal{T} é um conjunto discreto, composto pelos dias de inspeção e $S = \{0, 1, 2\}$. O componente estar insatisfatório significa que apesar dele ainda estar funcionando, não passa pelo controle de qualidade, indicando que em pouco tempo vai falhar e ser inutilizado. Um possível modelo para esse sistema seria inicialmente supor que no tempo t , o processo está no estado 0, ou seja, $X(t) = 0$, de modo que as probabilidades de no tempo $t + 1$ o sistema passar para o estado 0, 1 ou 2 são respectivamente $p_{00}(t, t + 1)$, $p_{01}(t, t + 1)$ e $p_{02}(t, t + 1)$, com $p_{00}(t, t + 1) + p_{01}(t, t + 1) + p_{02}(t, t + 1) = 1$. Se o processo está no estado 1 no tempo t , isto é, se $X(t) = 1$, as probabilidades de no tempo $t + 1$ o sistema passar para o estado 0, 1 ou 2 são respectivamente 0, $p_{11}(t, t + 1)$ e $p_{12}(t, t + 1)$, com $p_{11}(t, t + 1) + p_{12}(t, t + 1) = 1$. Note que é impossível o componente passar do estado insatisfatório para o satisfatório. Finalmente assume-se que se o sistema está no estado 2 no tempo t , certamente permanecerá no mesmo estado no tempo $t + 1$, ou seja, o estado 2 é absorvente. A propriedade de Markov, nesse caso, é caracterizada pelo fato do estado do componente no dia seguinte, depender somente do estado no dia imediatamente anterior. Temos então uma cadeia de Markov cuja matriz de probabilidades de transição pode ser representada por:

$$P(t, t + 1) = \begin{pmatrix} p_{00}(t, t + 1) & p_{01}(t, t + 1) & p_{02}(t, t + 1) \\ 0 & p_{11}(t, t + 1) & p_{12}(t, t + 1) \\ 0 & 0 & 1 \end{pmatrix}$$

Quando as probabilidades de transição são independentes do tempo, isto é, quando são independentes do valor de t , dizemos que a cadeia de Markov tem probabilidades de transição es-

tacionárias. Nesse caso, as probabilidades de transição e a matriz de probabilidades de transição são denotadas respectivamente por p_{ij} e P . No caso do exemplo 2.1, supor que as probabilidades de transição são estacionárias, é considerar a seguinte matriz de probabilidades de transição:

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} \\ 0 & p_{11} & p_{12} \\ 0 & 0 & 1 \end{pmatrix}$$

que significa que as probabilidades de transição são mantidas, independentemente do dia da inspeção, por exemplo, para qualquer dia, a probabilidade de passar do estado 0 para o estado 1 no dia seguinte será sempre p_{01} , e assim para todas as transições.

A cadeia de Markov é completamente caracterizada estocasticamente se as probabilidades de transição e a distribuição de probabilidades (ou valor) do estado inicial, são especificados. Dessa forma, se η_t é o vetor coluna de probabilidades marginais no tempo t , temos:

$$\eta_s' = \eta_0' \times \prod_{t=1}^{s-1} P(t-1, t) \quad (2.3)$$

e se $P(t-1, t)$ é independente do tempo, isto é, se o processo é estacionário, $\eta_s' = \eta_0' \times P^{s-1}$. A demonstração desse resultado é dada em Karlin & Taylor (1975). Utilizando o exemplo 2.1, para o caso estacionário, temos:

$$\eta_t = \begin{pmatrix} \Pr\{X(t) = 0\} \\ \Pr\{X(t) = 1\} \\ \Pr\{X(t) = 2\} \end{pmatrix}.$$

As probabilidades marginais no tempo $t = 2$, por exemplo, são obtidas através de:

$$\eta_2 = \begin{pmatrix} \Pr\{X(2) = 0\} \\ \Pr\{X(2) = 1\} \\ \Pr\{X(2) = 2\} \end{pmatrix} = P^2 \times \begin{pmatrix} \Pr\{X(0) = 0\} \\ \Pr\{X(0) = 1\} \\ \Pr\{X(0) = 2\} \end{pmatrix}.$$

Utilizando-se a notação matricial, P é a matriz de probabilidades de transição estacionárias considerando tempos sucessivos. Denotamos a matriz de transição em n etapas, como $P^{(n)}$, uma matriz quadrada de ordem k , cujos elementos são:

$$p_{ij}^{(n)} = \Pr\{X(m+n) = j \mid X(m) = i\} \quad i, j \in S. \quad (2.4)$$

A condição Markoviana permite-nos expressar $p_{ij}^{(n)}$ em termos de p_{ij} de acordo com a *equação de Chapman-Kolmogorov*:

$$p_{ij}^{(m+n)} = \sum_{z \in S} p_{iz}^{(n)} p_{zj}^{(m)} \quad i, j \in S. \quad (2.5)$$

ou em termos matriciais $\mathbf{P}^{(m+n)} = \mathbf{P}^{(m)} \mathbf{P}^{(n)}$.

Um importante conceito que definimos agora é o de *distribuição estacionária* ou *distribuição de equilíbrio* de uma cadeia de Markov, denotada por π_j :

$$\pi_j = \sum_{i=1}^k \pi_i p_{ij} \quad \forall j \in S \quad (2.6)$$

onde $\pi_j \geq 0$ e $\sum_{j=1}^k \pi_j = 1$.

Qualquer conjunto $\{\pi_j; j \in S\}$ satisfazendo (2.6) é chamado a *distribuição de probabilidade estacionária* da cadeia de Markov.

Vamos assumir que exista uma distribuição de probabilidade estacionária $\{\pi_j; j \in S\}$ e que

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j \quad i, j \in S.$$

Então, independentemente da distribuição inicial, $\{X(t); t \in T\}$ tem distribuição aproximadamente igual a $\{\pi_j; j \in S\}$ quando $n \rightarrow \infty$. A distribuição da cadeia de Markov é independente de t se e somente se a distribuição inicial é uma distribuição estacionária, ou seja,

$$\Pr\{X(t) = j\} = \pi_j \Leftrightarrow \pi_j = \Pr\{X(0) = j\}.$$

Podemos concluir que

$$\lim_{t \rightarrow \infty} \Pr\{X(t) = j\} = \pi_j \quad j \in S.$$

Isso implica que a distribuição estacionária é única. Em termos gerais, um sistema possuir essa característica significa que após um longo período de tempo ele terá um comportamento estatisticamente estável.

Em muitas aplicações, a distribuição estacionária é o aspecto de maior interesse no processo. Isso ocorre principalmente porque é mais interessante para o pesquisador trabalhar com um processo estável, uma vez que alcançada essa estabilidade os resultados são mais confiáveis. Por exemplo, no caso do exemplo 2.1, suponhamos que na fabricação do componente eletrônico esteja envolvido um determinado equipamento que leva um certo tempo para atingir uma temperatura ideal. Podemos supor que os componentes fabricados fora da temperatura ideal têm uma maior probabilidade de tornarem-se insatisfatórios ou serem inutilizados, podendo levar-nos à conclusões errôneas quanto à sua durabilidade, enquanto que na temperatura ideal, os resultados obtidos serão mais condizentes com a realidade. A distribuição de probabilidade estacionária pode ser interpretada, nesse caso, como as probabilidades de transição considerando os componentes fabricados com o equipamento na temperatura ideal.

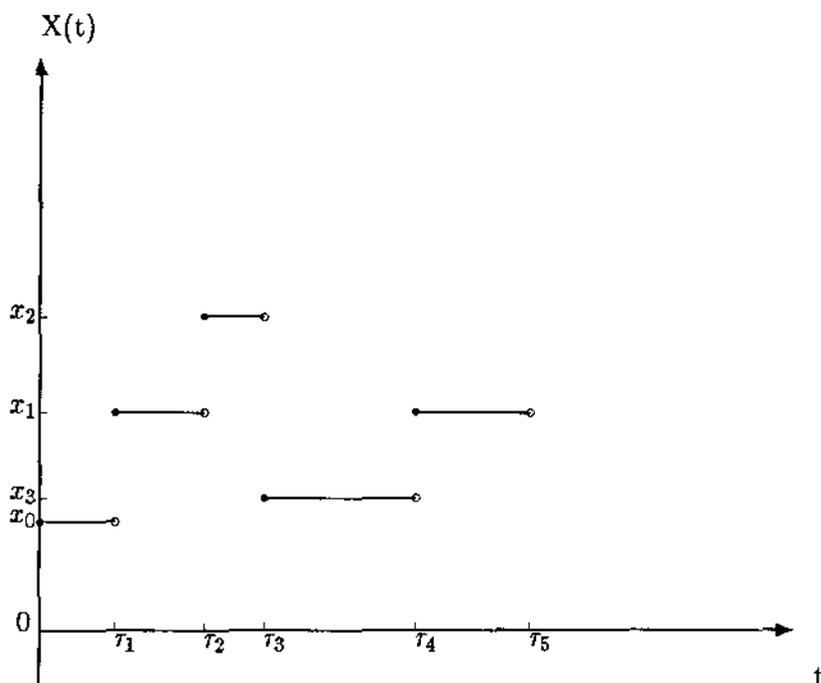
2.3 Processos de Markov ou de Saltos

Na seção 2.2 consideramos processos em tempo e espaço de estados discretos. Vamos agora discutir os casos onde o espaço de estados é discreto, porém o tempo é contínuo, também conhecidos como *processos de Markov* ou *processos de Saltos*.

Vamos considerar um sistema que está no estado x_0 inicialmente. Suponhamos que o sistema permaneça no estado x_0 até algum tempo positivo e aleatório τ_1 no qual ele salta para um novo estado $x_1 \neq x_0$, permanece em x_1 até algum tempo aleatório $\tau_2 > \tau_1$ quando salta para o estado $x_2 \neq x_1$. Note que não necessariamente $x_2 \neq x_0$. Esse procedimento é repetido indefinidamente. Deve-se observar que pode ocorrer do sistema alcançar um determinado estado e dele não mais sair, isto é, para algum l , $\tau_l = \infty$. O estado no qual o sistema se encontra no tempo t , denotado por $X(t)$ é definido por:

$$X(t) = \begin{cases} x_0 & 0 \leq t < \tau_1 \\ x_1 & \tau_1 \leq t < \tau_2 \\ x_2 & \tau_2 \leq t < \tau_3 \\ \vdots & \vdots \end{cases} \quad (2.7)$$

O processo definido em (2.7) é denominado um processo de Saltos. Vamos considerar processos de Saltos com a propriedade definida em (2.1) de modo que neste trabalho, processos com espaço de estados discreto e tempo contínuo são denominados processos de Markov (ou de Saltos). Um processo desse tipo pode ser representado graficamente como:



Assumimos que em um intervalo de tempo finito, o processo de Markov tem um número finito de saltos ou equivalentemente, $\lim_{n \rightarrow \infty} \tau_n = \infty$. Vamos especificar uma estrutura probabilística para tais processos. A cada estado i não absorvente, existe associada uma função distribuição $\mathcal{F}_i(t)$, $-\infty < t < \infty$, que assume valor 0 quando $t \leq 0$ e probabilidades de transição pr_{ij} , $i, j \in S$, não negativas e tais que:

$$\text{pr}_{ii} = 0$$

e

$$\sum_{j=1}^k \text{pr}_{ij} = 1.$$

Um processo iniciado em x_0 permanece nesse estado por um intervalo de tempo aleatório τ_1 com função distribuição $\mathcal{F}_{x_0}(t)$ e então salta para o estado x_1 com probabilidade

$\text{pr}_{x_0 x_1}, x_0, x_1 \in \mathcal{S}$. É assumido que τ_1 e $X(\tau_1)$ são escolhidos independentemente um do outro, ou seja,

$$\text{Pr}_{x_0} \{ \tau_1 \leq t, X(\tau_1) = x_1 \} = \mathcal{F}_{x_0}(t) \text{pr}_{x_0 x_1}. \quad (2.8)$$

A notação $\text{Pr}_{x_0} \{ \cdot \}$ denota a probabilidade de um evento definida em termos de um processo iniciado em x_0 . Sempre que o processo salta para um outro estado, ele atua como se o processo estivesse começando nesse estado, por exemplo, se x_0 e x_1 são estados não absorventes

$$\text{Pr}_{x_0} \{ \tau_1 \leq s, X(\tau_1) = x_1, \tau_2 - \tau_1 \leq t, X(\tau_2) = x_2 \} = \mathcal{F}_{x_0}(s) \text{pr}_{x_0 x_1} \mathcal{F}_{x_1}(t) \text{pr}_{x_1 x_2}.$$

Fórmulas semelhantes são obtidas para eventos definidos em termos de três ou mais saltos. Se i é um estado absorvente, $\text{pr}_{ij} = \delta_{ij}$, onde

$$\delta_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{caso contrário} \end{cases}$$

As probabilidades pr_{ij} são as probabilidades de transição de uma cadeia de Markov conhecida na literatura como *cadeia "embedded"* e que nós denominamos *cadeia embebida*. Essa cadeia pode ser utilizada para aproximar um processo contínuo no tempo, através de um processo em tempo discreto. Esse tipo de cadeia é importante quando se tem um processo contínuo onde as observações são realizadas em tempos discretos.

As probabilidades de transição do processo de Markov em tempo contínuo são definidas por:

$$p_{ij}(s, t) = \text{Pr}\{X(t) = j \mid X(s) = i\} \quad i, j \in \mathcal{S}, s < t. \quad (2.9)$$

Em termos matriciais

$$P(s, t) = \begin{pmatrix} p_{11}(s, t) & \cdots & p_{1k}(s, t) \\ p_{21}(s, t) & \cdots & p_{2k}(s, t) \\ \vdots & & \vdots \\ p_{k1}(s, t) & \cdots & p_{kk}(s, t) \end{pmatrix}. \quad (2.10)$$

Vamos assumir daqui em diante que as probabilidades de transição são estacionárias, isto é,

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

ou na notação matricial $\mathbf{P}(t) = \mathbf{P}(0, t) = \mathbf{P}(s, s + t)$, o que significa que $p_{ij}(s, t)$ só depende do intervalo de tempo, ou seja, da diferença $t - s$.

A propriedade de Markov garante as seguintes afirmações com respeito às probabilidades de transição:

$$(i) \quad p_{ij}(s, t) \geq 0;$$

$$(ii) \quad \sum_{j=1}^k p_{ij}(s, t) = 1;$$

$$(iii) \quad p_{ij}(0, s + t) = \sum_{l=1}^k p_{il}(0, s) p_{lj}(s, t) \quad \text{para } t, s \geq 0;$$

$$(iv) \quad \lim_{\Delta t \rightarrow 0} p_{ij}(t, t + \Delta t) = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j. \end{cases}$$

A afirmação (iii) representa a equação de Chapman-Kolmogorov, para o caso de processos de Markov e pode ser escrita na forma $\mathbf{P}(t + s) = \mathbf{P}(t) \times \mathbf{P}(s)$, $t, s \geq 0$.

Pode ser demonstrado que um processo de Saltos é Markoviano com função de probabilidade de transição estacionária se e somente se todos os estados não absorventes x são tais que:

$$\Pr_x\{\tau_1 > t + s \mid \tau_1 > s\} = \Pr_x\{\tau_1 > t\}$$

onde $s, t \geq 0$, ou em termos da função distribuição:

$$\frac{1 - \mathcal{F}_x(t + s)}{1 - \mathcal{F}_x(s)} = 1 - \mathcal{F}_x(t) \quad \text{com } s, t \geq 0. \quad (2.11)$$

Uma função distribuição $\mathcal{F}_x(t)$ satisfaz (2.11) se e somente se ela é uma função distribuição exponencial. Conclui-se portanto, que um processo de Saltos é Markoviano se e somente se $\mathcal{F}_x(t)$ é uma distribuição exponencial para todos estados não absorventes x .

Seja $\{X(t); t \geq 0\}$ um processo de Saltos Markoviano, ou seja, um processo de Markov. Se x é um estado não absorvente, então $\mathcal{F}_x(t)$ tem uma densidade exponencial $f_x(t)$. Denotamos

q_x , o parâmetro dessa densidade, conseqüentemente

$$f_x(t) = \begin{cases} q_x e^{-q_x t} & \text{se } t \geq 0 \\ 0 & \text{se } t < 0. \end{cases}$$

Se x é um estado absorvente assume-se que $q_x = 0$. Para o caso estacionário, é demonstrado que a probabilidade de transição $p_{ij}(t)$ satisfaz a seguinte equação:

$$p_{ij}(t) = \delta_{ij} e^{-q_i t} + \int_0^t q_i e^{-q_i s} \left(\sum_{l \neq i} p_{il} p_{lj}(t-s) \right) ds, \quad (2.12)$$

onde $t \geq 0$ (ver Hoel, Port & Stone (1972) - Cap. 3). De (2.12), temos que $p_{ij}(t)$ é contínua e diferenciável em t para $t \geq 0$ e a derivada de $p_{ij}(t)$ em relação a t é dada por:

$$\frac{\partial p_{ij}(t)}{\partial t} = -q_i p_{ij}(t) + q_i \sum_{l \neq i} p_{il} p_{lj}(t), \quad (2.13)$$

onde $t \geq 0$. Em particular,

$$\left. \frac{\partial p_{ij}(t)}{\partial t} \right|_{t=0} = -q_i \delta_{ij} + q_i p_{ij}.$$

Seja $q_{ij}(t) = (\partial p_{ij}(t)/\partial t)|_{t=0}$, $i, j \in S$, então

$$q_{ij}(t) = \begin{cases} -q_i & \text{se } j = i \\ q_i p_{ij} & \text{caso contrário.} \end{cases} \quad (2.14)$$

Segue de (2.14) que

$$\sum_{j \neq i} q_{ij}(t) = q_i = -q_{ii}.$$

As quantidades $q_{ij}(t)$, $i, j \in S$ são denominadas *parâmetros infinitesimais* ou *intensidades de transição* do processo de Markov. Outra forma de defini-las é através de:

$$\begin{aligned} q_{ij}(t) &= \lim_{\Delta t \rightarrow 0} \frac{p_{ij}(t, t + \Delta t) - p_{ij}(t)}{\Delta t}, & \text{se } i \neq j = 1, \dots, k \\ q_{ii}(t) &= -\sum_{j \neq i} q_{ij}(t), & \text{se } i \in S. \end{aligned} \quad (2.15)$$

Esses parâmetros determinam q_i e p_{ij} e portanto determinam o processo de Saltos Markoviano.

Denotamos por $Q(t)$, a matriz quadrada de ordem k , cujo (i, j) -ésimo elemento corresponde à $q_{ij}(t)$, que por sua vez é denominada *matriz de intensidades de transição*. Para

processos estacionários $q_{ij}(t) = q_{ij}$ e $Q(t) = Q$. As intensidades de transição podem ser interpretadas como taxas de transição num período muito pequeno de tempo e satisfazem $q_{ij}(t) \geq 0$ para $i \neq j$ e $\sum_{j=1}^k q_{ij}(t) = 0$. É interessante observar que a intensidade de transição de i para j ser igual a 0, não implica que a probabilidade de transição $p_{ij}(s, t)$ seja igual a 0. Entretanto, se $p_{ij}(s, t) = 0$, a intensidade de transição $q_{ij}(t)$ certamente é nula. Em termos práticos o que essa afirmação quer dizer é que às vezes, apesar de ser possível a transição do estado i para o estado j , essa passagem não ocorre de forma instantânea, isto é, em um intervalo de tempo muito pequeno.

Qualquer característica do processo pode ser definida utilizando-se as intensidades de transição, alguns exemplos são:

- $q_{ii}(t)$: taxa, no tempo t , de saída do estado i
- $q_{ii}^{-1}(t)$: tempo médio de permanência no estado i ;
- $\frac{-q_{ij}(t)}{q_{ii}(t)}$: probabilidade que um indivíduo no estado i mova-se para o estado j , dado que houve uma transição.

A equação (2.13) pode ser escrita em termos das intensidades de transição como:

$$\frac{\partial p_{ij}(t)}{\partial t} = \sum_{l=1}^k q_{il} p_{lj}(t), \quad t \geq 0. \quad (2.16)$$

Essa equação é denominada *equação retrospectiva* ("backward"). Em termos matriciais e assumindo estacionaridade, ou seja, $P(s, t) = P(t - s)$, temos

$$\frac{\partial P(s, t)}{\partial t} = Q(t)P(s, t). \quad (2.17)$$

Esse sistema de equações diferenciais para o caso estacionário tem a seguinte solução:

$$P(s, t) = \sum_{m=0}^{\infty} \frac{Q^m (t - s)^m}{m!} = e^{(t-s)Q} \quad (2.18)$$

onde Q^0 é a matriz identidade (para obtenção desse resultado ver Karlin & Taylor(1975)). Para processos não estacionários, não existe uma solução simples para $P(s, t)$ em termos de $Q(t)$ e essa é uma das razões para supormos inicialmente a estacionaridade do processo.

A distribuição estacionária para processos de Markov, é obtida através das intensidades de transição como a solução única para o sistema de equações $\sum_{i=1}^k \pi_i q_{ij} = 0$, $j = 1, \dots, k$. Utilizando-se notação matricial, $Q' \pi = 0$, onde $\pi = (\pi_1, \dots, \pi_k)'$ e $\sum_{j=1}^k \pi_j = 1$.

Se $X(0)$ tem uma distribuição π como sua distribuição inicial, então

$$\Pr\{X(t) = j\} = \sum_{i=1}^k \pi_i p_{ij}(t) = \pi_j \quad \forall j \in S$$

de modo que $X(t)$ tem distribuição π para todo $t \geq 0$. A idéia por trás da distribuição estacionária, no caso de processos ocorrendo em tempo contínuo, é a mesma descrita no caso de cadeias de Markov.

2.4 Inferência Estatística Para Cadeias de Markov

2.4.1 Introdução

A teoria de inferência estatística em processos estocásticos é bastante ampla. Uma referência sobre o assunto é o texto de Basawa & Prakasa Rao (1980). Uma outra mais específica, dedicada à inferência estatística para processos de Markov de um modo geral é a monografia de Billingsley (1961a). Os trabalhos de Anderson & Goodman (1957) e Billingsley (1961b) discutem inferência estatística para cadeias de Markov.

Como foi mencionado na seção 2.1, neste trabalho lidamos com dois tipos de processos: processos em tempo discreto e processos em tempo contínuo, porém observados em pontos discretos no tempo. No primeiro caso estamos lidando com cadeias de Markov diretamente. Todavia, no segundo não temos uma cadeia de Markov e sim um processo de Saltos e obviamente a forma de se fazer inferências sobre os parâmetros de interesse nesse caso, não pode ser baseada diretamente em uma cadeia de Markov. Entretanto, considerando a teoria descrita na seção 2.3, podemos mostrar que o comportamento do processo em tempo contínuo pode ser estudado através de um processo em tempo discreto, justificando assim, a utilização de cadeias de Markov para fazer inferências, mesmo que tenhamos um processo de Markov em tempo contínuo.

Vamos supor $\{X(t); t \geq 0\}$ um processo de Markov em tempo contínuo com espaço de

estados $S = \{1, 2, \dots, k\}$, definido de acordo com a seção 2.3. As quantidades q_i e q_{ij} conforme definidas em (2.14) têm o seguinte importante significado probabilístico. O processo inicia em algum estado $x_0 = i$, escolhido aleatoriamente de acordo com uma certa distribuição inicial, permanece em i por um período $\rho_1 = \tau_1$, onde ρ_1 é uma variável aleatória exponencialmente distribuída com parâmetro q_i . No tempo τ_1 o processo salta para um diferente estado j , escolhido de acordo com a distribuição q_{ij}/q_i ($j \neq i$) e permanece nesse estado por um período $\rho_2 = \tau_2 - \tau_1$ de tempo, que é exponencialmente distribuído com parâmetro q_j . No tempo $\tau_2 = \rho_1 + \rho_2$ o processo dá um novo salto para um estado $l \neq j$ conforme a distribuição q_{jl}/q_j e assim sucessivamente. Seja z_1, z_2, \dots , a sequência de estados pelos quais o processo $\{X(t); t \geq 0\}$ passa e ρ_1, ρ_2, \dots os períodos de tempo que o processo permanece em cada um desses estados. Se v_t é o número de saltos que ocorrem até o tempo t , isto é, se $v_t = \max(m : \rho_1 + \dots + \rho_m < t)$, então $X(t) = z_{v_t}$. O fato importante é que o processo definido pelo par $\{(z_m, \rho_m); m = 1, 2, \dots\}$ que denominamos na seção 2.3 como a cadeia embebida, é um processo de Markov em tempo discreto com espaço de estados definido no eixo cartesiano $X(t) \times (0, \infty)$ e probabilidades de transição

$$\Pr\{z_{m+1} = j, \rho_{m+1} \geq \alpha \mid z_m = i, \rho_m = \beta\} = \left(\frac{q_{ij}}{q_i}\right) e^{-q_i \alpha} = \text{pr}_{ij}(1 - \mathcal{F}_i(\alpha)),$$

conforme pode ser visto em Billingsley (1961b).

Conclui-se portanto, que a estrutura do processo de Saltos $\{X(t); t \geq 0\}$ pode ser determinada pela estrutura do processo em tempo discreto $\{(z_m, \rho_m); m = 1, 2, \dots\}$ (cadeia embebida). Se temos disponível uma amostra $\{(z_1, \rho_1), \dots, (z_m, \rho_m)\}$ extraída de uma cadeia embebida, é possível fazer inferências sobre os parâmetros de interesse, baseando-se na teoria para cadeias de Markov.

Em vista do que foi exposto, neste trabalho para fins de inferência estatística, precisamos apenas da teoria relacionada com cadeias de Markov, cuja descrição é o objetivo desta seção, que por sua vez é dividida em duas partes: estimação e testes de hipóteses. As duas primeiras subseções estão relacionadas à parte de estimação, enquanto que as demais têm como objetivo principal descrever de uma forma sucinta, vários testes de hipóteses para uma série de características da cadeia de Markov que podem ser de interesse do pesquisador. As referências citadas no início desta seção constituem a base para o desenvolvimento da mesma.

2.4.2 Estimação

Estimação dos Parâmetros

Nesta subseção, discutimos a teoria de estimação do vetor de probabilidades iniciais e da matriz de probabilidades de transição baseada no trabalho de Anderson & Goodman (1957).

Vamos assumir que o espaço de estados da cadeia de Markov tenha k elementos, ou seja, $\mathcal{S} = \{1, \dots, k\}$. Sejam $t = 0, 1, \dots, T$ os tempos de observação da cadeia de Markov $\{X(t); t \in \mathcal{T}\}$ e $X(0) = x_0, X(1) = x_1, \dots, X(T) = x_T$. Assumimos que existam $n_i(0)$ unidades experimentais no estado i no tempo $t = 0$. Vamos considerar ainda que a cada $n_i(0)$ existe associada uma probabilidade de ocorrência ε_i . Podemos pensar então nos $n_i(0)$'s como sendo variáveis aleatórias distribuídas multinomialmente com probabilidades ε_i . Suponhamos que são feitas observações de N unidades experimentais, onde $N = \sum_{i=1}^k n_i(0)$, de modo que uma observação de uma particular unidade experimental consiste da sequência de estados por ela ocupados em cada um dos tempos $t = 0, 1, \dots, T$ denotada por (x_0, x_1, \dots, x_T) . Dessa forma, a distribuição marginal do conjunto $\{n_i(0)\}$ é:

$$\left[\frac{N!}{\prod_{i=1}^k n_i(0)!} \prod_{i=1}^k \varepsilon_i^{n_i(0)} \right] \quad (2.19)$$

Temos que cada observação corresponde a uma cadeia de Markov e que dado o estado inicial x_0 , existem k^T possíveis sequências. Denotamos $n_{ij}(t)$, o número de unidades experimentais no estado i no tempo $t - 1$ e em j no tempo t . Pode ser facilmente demonstrado que o conjunto dos $n_{ij}(t)$'s ($i, j \in \mathcal{S}, t = 1, \dots, T$) com $k^2 T$ componentes, forma um conjunto de estatísticas suficientes para as sequências observadas. Devemos notar que

$$n_{.j}(t) = \sum_{i=1}^k n_{ij}(t) = n_j(t) \quad \text{e} \quad n_{i.}(t) = \sum_{j=1}^k n_{ij}(t) = n_i(t-1).$$

Sendo $p_{ij}(t-1, t)$ a probabilidade condicional de estar no estado j no tempo t , dado que estava em i no tempo $t - 1$, conforme (2.2) da seção 2.2, temos então, que para um valor fixo de i , a distribuição condicional de $n_{ij}(t)$ para $j = 1, \dots, k$, dado $n_{i.} = n_i(t-1)$ é

$$\frac{n_i(t-1)!}{\prod_{j=1}^k n_{ij}(t)!} \prod_{j=1}^k p_{ij}(t-1, t)^{n_{ij}(t)}, \quad (2.20)$$

que corresponde a $n_i(t-1)$ observações de uma distribuição multinomial com probabilidades $p_{ij}(t-1, t)$ e com frequências $n_{ij}(t)$. Se todas as transições são mutuamente independentes, então a distribuição de probabilidade conjunta de $n_i(0)$ e $n_{ij}(t)$ é

$$\left[\frac{N!}{\prod_{i=1}^k n_i(0)!} \prod_{i=1}^k \varepsilon_i^{n_i(0)} \right] \times \prod_{t=1}^T \left\{ \prod_{i=1}^k \left[\frac{n_i(t-1)!}{\prod_{j=1}^k n_{ij}(t)!} \prod_{j=1}^k p_{ij}(t-1, t)^{n_{ij}(t)} \right] \right\}. \quad (2.21)$$

A função de verossimilhança de $\{\varepsilon_i\}$ e de $\{p_{ij}(t-1, t)\}$ é portanto,

$$\mathcal{L} = \left[\prod_{i=1}^k \varepsilon_i^{n_i(0)} \right] \times \left[\prod_{t=1}^T \prod_{i=1}^k \prod_{j=1}^k p_{ij}(t-1, t)^{n_{ij}(t)} \right], \quad (2.22)$$

um produto de duas funções, uma envolvendo somente ε_i e a outra somente $p_{ij}(t-1, t)$. As estimativas de máxima verossimilhança para os parâmetros são:

$$\begin{aligned} \hat{\varepsilon}_i &= \frac{n_i(0)}{N}, \\ e \\ \hat{p}_{ij}(t-1, t) &= \frac{n_{ij}(t)}{n_i(t-1)} = \frac{n_{ij}(t)}{n_{i.}(t)}. \end{aligned} \quad (2.23)$$

Formalmente essas estimativas são as mesmas que seriam obtidas se para cada i e t tivéssemos $n_i(t-1)$ observações provenientes de uma distribuição multinomial com probabilidades $p_{ij}(t-1, t)$ e frequências $n_{ij}(t)$.

Para uma cadeia de Markov com probabilidades de transição estacionárias, um resultado mais forte sobre suficiência é que o conjunto $n_{ij} = \sum_{t=1}^T n_{ij}(t)$ forma um conjunto de estatísticas suficientes. Para probabilidades de transição estacionárias $p_{ij}(t)$, o conjunto dos $n_{ij}(t)$ forma um conjunto minimal de estatísticas suficientes. Nesse caso, $p_{ij}(t-1, t) = p_{ij}$ para $t = 1, \dots, T$, e o segundo fator da função de verossimilhança reduz-se a

$$\prod_{i=1}^k \prod_{j=1}^k p_{ij}^{n_{ij}}, \quad (2.24)$$

e as estimativas das probabilidades de transição estacionárias são:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i.}} = \frac{\sum_{t=1}^T n_{ij}(t)}{\sum_{t=1}^T n_{i.}(t-1)}. \quad (2.25)$$

As estimativas de máxima verossimilhança para as probabilidades iniciais permanecem as mesmas dadas em (2.23).

Podemos descrever essas estimativas da seguinte maneira: Na estrutura mais geral, quando as probabilidades de transição são não estacionárias, as transições observadas podem ser representadas em termos do conjunto de estatísticas suficientes $\{n_{ij}(t)\}$, e podem ser tratadas como se fossem as frequências observadas $x_{ijt} = n_{ij}(t)$ em uma tabela de contingência $k \times k \times T$, onde o total para cada nível da terceira variável é igual ao tamanho N da amostra. Quando as probabilidades de transição são estacionárias, podemos reduzir a tabela $k \times k \times T$ para uma tabela $k \times k$ com frequências $n_{ij} = x_{ij} = \sum_{t=1}^T x_{ijt}$.

Para facilitar a compreensão dessa idéia, vamos mostrar um exemplo extraído de Bishop *et al.* (1980). O experimento consistiu em fazer com que estudantes fechassem os olhos, e após um sinal escrevessem em uma folha de papel um dos números 1, 2 ou 3, o que viesse à sua cabeça. O procedimento foi repetido 7 vezes, dessa forma cada estudante escreveu uma sequência de 7 números. Para esse exemplo $k = 3$ e $t = 0, \dots, 6$. A tabela 2.1 mostra a tabela de contingência 3×3 considerando a suposição de estacionaridade, enquanto que a tabela 2.2 mostra a tabela de contingência $3 \times 3 \times 6$ para o mesmo exemplo, considerando o caso não estacionário.

Tabela 2.1: Número de transições no experimento de geração de números aleatórios (Probabilidades de transição estimadas para o caso estacionário)

| i -ésimo | $(i + 1)$ -ésimo | | | Total |
|------------|------------------|----------------|----------------|-------|
| | 1 | 2 | 3 | |
| 1 | 111 (0.273) | 131 (0.322) | 165 (0.405) | 407 |
| 2 | 165 (0.453) | 80 (0.220) | 119 (0.327) | 364 |
| 3 | 143 (0.407) | 140 (0.399) | 68 (0.194) | 351 |
| Total | 419 (0.374) | 350 (0.312) | 352 (0.314) | 1122 |

Podemos notar que na tabela 2.1 as probabilidades de transição correspondendo à diagonal principal da matriz de probabilidades de transição, são todas menores que 0.333, o que significa que números iguais aparecem menos frequentemente em ocasiões sucessivas do que o esperado. Já na tabela 2.2, pode-se observar que existe uma considerável variação entre

as matrizes de probabilidades de transição para cada intervalo de tempo. Isso pode ser um indicativo para a rejeição da hipótese de estacionaridade nesse experimento.

Tabela 2.2: Número de transições no experimento de geração de números aleatórios (Probabilidades de transição estimadas)

| a. 1º | 2º | | | Total | b. 2º | 3º | | | Total |
|----------|---------------|---------------|---------------|-------|----------|---------------|---------------|---------------|-------|
| | 1 | 2 | 3 | | | 1 | 2 | 3 | |
| 1 | 10 (0.167) | 16 (0.267) | 34 (0.567) | 60 | 1 | 23 (0.284) | 18 (0.222) | 40 (0.494) | 81 |
| 2 | 42 (0.575) | 10 (0.137) | 21 (0.288) | 73 | 2 | 16 (0.381) | 11 (0.262) | 15 (0.357) | 42 |
| 3 | 29 (0.537) | 16 (0.296) | 9 (0.167) | 54 | 3 | 22 (0.344) | 26 (0.406) | 16 (0.250) | 64 |
| Total | 81 (0.433) | 42 (0.225) | 64 (0.342) | 187 | Total | 61 (0.326) | 55 (0.294) | 71 (0.380) | 187 |
| c. 3º | 4º | | | Total | d. 4º | 5º | | | Total |
| | 1 | 2 | 3 | | | 1 | 2 | 3 | |
| 1 | 11 (0.180) | 33 (0.541) | 17 (0.279) | 61 | 1 | 20 (0.299) | 21 (0.313) | 26 (0.388) | 67 |
| 2 | 22 (0.400) | 10 (0.182) | 23 (0.418) | 55 | 2 | 32 (0.457) | 18 (0.257) | 20 (0.287) | 70 |
| 3 | 34 (0.479) | 27 (0.380) | 10 (0.141) | 71 | 3 | 24 (0.480) | 17 (0.340) | 9 (0.180) | 50 |
| Total | 67 (0.358) | 70 (0.374) | 50 (0.267) | 187 | Total | 76 (0.406) | 56 (0.299) | 55 (0.294) | 187 |
| e. 5º | 6º | | | Total | f. 6º | 7º | | | Total |
| | 1 | 2 | 3 | | | 1 | 2 | 3 | |
| 1 | 25 (0.329) | 26 (0.342) | 25 (0.329) | 76 | 1 | 22 (0.355) | 17 (0.274) | 23 (0.371) | 62 |
| 2 | 24 (0.429) | 13 (0.232) | 19 (0.339) | 56 | 2 | 29 (0.426) | 18 (0.265) | 21 (0.309) | 68 |
| 3 | 13 (0.236) | 29 (0.527) | 13 (0.236) | 55 | 3 | 21 (0.368) | 25 (0.439) | 11 (0.193) | 57 |
| Total | 62 (0.332) | 68 (0.364) | 57 (0.305) | 187 | Total | 72 (0.385) | 60 (0.321) | 55 (0.294) | 187 |

Propriedades Assintóticas dos Estimadores

Em função desses estimadores não terem uma distribuição exata, é necessário conhecer suas propriedades assintóticas com a finalidade de obter a distribuição aproximada para amostras suficientemente grandes, isto é, sua distribuição assintótica.

Consideramos $\{X(t); t \in \mathcal{T}\}$, onde $\mathcal{T} = \{0, \dots, T\}$, uma cadeia de Markov estacionária com espaço de estados $\mathcal{S} = \{1, 2, \dots, k\}$, probabilidades de transição p_{ij} , probabilidades iniciais $\varepsilon_i = \Pr\{X(0) = i\}$, $i \in \mathcal{S}$ e distribuição de probabilidade estacionária $\{\pi_j; j \in \mathcal{S}\}$. Consistência e normalidade assintóticas dos estimadores de máxima verossimilhança podem ser deduzidas usando a analogia com a distribuição multinomial conforme demonstrado em Anderson & Goodman (1957) e descrito na seção 2.4.2. Vamos assumir por simplicidade que o processo ocorre por um período muito longo de tempo, de modo que ele seja estacionário (conforme seção 2.2). Então podemos assumir que $\varepsilon_i = \pi_i$. Consequentemente $\Pr\{X(t) = i\} = \pi_i$ para qualquer t . Deve ser notado que a suposição $\varepsilon_i = \pi_i$ contém informação sobre as probabilidades de transição p_{ij} uma vez que $\{\pi_j; j \in \mathcal{S}\}$ satisfaz a equação (2.6).

Temos então um importante resultado relacionado à distribuição assintótica de \hat{p}_{ij} . Seja

$$\boldsymbol{\psi} = (p_{11}, p_{12}, \dots, p_{1k}, \dots, p_{k1}, \dots, p_{kk})'$$

o vetor $k^2 \times 1$ de parâmetros e

$$\hat{\boldsymbol{\psi}} = (\hat{p}_{11}, \hat{p}_{12}, \dots, \hat{p}_{1k}, \dots, \hat{p}_{k1}, \dots, \hat{p}_{kk})'$$

o vetor $k^2 \times 1$ de estimativas dos parâmetros, onde $\hat{p}_{ij} = n_{ij}/n_i$ conforme equação (2.25). Então

$$\sqrt{N}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \xrightarrow{\mathcal{L}} \mathcal{N}_{k^2}(\mathbf{0}, \boldsymbol{\Sigma}) \quad (2.26)$$

onde os componentes da matriz $\boldsymbol{\Sigma}$ são dados por

$$\sigma_{i_j, i'_j} = \frac{1}{\pi_i} [\delta_{ii'} (\delta_{jj'} p_{ij} - p_{ij} p_{i'j'})]$$

onde

$$\delta_{uv} = \begin{cases} 0 & u \neq v \\ 1 & u = v. \end{cases}$$

Para demonstração desse resultado, ver Basawa & Prakasa Rao (1980) - Cap. 2.

Vamos considerar agora o caso onde p_{ij} são funções conhecidas de um certo vetor de parâmetros desconhecidos, $\theta = (\theta_1, \dots, \theta_b)$ de modo que $p_{ij}(\theta)$ assume valores em \mathfrak{R}^b , que denominamos caso paramétrico. O estimador de máxima verossimilhança $\hat{\theta}$ de θ pode ser obtido através da solução do sistema de equações

$$\frac{\partial \log \mathcal{L}(\theta)}{\partial \theta_u} = 0, \quad u = 1, \dots, b, \quad (2.27)$$

onde

$$\log \mathcal{L}(\theta) = \sum_{i,j=1}^k n_{ij} \log p_{ij}. \quad (2.28)$$

Derivando (2.28) em relação a θ_u temos que (2.27) pode ser escrito como

$$\sum_{i,j=1}^k \left[n_{ij} \times \frac{\partial p_{ij}(\theta)}{\partial \theta_u} \times \frac{1}{p_{ij}(\theta)} \right] = 0 \quad (2.29)$$

simultaneamente para $\theta_1, \theta_2, \dots, \theta_b$. Seja $\hat{\theta}$ uma raiz de (2.29). A distribuição assintótica do estimador de máxima verossimilhança no caso paramétrico pode ser especificada através do seguinte teorema:

TEOREMA 1 : Suponha que as seguintes condições de regularidade são satisfeitas:

- (i) $p_{ij}(\theta) = p_{ij}(\zeta)$ para todo i e j , implica que $\theta = \zeta$,
- (ii) $p_{ij}(\theta)$ admite derivadas segundas contínuas com relação a θ para o verdadeiro valor do parâmetro θ_0 ,
- (iii) pelo menos um $(\partial p_{ij}(\theta)/\partial \theta)|_{\theta=\theta_0}$ é diferente de zero.

Temos então que

$$\sqrt{N} (\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}_b(\mathbf{0}, \mathbf{F}^{-1}) \quad (2.30)$$

onde \mathbf{F} é a matriz quadrada de ordem b com elementos

$$\sigma_{uv} = -E_{\theta} \left\{ \frac{\partial \log p_{x_0 x_1}(\theta)}{\partial \theta_u} \times \frac{\partial \log p_{x_0 x_1}(\theta)}{\partial \theta_v} \right\}$$

onde E_{θ} denota a esperança matemática em função de θ , denominada a matriz de informação de Fisher assumida ser não singular.

2.4.3 Testes de Hipóteses

Os resultados da seção anterior podem ser utilizados para derivar testes de significância para várias hipóteses relacionadas à cadeia de Markov em questão. Vamos descrever alguns deles.

Teste para uma matriz de transição específica

Considere a hipótese $\mathcal{H}_0 : p_{ij} = p_{ij}^0, i, j \in \mathcal{S}$. Para qualquer i fixo, $\{p_{ij}^0; j \in \mathcal{S}\}$ é uma distribuição de probabilidades. Sob a hipótese nula, a estatística $\sum_{j=1}^k (n_{ij} - n_i p_{ij}^0)^2 / n_i p_{ij}^0$ é, de fato, a estatística qui-quadrado para adequabilidade do ajuste para a distribuição multinomial com valores n_{ij} e probabilidades $n_i p_{ij}^0$. Consequentemente, a estatística

$$\phi = \sum_{i,j=1}^k \frac{(n_{ij} - n_i p_{ij}^0)^2}{n_i p_{ij}^0} \quad (2.31)$$

é a soma de k estatísticas qui-quadrado independentes correspondentes às k linhas da matriz de transição. Usando a analogia com a distribuição multinomial, temos que ϕ tem, assintoticamente, distribuição χ^2 com $k(k-1)$ graus de liberdade, assumindo $p_{ij} > 0, \forall i, j \in \mathcal{S}$. No caso mais geral, consideraremos o conjunto \mathcal{D} de pares (i, j) tais que $p_{ij} > 0$. Então, sob a hipótese nula,

$$\phi = \sum_{\mathcal{D}} \frac{(n_{ij} - n_i p_{ij}^0)^2}{n_i p_{ij}^0} \quad (2.32)$$

tem assintoticamente distribuição χ^2 com $d - k$ graus de liberdade.

Utilizando-se o critério de Neyman-Pearson, temos como estatística da razão de verossimilhança para o teste,

$$\lambda = -2 \sum_{\mathcal{D}} n_{ij} \log \left(\frac{n_i p_{ij}^0}{n_{ij}} \right) \quad (2.33)$$

que, sob \mathcal{H}_0 tem assintoticamente distribuição χ^2 com $d - k$ graus de liberdade.

As estatísticas ϕ e λ são critérios bastante razoáveis para se testar a hipótese sobre probabilidades de transição específicas. O teste rejeita \mathcal{H}_0 sempre que o valor de ϕ ou λ é maior que o valor de χ^2 tabelado para o nível de significância α especificado. Pode ser demonstrado que as estatísticas ϕ e λ são assintoticamente equivalentes.

Teste para independência

A hipótese de independência das variáveis aleatórias $\{X(t); t \in \mathcal{T}\}$ baseada na suposição de dependência de Markov e estacionaridade, assume a seguinte forma:

$$\mathcal{H}_0 : p_{ij} = \pi_j = \Pr\{X(t) = j\} \quad i, j \in \mathcal{S},$$

onde as probabilidades estacionárias π_j são desconhecidas. Sob essa hipótese, temos apenas $k-1$ parâmetros que são $(\pi_1, \pi_2, \dots, \pi_k)$ sujeitos à restrição $\sum_{j=1}^k \pi_j = 1$.

O estimador de máxima verossimilhança de π_j , sob \mathcal{H}_0 é $n_{.j}/n$, onde $n_{.j} = \sum_{i=1}^k n_{ij}$. O estimador de máxima verossimilhança de p_{ij} é n_{ij}/n_i (ignorando o estado inicial). Os critérios qui-quadrado e razão de verossimilhança para testar \mathcal{H}_0 são respectivamente:

$$\phi = \sum_{i,j=1}^k \frac{(n_{ij} - n_i \cdot n_{.j}/N)^2}{\frac{n_i \cdot n_{.j}}{N}} \quad (2.34)$$

e

$$\lambda = -2 \sum_{i,j=1}^k n_{ij} \log \left(\frac{n_i \cdot n_{.j}/N}{n_{ij}} \right). \quad (2.35)$$

Novamente utilizando a analogia com a distribuição multinomial, temos que a estatística ϕ tem distribuição assintótica χ^2 com $(k^2 - k) - (k - 1) = (k - 1)^2$ graus de liberdade, e valores grandes da estatística ϕ correspondem à rejeição da hipótese nula, o mesmo ocorrendo com λ . Ambas as estatísticas podem ser utilizadas como critério para o teste.

Teste para homogeneidade em várias amostras

Suponha que temos m amostras independentes provenientes de m cadeias de Markov com probabilidades de transição $p_{ij(h)}$, $i, j = 1, \dots, k$ e $h = 1, \dots, m$. Cada amostra tem sua frequência de transições $n_{ij(h)}$, $h = 1, \dots, m$. Então o estimador de máxima verossimilhança de $p_{ij(h)}$ obtido a partir da amostra correspondente é $n_{ij(h)}/n_{i(h)}$.

Sob $\mathcal{H}_0 : p_{ij(1)} = p_{ij(2)} = \dots = p_{ij(m)} = p_{ij}$; $i, j \in \mathcal{S}$, isto é, as m matrizes de transição são idênticas. O estimador de máxima verossimilhança da probabilidade de transição comum a

todas as amostras p_{ij} é

$$\hat{p}_{ij} = \frac{\sum_{h=1}^m n_{ij(h)}}{\sum_{h=1}^m n_{i \cdot (h)}}. \quad (2.36)$$

Suponha que as amostras têm tamanhos N_1, N_2, \dots, N_m , onde $\sum_{h=1}^m N_h = N$. Os critérios qui-quadrado e razão de verossimilhança para testar \mathcal{H}_0 são respectivamente:

$$\phi = \sum_{h=1}^m \sum_{i,j=1}^k \frac{(n_{ij(h)} - n_{i \cdot (h)} \hat{p}_{ij})^2}{n_{i \cdot (h)} \hat{p}_{ij}} \quad (2.37)$$

e

$$\lambda = -2 \sum_{h=1}^m \sum_{i,j=1}^k n_{ij(h)} \log \left(\frac{n_{i \cdot (h)} \hat{p}_{ij}}{n_{ij(h)}} \right). \quad (2.38)$$

Essas estatísticas, sob \mathcal{H}_0 , têm assintoticamente distribuição χ^2 com $mk(k-1) - k(k-1) = (m-1)k(k-1)$ graus de liberdade.

Teste para a ordem da cadeia

Uma cadeia de Markov foi definida na seção 2.2 considerando tempos sucessivos, isto é, o estado no tempo t dependia apenas do estado no tempo $t-1$, porém isso nem sempre ocorre. Às vezes a probabilidade do sistema estar em um determinado estado no tempo t depende dos estados nos z tempos anteriores, ou seja, nos tempos $t-1, t-2, \dots, t-z$. O número z de estados prévios necessários para se determinar a probabilidade de transição é denominado a *ordem* da cadeia de Markov.

Uma cadeia de Markov de ordem z é definida pela seguinte propriedade:

$$\begin{aligned} & \Pr\{X(t) = a_{z+1} \mid X(t-1) = a_z, \dots, X(t-z) = a_1, \dots, X(1) = x_1, X(0) = x_0\} \\ &= \Pr\{X(t) = a_{z+1} \mid X(t-1) = a_z, \dots, X(t-z) = a_1\} \\ &= p_{a_1 \dots a_z a_{z+1}}, \end{aligned} \quad (2.39)$$

que significa que a probabilidade do estado atual, dado os estados passados, depende somente dos últimos z estados.

Quando $z = 1$, temos o caso usual de dependência de primeira ordem, isto é, uma cadeia de Markov de 1ª ordem. Qualquer cadeia de Markov de z -ésima ordem, contudo, pode

ser transformada em uma cadeia de Markov de 1ª ordem equivalente. Vamos considerar primeiramente o caso de uma cadeia de Markov de 2ª ordem, de modo a tornar mais clara a notação quando no caso de z -ésima ordem. Considere o par de estados sucessivos i e j definindo um estado composto (i, j) . Então a probabilidade do estado composto (j, l) dado o estado composto (i, j) é p_{ijl} . Obviamente, a probabilidade de (h, l) , $h \neq j$, dado (i, j) é igual a 0. Os estados compostos são facilmente vistos como formando uma cadeia de Markov de primeira ordem com k^2 estados e com algumas probabilidades de transição nulas.

Para o caso geral, seja $\{Y(t)\}$, o processo com espaço de estados z -dimensional S^z com elementos $y(t) = (a_t, a_{t+1}, \dots, a_{t+z-1})$. Então

$$\begin{aligned} & \Pr\{Y(t) = (a_1, \dots, a_z) \mid Y(t-1) = (b_1, \dots, b_z)\} \\ &= \Pr\{X(t) = a_1, \dots, X(t+z-1) = a_z \mid X(t-1) = b_1, \dots, X(t+z-2) = b_z\} \\ &= \Pr\{X(t+z-1) = a_z \mid X(t-1) = b_1, X(t) = b_2 = a_1, \dots, X(t+z-2) = b_z = a_{z-1}\} \\ &= \begin{cases} p_{b_1 b_2 \dots b_z; a_z} & \text{se } a_i = b_{i+1}, i = 1, \dots, z-1 \\ 0 & \text{caso contrário.} \end{cases} \end{aligned} \tag{2.40}$$

Dessa forma, $\{Y(t)\}$ forma uma cadeia de Markov de 1ª ordem com probabilidades de transição dadas por

$$p_{(b_1, \dots, b_z)(a_1, \dots, a_z)} = \begin{cases} p_{b_1 \dots b_z; a_z} & \text{se } a_i = b_{i+1}, i = 1, \dots, z-1 \\ 0 & \text{caso contrário.} \end{cases} \tag{2.41}$$

O número de elementos positivos em cada linha da matriz de transição P para $\{Y(t)\}$ é k e o número total de elementos positivos é k^{z+1} , uma vez que existem k^z linhas em P . Denotamos a frequência da cadeia a_1, \dots, a_z como $n_{a_1 \dots a_z}$ e portanto, $n_{a_1 \dots a_{z+1}}$ corresponde à frequência de transições de (a_1, \dots, a_z) para (a_2, \dots, a_{z+1}) . Logo, $n_{a_1 \dots a_z}$ e $n_{a_1 \dots a_{z+1}}$ correspondem a n_i e n_{ij} na cadeia de 1ª ordem. Suponha que queremos testar a seguinte hipótese:

- \mathcal{H}_0 : o processo é de w -ésima ordem,
- \mathcal{H}_1 : o processo é de z -ésima ordem ($z > w$).

A hipótese nula pode ser representada como $p_{a_1 \dots a_z; a_{z+1}} = p_{a_{z-w+1} \dots a_z; a_{z+1}}$, isto é, as probabilidades de transição são especificadas por k^{w+1} parâmetros $\{p_{a_{z-w+1} \dots a_z; a_{z+1}}\}$. Esses parâmetros

contudo, são restritos por k^w somas de linhas que são iguais a unidade, de modo que temos $k^{w+1} - k^w = k^w(k - 1)$ parâmetros independentes.

Sob \mathcal{H}_0 , o estimador de máxima verossimilhança é dado por

$$\hat{p}_{a_1 \cdots a_z; a_{z+1}} = \frac{n_{a_z - w + 1 \cdots a_{z+1}}}{n_{a_z - w + 1 \cdots a_z}}. \quad (2.42)$$

A estatística qui-quadrado apropriada para testar \mathcal{H}_0 é dada então por

$$\phi = \sum_{a_1, \dots, a_{z+1}} \frac{(n_{a_1 \cdots a_{z+1}} - n_{a_1 \cdots a_{z+1}} \hat{p}_{a_1 \cdots a_z; a_{z+1}})^2}{n_{a_1 \cdots a_{z+1}} \hat{p}_{a_1 \cdots a_z; a_{z+1}}} \quad (2.43)$$

e para o critério da razão de verossimilhança, a estatística apropriada é:

$$\lambda = -2 \sum_{a_1, \dots, a_{z+1}} n_{a_1 \cdots a_{z+1}} \log \left(\frac{n_{a_1 \cdots a_{z+1}} \hat{p}_{a_1 \cdots a_z; a_{z+1}}}{n_{a_1 \cdots a_{z+1}}} \right). \quad (2.44)$$

Essas estatísticas têm, sob \mathcal{H}_0 , assintoticamente distribuição χ^2 com $k^z(k - 1) - k^w(k - 1) = k^w(k - 1)(k^{z-w} - 1)$ graus de liberdade.

Teste de hipóteses sobre os parâmetros

Suponha que queremos testar a seguinte hipótese nula:

$$\mathcal{H}_0 : p_{ij} = p_{ij}(\theta)$$

onde $\theta = (\theta_1, \dots, \theta_b)$ é um vetor de parâmetros desconhecidos $\in \mathbb{R}^b$. As estatísticas apropriadas para essa situação são:

$$\phi = \sum_{i,j=1}^k \frac{(n_{ij} - n_i p_{ij}(\hat{\theta}))^2}{n_i p_{ij}(\hat{\theta})} \quad (2.45)$$

e

$$\lambda = -2 \sum_{i,j=1}^k n_{ij} \log \left(\frac{n_i p_{ij}(\hat{\theta})}{n_{ij}} \right), \quad (2.46)$$

onde $\hat{\theta}$ é o estimador de máxima verossimilhança de θ obtido resolvendo-se (2.29).

Assumindo que a matriz de dimensão $k^2 \times b$ com elementos $(\partial p_{ij}(\theta) / \partial \theta_u)$, tem "rank" b e que $p_{ij}(\theta)$ têm derivadas parciais contínuas de primeira e segunda ordens com relação a θ_u ,

$u = 1, \dots, b$, pode ser feita a analogia com a distribuição multinomial e ver que a distribuição assintótica das estatísticas em questão, sob a hipótese nula, é χ^2 com $k(k-1) - b$ graus de liberdade.

Isso ocorre quando todos os p_{ij} 's são positivos. Se houverem apenas d elementos $p_{ij} > 0$, os graus de liberdade tornam-se $d - k - b$. Utilizando-se o critério de Neyman-Pearson, a estatística da razão de verossimilhança para o teste é:

$$\lambda = -2 \sum_D n_{ij} \log \left(\frac{n_i p_{ij}(\hat{\theta})}{n_{ij}} \right). \quad (2.47)$$

Sob a hipótese nula, a estatística tem distribuição assintótica χ^2 com $d - k - r$ graus de liberdade.

Teste de hipótese de estacionaridade

Até o momento, sempre foi feita a suposição de que as probabilidades de transição são independentes do tempo, o que nem sempre é verdade. Uma alternativa para essa suposição é que elas dependem do tempo. Denotamos as probabilidades de transição dependentes do tempo como $p_{ij}(t)$. A hipótese que se quer testar é $\mathcal{H}_0 : p_{ij}(t) = p_{ij}, t = 1, \dots, T$. Esse teste é muito semelhante àquele para homogeneidade em várias amostras, descrito anteriormente. O número de transições do estado i para o estado j , entre os tempos $t-1$ e t é denotado por $n_{ij}(t)$ e o número de indivíduos no estado i no tempo $t-1$ é denotado por $n_i(t-1) = \sum_{j=1}^k n_{ij}(t)$. Sob a hipótese alternativa, as estimativas das probabilidades de transição no tempo t são

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{n_i(t-1)}.$$

Para um dado i , o conjunto $\hat{p}_{ij}(t)$ tem a mesma distribuição assintótica que as estimativas das probabilidades multinomiais $p_{ij}(t)$ para T amostras independentes. Uma tabela $k \times T$, que tem a mesma aparência formal de uma tabela de contingência, pode ser usada para representar as estimativas conjuntas $\hat{p}_{ij}(t)$ para um i dado, $j = 1, \dots, k$ e $t = 1, \dots, T$, conforme a tabela 2.3 a seguir:

Tabela 2.3: Representação tabular para as estimativas das probabilidades de transição

| t | j | | | |
|---|-------------------|-------------------|-----|-------------------|
| | 1 | 2 | ... | k |
| 1 | $\hat{p}_{i1}(1)$ | $\hat{p}_{i2}(1)$ | ... | $\hat{p}_{ik}(1)$ |
| 2 | $\hat{p}_{i1}(2)$ | $\hat{p}_{i2}(2)$ | ... | $\hat{p}_{ik}(2)$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| T | $\hat{p}_{i1}(T)$ | $\hat{p}_{i2}(T)$ | ... | $\hat{p}_{ik}(T)$ |

A hipótese de interesse é que as variáveis aleatórias, representadas pelas T linhas, têm a mesma distribuição, de modo que os dados sejam homogêneos nesse aspecto. Isso é equivalente a hipótese que existam k constantes $p_{i1}, p_{i2}, \dots, p_{ik}$ com $\sum_{j=1}^k p_{ij} = 1$ tais que a probabilidade associada com a j -ésima coluna seja igual a p_{ij} em todas as T linhas, isto é, queremos testar a seguinte hipótese nula:

$$\mathcal{H}_0 : p_{ij}(t) = p_{ij}$$

para $t = 1, \dots, T$.

A estatística qui-quadrado adequada para esse teste é:

$$\phi_i = \sum_{t=1}^T \sum_{j=1}^k \frac{n_i(t-1) [\hat{p}_{ij}(t) - p_{ij}]^2}{\hat{p}_{ij}} \quad (2.48)$$

Sob a hipótese nula, ϕ_i tem distribuição assintótica χ^2 com $(k-1)(T-1)$ graus de liberdade.

Utilizando o critério de Neyman-Pearson, temos como estatística da razão de verossimilhança

$$\lambda_i = -2 \sum_{t=1}^T \sum_{j=1}^k n_{ij}(t) \log \left(\frac{\hat{p}_{ij}}{\hat{p}_{ij}(t)} \right), \quad (2.49)$$

que tem distribuição assintótica χ^2 com $(k-1)(T-1)$ graus de liberdade. O teste acima lida com um valor fixo de i e portanto a hipótese pode ser testada separadamente para cada valor de i .

Considerando a hipótese geral $\mathcal{H}_0 : p_{ij}(t) = p_{ij}$ para todo $i = 1, \dots, k; j = 1, \dots, k$ e $t = 1, \dots, T$, temos que um teste dessa hipótese segue diretamente do fato que as variáveis aleatórias $\hat{p}_{ij}(t)$ e \hat{p}_{ij} para dois valores distintos de i , são assintoticamente independentes. Portanto, sob

λ_0 , os valores de ϕ_i calculados para cada $i = 1, 2, \dots, k$ são assintoticamente independentes e o somatório

$$\phi = \sum_{i=1}^k \phi_i = \sum_{t=1}^T \sum_{i,j=1}^k \frac{n_i(t-1)[\hat{p}_{ij}(t) - \hat{p}_{ij}]^2}{\hat{p}_{ij}} \quad (250)$$

tem distribuição assintótica χ^2 com $k(k-1)(T-1)$ graus de liberdade.

Da mesma forma,

$$\lambda = \sum_{i=1}^k \lambda_i = -2 \sum_{t=1}^T \sum_{i,j=1}^k n_{ij}(t) \log \left(\frac{\hat{p}_{ij}}{\hat{p}_{ij}(t)} \right) \quad (251)$$

tem distribuição assintótica χ^2 com $k(k-1)(T-1)$ graus de liberdade.

Capítulo 3

Análise Utilizando Modelos para Processos em Tempo Discreto

3.1 Introdução

De uma forma geral, quando se tem dados obtidos longitudinalmente e o interesse está concentrado em estudar as mudanças individuais (transições) no tempo, estamos lidando com modelos de transição. Neste capítulo estudamos esses modelos quando o tempo é discreto e as variáveis resposta são categóricas ordinais. O que determina o fato do tempo ser discreto é a forma como a resposta foi obtida, por exemplo, um questionário aplicado mensalmente ou um exame realizado em pacientes cujo resultado é obtido trimestralmente e assim por diante. É importante notar que para se ter respostas em tempo discreto, elas só podem ser obtidas na ocasião especificada, ou seja, definem-se os tempos de observação de modo que fora deles não existe resposta.

Para o estudo de transições, o enfoque mais utilizado é baseado na suposição de Markov, conforme descrito na seção 2.2 do capítulo 2. Se a amostra utilizada no estudo é assumida como sendo homogênea, isto é, se temos um único grupo com as mesmas probabilidades de transição, as estimativas dadas na seção 2.4.2 podem ser utilizadas diretamente. Todavia, em casos práticos, nem sempre a amostra pode ser considerada homogênea. Se variáveis externas

ou exógenas que influenciam na resposta são introduzidas na análise na forma de tratamentos, blocos ou covariáveis, caracterizando a formação de diferentes grupos de acordo com os valores assumidos por essas variáveis, podemos utilizá-las considerando cada grupo em separado, assumindo homogeneidade dentro de cada grupo. As probabilidades de transição para cada um dos grupos são estimadas conforme descrito na seção 2.4.2.

Vamos supor por exemplo, que foi realizado um experimento para verificar a influência da poluição atmosférica na incidência do câncer no pulmão em uma determinada cidade. A amostra consistiu de indivíduos selecionados aleatoriamente de vários bairros dessa cidade e a variável de interesse foi dividida em categorias de acordo com o estágio cancerígeno do indivíduo. Essa divisão foi determinada pela equipe responsável pela pesquisa. Definiu-se uma medida representando a distância entre o bairro e a indústria mais próxima, que foi considerada nesse caso uma covariável. Claramente a amostra como um todo não pode ser considerada um grupo homogêneo, porém pode-se assumir a homogeneidade dentro de cada bairro. Dessa forma, conforme exposto anteriormente, podemos obter estimativas para as probabilidades de transição dentro de cada grupo (bairro) e através de testes específicos, tirar as conclusões desejadas.

Entretanto, à medida que aumenta o número de variáveis exógenas envolvidas na análise, ocorre uma diminuição de observações dentro dos grupos e conseqüentemente para alguns deles, o tamanho de amostra pode ser insuficiente para obtenção de estimativas consistentes dos parâmetros. No caso do exemplo descrito, se além da distância acrescentarmos uma outra covariável representando o hábito de fumar do indivíduo, cada bairro será subdividido de acordo com o valor assumido por essa covariável e assim teríamos amostras cada vez menores para cada variável exógena inserida na análise.

Neste capítulo abordamos o enfoque no qual o interesse é estudar o efeito das variáveis exógenas nas probabilidades de transição através de modelos que permitem descrever as probabilidades de transição, conforme os valores assumidos pelas variáveis exógenas. Esses modelos têm a vantagem de geralmente trabalhar com um número menor de parâmetros. Sob esse enfoque, para a análise do exemplo anterior, em vez de estimar todas as probabilidades de transição

para cada grupo, fazemos uma modelagem das probabilidades de transição baseada nas variáveis exógenas. Dessa forma, o que se está fazendo é estimando um parâmetro para cada variável exógena, que pode ser interpretado como o peso que a variável tem na estimação da probabilidade de transição.

Para tal propósito, utilizamos o método descrito em Ware, Lipsitz & Speizer (1988), que é uma extensão do método desenvolvido por Stram, Wei & Ware (1988) para modelos marginais com respostas ordinais.

Este capítulo está dividido em três seções. A seção 3.2 é dedicada à descrição, em uma forma resumida, do método de Stram, Wei & Ware para modelos marginais com respostas ordinais categorizadas. Na seção 3.3 descrevemos o método adaptado para modelos de transição desenvolvido em Ware, Lipsitz & Speizer (1988). Na última seção vamos através de um exemplo extraído da literatura, dar uma idéia de como esses métodos podem ser utilizados na prática.

3.2 Método de Stram, Wei & Ware para modelos marginais com respostas ordinais

Vamos assumir que as respostas são do tipo ordinal com valores entre 1 e k onde $1 < 2 < \dots < k$, ou seja, o nosso espaço de estados é $S = \{1, \dots, k\}$. O objetivo é modelar as probabilidades marginais para esse tipo de resposta levando-se em conta a existência de variáveis exógenas influenciando na resposta. Os autores assumem que as probabilidades acumuladas da resposta em cada ocasião seguem o modelo de “odds” proporcionais, que é um membro da família dos modelos lineares generalizados descrita por McCullagh (1980). Os parâmetros desse modelo são assumidos específicos para cada ocasião e são estimados maximizando-se as verossimilhanças de cada uma das ocasiões em separado.

Para a t -ésima ocasião com $t = 0, 1, \dots, T$, a resposta para o v -ésimo indivíduo é representada pelo vetor $\mathbf{Z}_{tv} = (Z_{1tv}, Z_{2tv}, \dots, Z_{ktv})'$, onde Z_{jtv} , ($j = 1, \dots, k$) denotam variáveis indicadoras para a categoria da variável resposta, isto é, $Z_{jtv} = 1$ se o indivíduo v está na j -

ésima categoria no tempo t e $Z_{jtv} = 0$ caso contrário. As variáveis exógenas de interesse são representadas pelo vetor $\mathbf{W}_{tv} = (W_{1tv}, W_{2tv}, \dots, W_{ptv})'$ onde p é o número de variáveis exógenas na análise, que por sua vez, podem variar no tempo. A probabilidade marginal da ocorrência de Z_{jtv} é denotada por $\xi_{jt}(\mathbf{w})$, onde $j = 1, \dots, k$ e as probabilidades acumuladas por $\gamma_{1t}(\mathbf{w}) = \xi_{1t}(\mathbf{w})$, $\gamma_{2t}(\mathbf{w}) = \xi_{1t}(\mathbf{w}) + \xi_{2t}(\mathbf{w})$, \dots , $\gamma_{k-1,t}(\mathbf{w}) = \xi_{1t}(\mathbf{w}) + \dots + \xi_{k-1,t}(\mathbf{w})$ e $\gamma_{kt}(\mathbf{w}) = 1$, onde $\mathbf{w} = (w_1, w_2, \dots, w_p)'$ é o vetor de valores assumidos pelos componentes do vetor \mathbf{W}_{tv} . O modelo de “odds” proporcionais especifica que:

$$\gamma_{jt}(\mathbf{w}) = \phi(\lambda_{jt} - \beta_t' \mathbf{w}) \quad j = 1, \dots, k-1, \quad (3.1)$$

onde $\phi(y) = e^y / (1 + e^y)$, λ_{jt} é o parâmetro de perturbação do modelo e que denominamos “intercept” e $\beta_t = (\beta_{1t}, \dots, \beta_{pt})'$ é um vetor de parâmetros desconhecidos que representam o efeito das variáveis exógenas nas probabilidades $\xi_{jt}(\mathbf{w})$. Caso existam apenas duas categorias, o modelo reduz-se a uma regressão logística para a probabilidade da primeira resposta.

Em função dos parâmetros $\gamma_{jt}(\mathbf{w})$ serem obtidos acumulando do menor estado para o maior e os termos envolvendo as variáveis exógenas em (3.1) terem sinal negativo, um coeficiente positivo para uma variável exógena indica que a probabilidade de estar no estado maior aumenta à medida que o valor dessa variável aumenta.

Vamos definir uma variável que indica a presença de valores perdidos (“missing values”) nas observações, da seguinte forma:

$$\delta_{tv} = \begin{cases} 1 & \text{se } Z_{tv} \text{ e } \mathbf{W}_{tv} \text{ são observados} \\ 0 & \text{caso contrário.} \end{cases}$$

É assumido que dado \mathbf{W}_{tv} , Z_{tv} e δ_{tv} são independentes.

A proposta de Stram, Wei & Ware consiste em maximizar a função de verossimilhança para cada ocasião em separado, utilizando-se o modelo (3.1) para modelagem das probabilidades marginais. Os parâmetros do modelo são assumidos serem específicos para cada ocasião. O logaritmo da função de verossimilhança no tempo t é proporcional a

$$\mathcal{L}_t = \sum_{v=1}^N \delta_{tv} \sum_{j=1}^k Z_{jtv} \{ \log [\phi(\lambda_{jt} - \mathbf{W}_{tv}' \beta_t) - \phi(\lambda_{j-1,t} - \mathbf{W}_{tv}' \beta_t)] \} \quad t = 0, \dots, T, \quad (3.2)$$

onde $\phi(\lambda_{0t} - \mathbf{W}'_{t0}\beta_t) = 0$ e $\phi(\lambda_{kt} - \mathbf{W}'_{t0}\beta_t) = 1$. Desse modo, os parâmetros β_t e λ_{jt} são estimados através da maximização da equação (3.2). Devido à não existência de solução explícita para esse problema, deve-se utilizar um método iterativo, por exemplo o método de Newton-Raphson, para obtenção de $\hat{\beta}_t$ e $\hat{\lambda}_{jt}$. A partir dessas estimativas, pode-se obter as estimativas para as probabilidades marginais como segue:

$$\begin{aligned}\hat{\xi}_{1t}(\mathbf{w}) &= \hat{\gamma}_{1t}(\mathbf{w}) = \phi(\hat{\lambda}_{1t} - \hat{\beta}'_t \mathbf{w}), \\ \hat{\xi}_{2t}(\mathbf{w}) &= \hat{\gamma}_{2t}(\mathbf{w}) - \hat{\xi}_{1t}(\mathbf{w}) = \phi(\hat{\lambda}_{2t} - \hat{\beta}'_t \mathbf{w}) - \hat{\xi}_{1t}(\mathbf{w}), \\ &\vdots \\ \hat{\xi}_{k-1,t}(\mathbf{w}) &= \hat{\gamma}_{k-1,t}(\mathbf{w}) - \hat{\xi}_{1t}(\mathbf{w}) - \hat{\xi}_{2t}(\mathbf{w}) - \dots - \hat{\xi}_{k-2,t}(\mathbf{w}) \\ &= \phi(\hat{\lambda}_{k-1,t} - \hat{\beta}'_t \mathbf{w}) - \hat{\xi}_{1t}(\mathbf{w}) - \hat{\xi}_{2t}(\mathbf{w}) - \dots - \hat{\xi}_{k-2,t}(\mathbf{w}), \\ \hat{\xi}_{kt}(\mathbf{w}) &= 1 - \hat{\gamma}_{k-1,t}(\mathbf{w}) = 1 - \phi(\hat{\lambda}_{k-1,t} - \hat{\beta}'_t \mathbf{w}).\end{aligned}$$

No seu trabalho, os autores demonstram que $\hat{\beta}_{ct} = (\hat{\beta}_{c0}, \hat{\beta}_{c1}, \dots, \hat{\beta}_{cT})'$ tem assintoticamente uma distribuição normal multivariada com vetor de médias $(\beta_{c0}, \beta_{c1}, \dots, \beta_{cT})'$ e matriz de covariância Σ . Eles ainda descrevem como obter um estimador consistente $\hat{\Sigma}$ de Σ . Para o cálculo das variâncias assintóticas das estimativas das probabilidades marginais, utiliza-se o método δ multivariado (ver Apêndice B-1). Aos leitores interessados em uma descrição mais detalhada do método, recomenda-se o trabalho de Dall'Agnol (1990) - Cap. 3.

3.3 Adaptação do método de Stram, Wei & Ware para modelos de transição

Como já foi citado no capítulo 1, os modelos marginais não utilizam toda a informação contida nos dados obtidos longitudinalmente. Em particular, através deles não é possível modelar as mudanças individuais no tempo, nem os efeitos de possíveis variáveis externas influenciando nessas mudanças. Os modelos que lidam com mudanças individuais são os que denominamos modelos de transição.

Nesta seção descrevemos uma adaptação do método descrito na seção anterior, cujo objetivo é modelar as probabilidades de transição entre pares de estados, que representam as mudanças individuais no tempo.

Uma observação consiste na sequência de respostas individuais no tempo, por exemplo, supondo $S = \{1, 2, 3\}$ e $t = 0, 1, 2, 3$, a observação v consistiria na classe a qual o v -ésimo indivíduo pertence nos tempos 0, 1, 2 e 3 respectivamente. O conjunto das N observações individuais pode ser disposto, por exemplo, conforme mostrado na tabela 3.1 abaixo:

Tabela 3.1: Exemplo de um possível conjunto de N observações em 4 tempos e 3 categorias

| Indivíduo | Tempo | | | |
|-----------|-------|---|---|---|
| | 0 | 1 | 2 | 3 |
| 1 | 1 | 1 | 1 | 2 |
| 2 | 3 | 3 | 3 | 3 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| v | 3 | 3 | 2 | 2 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| N | 2 | 1 | 1 | 2 |

Dessa tabela pode-se observar, por exemplo, que a v -ésima observação indica que o indivíduo v estava inicialmente na categoria 3, na ocasião seguinte continuou na categoria 3, passou para a categoria 2 na terceira ocasião e manteve-se em 2 na quarta ocasião. A resposta para esse indivíduo no tempo 1 é representada por $Z_{1v} = (0, 0, 1)'$ e no tempo 2 por $Z_{2v} = (0, 1, 0)'$, conforme notação definida na seção anterior.

A suposição inicial é que cada observação forma uma cadeia de Markov de 1^a ordem e o interesse principal é encontrar estimativas consistentes para as probabilidades de transição em cada intervalo de tempo.

De acordo com o trabalho de Ware, Lipsitz & Speizer (1988), a suposição de Markov pode ser operacionalmente caracterizada no modelo, através da inclusão da categoria no tempo anterior como variável exógena correspondendo a uma variável independente em um modelo de

regressão. A partir dessa idéia, os autores desenvolveram um método para estimar as probabilidades de transição influenciadas por variáveis exógenas, quando o tempo é discreto e a resposta é ordinal, o qual discutimos nesta seção.

Utilizando a notação da seção anterior com $t = 0, 1, \dots, T$, podemos escrever as probabilidades de transição levando-se em consideração a existência de variáveis exógenas influenciando na resposta como:

$$p_{lm}(t-1, t) = \Pr\{Z_{mtv} \mid Z_{l,t-1,v}, \mathbf{W}_{tv}\} \quad l, m = 1, \dots, k, t = 1, \dots, T. \quad (3.3)$$

O método descrito na seção 3.2 pode ser utilizado para a modelagem de (3.3) para as T transições em separado, ou seja, de $t = 0$ para $t = 1$, de $t = 1$ para $t = 2$, e assim sucessivamente até de $t = T - 1$ para $t = T$. Para o modelo de transição, o vetor de variáveis exógenas é denotado por:

$$\mathbf{W}_{tv}^* = (W_{1tv}, W_{2tv}, \dots, W_{p+1,t,v})',$$

onde W_{1tv} é o valor assumido por $Z_{l,t-1,v}$, ou seja, o vetor \mathbf{W}_{tv}^* tem um elemento a mais que \mathbf{W}_{tv} definido na seção 3.2, elemento esse que corresponde ao indicador da categoria no estado anterior. A probabilidade de transição $p_{lm}(t-1, t) = \Pr\{Z_{mtv} \mid \mathbf{W}_{tv}^*\}$ é denotada por $\xi_{lmt}(\mathbf{w}^*)$, onde $l, m = 1, \dots, k$ e as probabilidades de transição acumuladas por:

$$\gamma_{11t}(\mathbf{w}^*) = \xi_{11t}(\mathbf{w}^*),$$

$$\gamma_{12t}(\mathbf{w}^*) = \xi_{11t}(\mathbf{w}^*) + \xi_{12t}(\mathbf{w}^*),$$

⋮

$$\gamma_{l,k-1,t}(\mathbf{w}^*) = \xi_{11t}(\mathbf{w}^*) + \xi_{12t}(\mathbf{w}^*) + \dots + \xi_{l,k-1,t}(\mathbf{w}^*)$$

$$\gamma_{kt}(\mathbf{w}^*) = 1,$$

onde $\mathbf{w}^* = (w_1^*, w_2^*, \dots, w_{p+1}^*)'$ é o vetor de valores assumidos pelos elementos do vetor \mathbf{W}_{tv}^* . Essas probabilidades de transição acumuladas, especificadas pelo modelo de "odds" proporcionais definido em (3.1), nesse caso são definidas como:

$$\gamma_{lmt}(\mathbf{w}^*) = \phi(\lambda_{lmt} - \beta_t^* \mathbf{w}^*) \quad m = 1, \dots, k-1.$$

Nesse modelo, λ_{lmt} é o “intercept” considerando que a variável aleatória assume valor l no tempo anterior e $\beta_t^* = (\beta_{1t}^*, \dots, \beta_{p+1,t}^*)'$ é um vetor de parâmetros desconhecidos que representam o efeito das variáveis exógenas nas probabilidades $\xi_{lmt}(\mathbf{w}^*)$. A variável que indica a presença de valores perdidos é definida da seguinte maneira:

$$\delta_{lv}^* = \begin{cases} 1 & \text{se } \mathbf{Z}_{lv} \text{ e } \mathbf{W}_{lv}^* \text{ são observados} \\ 0 & \text{caso contrário.} \end{cases}$$

Os parâmetros são estimados através da maximização da seguinte função de verossimilhança:

$$\mathcal{L}_t = \sum_{v=1}^N \delta_{lv}^* \sum_{j=1}^k \mathbf{Z}_{jlv} \{ \log [\phi(\lambda_{lmt} - \mathbf{W}_{lv}^{*'} \beta_t^*) - \phi(\lambda_{l,m-1,t} - \mathbf{W}_{lv}^{*'} \beta_t^*)] \} \quad t = 1, \dots, T, \quad (3.5)$$

onde $\phi(\lambda_{l0t} - \mathbf{W}_{lv}^{*'} \beta_t^*) = 0$ e $\phi(\lambda_{lkt} - \mathbf{W}_{lv}^{*'} \beta_t^*) = 1$.

A maximização de (3.5) pode ser feita através de um método iterativo, por exemplo Newton-Raphson, obtendo-se dessa forma as estimativas $\hat{\lambda}_{lt} = (\hat{\lambda}_{l1t}, \dots, \hat{\lambda}_{lkt})'$ e $\hat{\beta}_t^*$ de λ_{lt} e β_t^* respectivamente. As estimativas das probabilidades de transição podem ser obtidas como segue:

$$\begin{aligned} \hat{\xi}_{l1t}(\mathbf{w}^*) &= \hat{\gamma}_{l1t}(\mathbf{w}^*) = \phi(\hat{\lambda}_{l1t} - \hat{\beta}_t^{*'} \mathbf{w}^*), \\ \hat{\xi}_{l2t}(\mathbf{w}^*) &= \hat{\gamma}_{l2t}(\mathbf{w}^*) - \hat{\xi}_{l1t}(\mathbf{w}^*) = \phi(\hat{\lambda}_{l2t} - \hat{\beta}_t^{*'} \mathbf{w}^*) - \hat{\xi}_{l1t}(\mathbf{w}^*), \\ &\vdots \\ \hat{\xi}_{l,k-1,t}(\mathbf{w}^*) &= \hat{\gamma}_{l,k-1,t}(\mathbf{w}^*) - \hat{\xi}_{l1t}(\mathbf{w}^*) - \hat{\xi}_{l2t}(\mathbf{w}^*) - \dots - \hat{\xi}_{l,k-2,t}(\mathbf{w}^*) \\ &= \phi(\hat{\lambda}_{l,k-1,t} - \hat{\beta}_t^{*'} \mathbf{w}^*) - \hat{\xi}_{l1t}(\mathbf{w}^*) - \hat{\xi}_{l2t}(\mathbf{w}^*) - \dots - \hat{\xi}_{l,k-2,t}(\mathbf{w}^*), \\ \hat{\xi}_{lkt}(\mathbf{w}^*) &= 1 - \hat{\gamma}_{l,k-1,t}(\mathbf{w}^*) = 1 - \phi(\hat{\lambda}_{l,k-1,t} - \hat{\beta}_t^{*'} \mathbf{w}^*). \end{aligned}$$

As variâncias assintóticas para essas estimativas podem ser calculadas através do método δ multivariado (ver Apêndice B-1). A matriz de covariância assintótica entre $(\hat{\lambda}_{lr}, \hat{\beta}_r^*)'$ e $(\hat{\lambda}_{lt}, \hat{\beta}_t^*)'$ com $(1 \leq r \leq t \leq T)$ é estimada utilizando-se o mesmo procedimento descrito em Stram, Wei & Ware (1988).

3.4 Exemplo

Vamos ilustrar os métodos descritos nas seções anteriores, através de um exemplo extraído de Ware, Lipsitz & Speizer (1988). Os dados originais são provenientes de um estudo desenvolvido por Ware *et al.* (1984), sobre os efeitos da poluição atmosférica em doenças respiratórias. Crianças foram examinadas anualmente dos 9 aos 12 anos de idade e de acordo com os exames, foram classificadas em uma das 3 categorias: 1- sem chiado, 2- chiado com resfriado ou 3- chiado independente de resfriado. Observe que este exemplo lida na verdade com um processo em tempo contínuo com observações periódicas, porém os autores a título de ilustração, utilizaram-no como sendo em tempo discreto. Uma análise considerando o tempo contínuo seria mais apropriada, entretanto como não tivemos acesso aos dados originais, isso não nos foi possível.

Foram consideradas duas variáveis exógenas: cidade de residência, que são Kingston-Harriman, TN e Portage, WI que assume valores 1 e 0 respectivamente, e nível de fumante da mãe medido em maços por dia a qual foi atribuída três possíveis valores:

0- menos que $1/2$, 1- entre $1/2$ e $1\frac{1}{2}$ e 2- mais que $1\frac{1}{2}$.

As regressões para as 4 idades em estudo, quando utilizado o modelo marginal, são dadas na tabela 3.2. Todos os resultados expostos nesta seção foram extraídos do trabalho de Ware *et al.* (1988).

Tabela 3.2: Estimativas dos coeficientes da regressão para cada idade em separado (erro padrão)

| | Idade | | | |
|---------------------------|---------------|---------------|---------------|---------------|
| | 9 | 10 | 11 | 12 |
| <i>"Intercept"</i> | | | | |
| Categoria 2 + 3 | -1.27 (0.075) | -1.39 (0.072) | -1.46 (0.075) | -1.71 (0.088) |
| Categoria 3 | -2.32 (0.091) | -2.33 (0.086) | -2.32 (0.088) | -2.56 (0.088) |
| <i>Variáveis Exógenas</i> | | | | |
| Cidade | 0.26 (0.10) | 0.27 (0.10) | 0.30 (0.10) | 0.45 (0.11) |
| Nível de Fumante | 0.015 (0.004) | 0.014 (0.004) | 0.018 (0.004) | 0.025 (0.004) |

A tabela 3.3 mostra as estimativas das probabilidades calculadas através das observações (O) e através do modelo (E) para as crianças filhas de mães não fumantes nas duas cidades, a fim de observar a adequabilidade do ajuste. É importante fazer uma observação com respeito a este exemplo. O modelo de “odds” proporcionais foi especificado com $\gamma_{jt}(\mathbf{w}) = \phi(\lambda_{jt} + \beta_j^t \mathbf{w})$. Note que o sinal no argumento da função é positivo. Essa formulação não altera as conclusões, porém enquanto nesse caso, $\gamma_{1t}(\mathbf{w})$ é a probabilidade marginal para a categoria 3 e $\gamma_{2t}(\mathbf{w})$ é a probabilidade acumulada das categorias 2 e 3, se fosse utilizado o sinal negativo no argumento $\gamma_{1t}(\mathbf{w})$ seria a probabilidade marginal para a categoria 1 e $\gamma_{2t}(\mathbf{w})$ a probabilidade acumulada das categorias 1 e 2 conforme definido em (3.1), desde que mantida a ordem das respostas nos dois casos.

Tabela 3.3: Estimativas das probabilidades marginais para as crianças de mães não fumantes via modelo marginal
O- pelas observações E- pelo modelo

| Kingston-Harriman, TN | | | | | | | | |
|-----------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Idade | | | | | | | | |
| | 9 | | 10 | | 11 | | 12 | |
| N | 514 | | 547 | | 540 | | 553 | |
| Categoria | O | E | O | E | O | E | O | E |
| 1 | 0.751 | 0.734 | 0.761 | 0.755 | 0.787 | 0.760 | 0.785 | 0.779 |
| 2 | 0.134 | 0.153 | 0.141 | 0.132 | 0.115 | 0.121 | 0.119 | 0.113 |
| 3 | 0.115 | 0.113 | 0.099 | 0.113 | 0.098 | 0.117 | 0.096 | 0.108 |
| Portage, WI | | | | | | | | |
| Idade | | | | | | | | |
| | 9 | | 10 | | 11 | | 12 | |
| N | 760 | | 918 | | 875 | | 730 | |
| Categoria | O | E | O | E | O | E | O | E |
| 1 | 0.764 | 0.784 | 0.802 | 0.801 | 0.802 | 0.812 | 0.845 | 0.847 |
| 2 | 0.136 | 0.129 | 0.107 | 0.110 | 0.089 | 0.099 | 0.070 | 0.081 |
| 3 | 0.100 | 0.090 | 0.092 | 0.089 | 0.109 | 0.089 | 0.085 | 0.072 |

Para se calcular as probabilidades marginais, o procedimento é o seguinte: por exemplo, a probabilidade de uma criança filha de mãe não fumante estar na categoria 3 aos 12 anos de idade, na cidade de Kingston-Harriman, TN, denotada por $p_3(12)$, é igual a $\phi(\lambda_{3,12} + \beta_{12}^3 \mathbf{w})$,

com $\lambda_{3,12} = -2.56$, $\beta_{12} = (0.45 \ 0.025)'$ e $w = (1 \ 0)'$. Substituindo os valores, temos $p_3(12) = e^{-2.56+0.45}/(1 + e^{-2.56+0.45})$, ou seja, a probabilidade de uma criança que mora em Kingston-Harriman, TN aos 12 anos ter chiado, independente de estar resfriada é igual a 0.108. Já a probabilidade de estar na categoria 2 ou na categoria 3 é igual a $p_{(2+3)}(12) = \phi(\lambda_{(2+3),12} + \beta_{12}'w)$, com $\lambda_{(2+3),12} = -1.71$, $\beta_{12} = (0.45 \ 0.025)'$ e $w = (1 \ 0)'$, ou seja, $p_{(2+3)}(12) = e^{-1.71+0.45}/(1 + e^{-1.71+0.45}) = 0.221$ e portanto, $p_2(12) = p_{(2+3)}(12) - p_3(12) = 0.221 - 0.108 = 0.113$, conseqüentemente $p_1(12) = 1 - p_{(2+3)}(12) = 1 - 0.221 = 0.779$. Dessa forma, todas as probabilidades marginais para as 4 ocasiões podem ser calculadas.

Para o modelo de transição, foram obtidas as regressões para as 3 transições ocorridas durante os 4 anos do estudo. Essas regressões estão na tabela 3.4. Note que enquanto na tabela 3.2 temos 4 colunas (9, 10, 11 e 12), na tabela 3.5 temos 3 colunas (10, 11 e 12) nas quais a primeira representa a transição entre 9 e 10 anos de idade, a segunda entre 10 e 11 anos e a terceira entre 11 e 12 anos.

Tabela 3.4: Estimativas dos coeficientes da regressão para o modelo de transição (erro padrão)

| | Idade | | |
|---------------------------|--------------|--------------|--------------|
| | 10 | 11 | 12 |
| <i>"Intercept"</i> | | | |
| Categoria 2 + 3 | -2.27 (0.11) | -2.23 (0.11) | -2.51 (0.12) |
| Categoria 3 | -3.63 (0.14) | -3.40 (0.13) | -3.61 (0.14) |
| <i>Variáveis Exógenas</i> | | | |
| Cidade | 0.05 (0.13) | 0.21 (0.12) | 0.33 (0.13) |
| Nível de Fumante | 0.01 (0.01) | 0.01 (0.01) | 0.01 (0.01) |
| $I_{X(t-1)=2}$ | 2.17 (0.14) | 2.00 (0.15) | 2.13 (0.16) |
| $I_{X(t-1)=3}$ | 3.54 (0.18) | 3.42 (0.17) | 3.27 (0.18) |

As variáveis exógenas $I_{X(t-1)=2}$ e $I_{X(t-1)=3}$ são variáveis indicadoras para a categoria no tempo imediatamente anterior definidas da seguinte maneira:

$$I_{X(t-1)=i} = \begin{cases} 1 & \text{se } X(t-1) = i \\ 0 & \text{caso contrário} \end{cases} \quad i = 2 \text{ ou } 3.$$

Para a obtenção das estimativas das probabilidades de transição, basta utilizar os

coeficientes da regressão da tabela 3.4 em conjunto com o vetor de variáveis exógenas. Por exemplo, a probabilidade de estar na categoria 3 aos 10 anos de idade, dado que estava na categoria 2 aos 9 anos, é calculada através de $\phi(-3.63 + 0.05w_1 + 0.01w_2 + 2.17)$. Onde w_1 e w_2 recebem valores conforme a cidade e em que nível de fumante se encontra a mãe da criança respectivamente. A probabilidade de estar na categoria 2 aos 10 anos de idade, dado que estava na categoria 1 aos 9 anos, é calculada através de $\phi(-2.27 + 0.049w_1 + 0.014w_2) - \phi(-3.63 + 0.049w_1 + 0.014w_2)$. A tabela 3.5 mostra as estimativas das probabilidades de transição calculadas através dos valores observados (O) e através do modelo (E) de 9 para 10 anos de idade nas duas cidades, para as crianças cujas mães são não fumantes. Pode-se observar nessa tabela que essas probabilidades não são tão próximas quanto no modelo marginal, o que mostra que o ajuste não foi tão adequado quando utilizou-se modelos de transição.

Tabela 3.5: Probabilidades de transição entre 9 e 10 anos para filhos de mães não fumantes
O- pelas observações E- pelo modelo

| Kingston-Harriman, TN | | | | | | | | |
|-----------------------|---|-------|-------|-------|-------|-------|-------|-----|
| 10 anos | | | | | | | | |
| Categoria | 1 | | 2 | | 3 | | N | |
| | O | E | O | E | O | E | | |
| 9 anos | 1 | 0.889 | 0.902 | 0.068 | 0.070 | 0.043 | 0.027 | 324 |
| | 2 | 0.458 | 0.514 | 0.492 | 0.290 | 0.051 | 0.196 | 59 |
| | 3 | 0.391 | 0.211 | 0.174 | 0.298 | 0.435 | 0.491 | 46 |
| Portage, WI | | | | | | | | |
| 10 anos | | | | | | | | |
| Categoria | 1 | | 2 | | 3 | | N | |
| | O | E | O | E | O | E | | |
| 9 anos | 1 | 0.935 | 0.907 | 0.046 | 0.067 | 0.019 | 0.026 | 523 |
| | 2 | 0.430 | 0.527 | 0.440 | 0.285 | 0.130 | 0.189 | 100 |
| | 3 | 0.200 | 0.219 | 0.167 | 0.302 | 0.633 | 0.479 | 60 |

Apesar do modelo de Markov de primeira ordem não ter sido muito adequado nesse exemplo, podemos constatar, verificando os resultados expostos na tabela 3.6, que as formas condicional e não condicional do modelo de McCullagh dão estimativas das probabilidades marginais bastante semelhantes.

Tabela 3.6: Probabilidades marginais estimadas pelos modelos marginal (M) e de transição (C) para crianças filhas de mães não fumantes

| Kingston-Harriman, TN | | | | | | | | |
|-----------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Idade | | | | | | | | |
| | 9 | | 10 | | 11 | | 12 | |
| Categoria | M | C | M | C | M | C | M | C |
| 1 | 0.734 | 0.734 | 0.755 | 0.765 | 0.762 | 0.762 | 0.779 | 0.775 |
| 2 | 0.153 | 0.153 | 0.132 | 0.130 | 0.121 | 0.119 | 0.112 | 0.110 |
| 3 | 0.113 | 0.113 | 0.113 | 0.105 | 0.117 | 0.119 | 0.109 | 0.115 |
| Portage, WI | | | | | | | | |
| Idade | | | | | | | | |
| | 9 | | 10 | | 11 | | 12 | |
| Categoria | M | C | M | C | M | C | M | C |
| 1 | 0.781 | 0.781 | 0.801 | 0.796 | 0.812 | 0.804 | 0.847 | 0.834 |
| 2 | 0.129 | 0.129 | 0.110 | 0.116 | 0.099 | 0.103 | 0.081 | 0.087 |
| 3 | 0.090 | 0.090 | 0.089 | 0.088 | 0.089 | 0.093 | 0.072 | 0.079 |

De posse desses resultados, testes de hipóteses, conforme descritos na seção 2.4.3, podem ser feitos a fim de tornar o modelo mais adequado. Podemos testar por exemplo, se a utilização de uma cadeia de Markov de 2ª ordem melhoraria as estimativas das probabilidades de transição. Devemos enfatizar ainda que esses modelos são usados de forma ilustrativa e não definitiva, ou seja, são modelos exploratórios, em outras palavras, eles servem para dar uma idéia de como estão acontecendo as mudanças de estado no decorrer do tempo em termos das probabilidades de transição.

Capítulo 4

Análise Utilizando Modelos para Processos em Tempo Contínuo

4.1 Introdução

Na maioria dos processos estocásticos que se deseja estudar, as transições ocorrem em tempo contínuo e não em tempo discreto. Apesar disso, na prática frequentemente é impossível observar tais processos continuamente por diversos motivos, por exemplo, escassez de recursos financeiros, uma vez que um acompanhamento dessa natureza pode ser muito dispendioso. O que se tem disponível geralmente, são processos em tempo contínuo observados periodicamente e o que se quer é utilizá-los de forma apropriada no sentido de tentar explicar o processo real através de um modelo discreto no qual as mudanças são tratadas como se tivessem ocorrido nos tempos das observações. Surge então a necessidade de se desenvolver métodos para estimação dos parâmetros de modelos contínuos a partir de observações discretas.

Neste capítulo descrevemos o método desenvolvido por Kalbfleisch & Lawless, que fornece uma maneira muito eficiente de obter os estimadores de máxima verossimilhança para os parâmetros. Na seção 4.2 descrevemos o algoritmo utilizado pelos autores, bem como a forma de obter a matriz de covariância assintótica dos estimadores dos parâmetros. O mesmo é feito para outras características de interesse no processo. A seção 4.3 é dedicada à generalização

para o uso de variáveis exógenas no modelo. Na seção 4.4 são discutidos exemplos a título de ilustração, com a finalidade de facilitar a compreensão das técnicas descritas.

4.2 Método de Kalbfleisch & Lawless

Vamos supor que as unidades experimentais movem-se independentemente entre k estados ou categorias denotadas por $1, \dots, k$, de acordo com um processo de Markov em tempo contínuo. Inicialmente consideramos que as probabilidades de transição são estacionárias, ou seja, o processo de Markov em questão é estacionário. Dessa forma, conforme foi descrito na seção 2.3 do capítulo 2 (ver equação (2.18)), temos que a matriz de probabilidades de transição para o intervalo (s, t) com $0 < s < t$, para processos de Markov é dada por:

$$P(s, t) = e^{(t-s)Q} \quad (4.1)$$

onde Q é a matriz de intensidades de transição, cujos elementos são definidos em (2.15). Vamos supor ainda que as intensidades de transição q_{ij} dependam de b parâmetros funcionalmente independentes $\theta_1, \dots, \theta_b$, isto é, $q_{ij} = q_{ij}(\theta)$ com $\theta = (\theta_1, \dots, \theta_b)'$ para cada $i, j = 1, \dots, k$, conseqüentemente $p_{ij}(t) = p_{ij}(t; \theta)$. Como ilustração usemos o seguinte exemplo: supor um processo cujo espaço de estados é composto por três estados 1, 2 e 3. Uma possível matriz de intensidades de transição seria

$$Q = \begin{pmatrix} -(\theta_1 + \theta_2) & \theta_1 & \theta_2 \\ \theta_3 & -(\theta_3 + \theta_4) & \theta_4 \\ \theta_5 & \theta_6 & -(\theta_5 + \theta_6) \end{pmatrix},$$

e portanto teríamos $\theta = (\theta_1, \dots, \theta_6)'$, $b = 6$ e $i, j = 1, 2, 3$ onde os θ_l 's ($l = 1, \dots, 6$) não têm qualquer relação funcional entre si. O problema passa a ser estimar o vetor θ , e como fazê-lo é descrito nas subseções que se seguem.

4.2.1 Estimação por máxima verossimilhança

Para o desenvolvimento do método, vamos inicialmente supor que todas as N unidades experimentais são observadas nos tempos t_0, t_1, \dots, t_T , ou seja, por enquanto não levamos em

consideração a possibilidade de existência de dados perdidos (“missing values”). Denotamos o número de unidades experimentais no estado j no tempo t_l que estavam no estado i no tempo t_{l-1} como $n_{ij}(l)$. A função de verossimilhança para θ é baseada na distribuição multinomial, conforme descrito na seção 2.4.2 (ver equação (2.28)) e na suposição de que as unidades experimentais movem-se independentemente entre os k estados. Ela pode ser representada por:

$$\mathcal{L}(\theta) = \prod_{l=1}^T \left\{ \prod_{i,j=1}^k p_{ij}(t_{l-1}, t_l)^{n_{ij}(l)} \right\}. \quad (4.2)$$

No caso estacionário com $y_l = t_l - t_{l-1}$, $l = 1, \dots, T$, e aplicando o logaritmo em (4.2), temos:

$$\log \mathcal{L}(\theta) = \sum_{l=1}^T \sum_{i,j=1}^k n_{ij}(l) \log p_{ij}(y_l). \quad (4.3)$$

Deve-se notar que em $p_{ij}(y_l)$ está implicitamente embutida a dependência de θ . Em outras palavras, $p_{ij}(y_l)$ representa $p_{ij}(y_l; \theta)$.

O estimador de máxima verossimilhança $\hat{\theta}$ de θ ou, equivalentemente, o estimador de máxima verossimilhança $\hat{Q} = Q(\hat{\theta})$ de Q , é obtido através da maximização de (4.3). Em geral, não existe solução explícita para esse problema e métodos iterativos devem ser utilizados para resolvê-lo. Vários algoritmos podem ser usados, um deles é o de Newton-Raphson. Os autores sugerem o método denominado “Scoring” de Fisher que consiste numa pequena modificação do Newton-Raphson. No primeiro, as derivadas primeira e segunda de (4.3) têm que ser calculadas explicitamente enquanto que no segundo, a derivada segunda é substituída por sua esperança, de modo que quando da implementação do algoritmo para a maximização de (4.3), somente a primeira derivada é efetivamente calculada, o que é, em termos computacionais, um bom motivo para a escolha do método. Ainda, segundo Kalbfleisch & Lawless, a convergência, caso ocorra, é mais rápida e o método dá uma estimativa para a matriz de covariância assintótica de $\hat{\theta}$ diretamente.

O cálculo da primeira derivada de (4.3) requer as derivadas de $P(s, t)$ com relação a θ . É necessário portanto, construir um algoritmo eficiente para o cálculo de $P(s, t)$ e suas derivadas com relação a θ . A matriz $P(s, t)$ pode ser escrita conforme equação (4.1). A matriz Q pode ser

decomposta em: $Q = ADA^{-1}$, com $D = \text{diag}(d_1, \dots, d_k)$ onde d_1, \dots, d_k são seus autovalores e A a matriz quadrada de ordem k cuja j -ésima coluna é um autovetor correspondente a d_j . Então,

$$P(s, t) = A \text{diag}(e^{d_1(t-s)}, \dots, e^{d_k(t-s)}) A^{-1}, \quad (4.4)$$

onde a dependência de Q , $P(s, t)$, A e dos d_j 's com relação a θ é suprimida por conveniência notacional. É importante observar que esse resultado é válido para autovalores distintos. Caso a matriz Q tenha autovalores repetidos utiliza-se uma decomposição de Jordan para a matriz (ver Strang (1980) - Apêndice B), porém isso raramente é necessário, uma vez que para a quase totalidade dos modelos de interesse, Q tem autovalores distintos para quase todo θ , e por isso não levamos em conta esse caso. A forma de obtenção do resultado (4.4) é dada no Apêndice A-1.

As derivadas podem ser calculadas de uma forma semelhante. Em particular, a matriz com elementos $\partial p_{ij}(s, t; \theta) / \partial \theta_u$ pode ser obtida por:

$$\frac{\partial P(s, t)}{\partial \theta_u} = AV_u A^{-1}, \quad u = 1, \dots, b, \quad (4.5)$$

onde V_u é uma matriz quadrada de ordem k com (i, j) -ésimo elemento dado por:

$$\begin{aligned} & \frac{g_{ij}^{(u)} [e^{d_i(t-s)} - e^{d_j(t-s)}]}{(d_i - d_j)} \quad \text{se } i \neq j \\ & g_{ii}^{(u)} (t-s) e^{d_i(t-s)} \quad \text{se } i = j, \end{aligned}$$

onde $g_{ij}^{(u)}$ é o (i, j) -ésimo elemento de $G^{(u)} = A^{-1}(\partial Q / \partial \theta_u)A$. A descrição de como esse resultado é obtido está no Apêndice A-2. As derivadas $\partial Q / \partial \theta_u$ são geralmente muito simples e o cálculo de (4.5) é fácil quando se tem A e D . Para o exemplo citado anteriormente, a derivada de Q em relação a θ_1 é

$$\left(\frac{\partial Q}{\partial \theta_1} \right) = \begin{pmatrix} -1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Uma das vantagens do método descrito é que ele dispensa o uso da representação explícita dos elementos $p_{ij}(s, t)$ como funções de θ , que em geral é bastante complicada para se trabalhar, como pode ser notado no primeiro exemplo da seção 4.4, exposto com essa finalidade.

4.2.2 Utilização do procedimento “Scoring” de Fisher

De (4.3) temos:

$$S_u(\boldsymbol{\theta}) = \frac{\partial \log \mathcal{L}(\boldsymbol{\theta})}{\partial \theta_u} = \sum_{l=1}^T \sum_{i,j=1}^k n_{ij}(l) \frac{\partial p_{ij}(y_l) / \partial \theta_u}{p_{ij}(y_l)}, \quad u = 1, \dots, b, \quad (4.6)$$

e

$$\frac{\partial^2 \log \mathcal{L}(\boldsymbol{\theta})}{\partial \theta_u \partial \theta_v} = \sum_{l=1}^T \sum_{i,j=1}^k n_{ij}(l) \times \left\{ \frac{\partial^2 p_{ij}(y_l) / \partial \theta_u \partial \theta_v}{p_{ij}(y_l)} - \frac{\partial p_{ij}(y_l) / \partial \theta_u \partial p_{ij}(y_l) / \partial \theta_v}{p_{ij}^2(y_l)} \right\}.$$

Como foi mencionado na seção 4.2.1, o método de “Scoring” de Fisher requer o cálculo da esperança da segunda derivada de (4.3) que pode ser feito da seguinte forma:

Seja $n_i(l) = \sum_{j=1}^k n_{ij}(l)$ o número de indivíduos no estado i no tempo t_{l-1} . Assumindo a esperança de $n_{ij}(l)$ condicional em $N_i(t_{l-1})$ e usando o fato que $\sum_{j=1}^k (\partial^2 p_{ij}(y_l) / \partial \theta_u \partial \theta_v) = 0$, temos que

$$E \left(-\frac{\partial^2 \log \mathcal{L}(\boldsymbol{\theta})}{\partial \theta_u \partial \theta_v} \right) = \sum_{l=1}^T \sum_{i,j=1}^k \frac{E[n_i(l)]}{p_{ij}(y_l)} \times \frac{\partial p_{ij}(y_l)}{\partial \theta_u} \times \frac{\partial p_{ij}(y_l)}{\partial \theta_v},$$

que pode ser estimada por

$$M_{uv}(\hat{\boldsymbol{\theta}}) = \sum_{l=1}^T \sum_{i,j=1}^k \frac{n_i(l)}{p_{ij}(y_l)} \times \frac{\partial p_{ij}(y_l)}{\partial \theta_u} \times \frac{\partial p_{ij}(y_l)}{\partial \theta_v} \Big|_{\theta_u = \hat{\theta}_u \text{ e } \theta_v = \hat{\theta}_v} \quad (4.7)$$

O cálculo de (4.6) e (4.7) para qualquer $\boldsymbol{\theta}$ dado é simplificado utilizando-se os resultados (4.4) e (4.5).

Seja $\boldsymbol{\theta}_0$ uma estimativa inicial de $\boldsymbol{\theta}$, $\mathbf{s}(\boldsymbol{\theta})$ o vetor $b \times 1$ com i -ésimo elemento dado por $S_u(\boldsymbol{\theta})$, e $\mathbf{M}(\boldsymbol{\theta})$ a matriz $b \times b$ com (u, v) -ésimo componente dado por $M_{uv}(\boldsymbol{\theta})$. Uma nova estimativa para $\boldsymbol{\theta}$ é obtida através de

$$\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 + \mathbf{M}(\boldsymbol{\theta}_0)^{-1} \mathbf{s}(\boldsymbol{\theta}_0), \quad (4.8)$$

onde é assumido que $\mathbf{M}(\boldsymbol{\theta}_0)$ é não singular. O processo é repetido com $\boldsymbol{\theta}_1$ no lugar de $\boldsymbol{\theta}_0$ gerando $\boldsymbol{\theta}_2$, e assim sucessivamente, até alcançar a convergência desejada, por exemplo, quando $(\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0\| / \|\boldsymbol{\theta}_0\|) \leq 10^{-3}$, onde $\|\mathbf{x}\|$ denota a norma do vetor \mathbf{x} . O algoritmo produz uma estimativa $\hat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$ e $\mathbf{M}(\hat{\boldsymbol{\theta}})^{-1}$ é uma estimativa da matriz de covariância assintótica de $\hat{\boldsymbol{\theta}}$. Se o

verdadeiro valor de θ é um ponto interior do espaço paramétrico, $\sqrt{N}(\hat{\theta} - \theta)$ tem assintoticamente distribuição normal multivariada conforme Teorema 1, descrito no capítulo 2. Esse resultado é fundamental para a realização de inferências sobre os parâmetros. É aconselhável parametrizar o modelo fazendo $q_{ij} = \exp(\alpha_{ij})$, $i \neq j$, uma vez que α_{ij} pode assumir qualquer valor real, enquanto que $q_{ij} \geq 0$. Essa reparametrização garante que o resultado de uma iteração não cairá fora do espaço paramétrico e, conseqüentemente, contorna problemas que podem surgir quando uma iteração resulta em vetores de parâmetros fora do espaço paramétrico.

Frequentemente, características específicas do modelo são de interesse para o pesquisador, por exemplo, $-q_{ii}(\theta)^{-1}$, o tempo médio de permanência no estado i . Nesses casos, tais quantidades são estimadas substituindo θ por $\hat{\theta}$, e a variância assintótica é estimada pelo teorema delta multivariado (ver Apêndice B-1). O tempo médio de permanência no estado i é estimado por $-q_{ii}(\hat{\theta})^{-1}$ e sua variância assintótica estimada é

$$q_{ii}(\hat{\theta})^{-4} \times \sum_{u=1}^b \sum_{v=1}^b \frac{\partial q_{ii}(\theta)}{\partial \theta_u} \times \frac{\partial q_{ii}(\theta)}{\partial \theta_v} \times M^{uv}(\theta) \Big|_{\theta=\hat{\theta}}$$

onde $M^{uv}(\theta)$ é o (u, v) -ésimo elemento de $\mathbf{M}(\theta)^{-1}$. No Apêndice B-2 damos alguns exemplos de estimação de outras características que podem ser de interesse do pesquisador.

4.2.3 Algumas extensões para o modelo

Muitas vezes, alguns dos indivíduos são observados somente sobre parte do período determinado para o estudo. Se os indivíduos que entram após o estudo ter iniciado ou saem antes do término do mesmo, são iguais àqueles que permaneceram durante todo o período, isto é, mesmas probabilidades de transição, então a metodologia pode ser aplicada sem alterações. Deve-se notar apenas que, $n_i(t)$ de (4.7) é o número de indivíduos no estado i no tempo t_{i-1} para o qual o estado ocupado no tempo t_i é conhecido.

Da mesma forma, é desnecessário que todos os indivíduos sejam observados sobre o mesmo conjunto de pontos no tempo. Porém não é aconselhável esse procedimento, uma vez que a álgebra requerida aumenta significativamente com o número de intervalos distintos de

tempo y_i na amostra.

Esses métodos podem ser estendidos de uma forma simples para o ajuste de certos modelos de Markov não estacionários. Suponha por exemplo, que $\{X(t); t = 0, \dots, T\}$ é um processo de Markov com matriz de intensidades de transição dependentes do tempo $Q(t) = Q_0 \times g(t; \lambda)$, onde Q_0 é uma matriz de intensidades fixada, com elementos desconhecidos q_{ij} e $g(t; \lambda)$ é uma função conhecida do tempo a menos do parâmetro não especificado λ . Nesse caso, $g(t; \lambda)$ define para um dado λ , um tempo operacional. Para dado λ , seja $s = \int_0^t g(u; \lambda) du$ e defina $Y(s) = X(t)$. O processo $\{Y(s); 0 < s < \infty\}$ é portanto, um processo de Markov estacionário com matriz de intensidades de transição Q_0 . Para um dado λ , t_i será substituído por $s_i = \int_0^{t_i} g(u; \lambda) du$ e y_i por $y_i^* = s_i - s_{i-1}$. Os parâmetros de Q_0 são estimados pelos métodos descritos anteriormente. Além disso, o logaritmo da verossimilhança maximizado pode ser obtido para aquele λ dado. Variando λ , esse parâmetro adicional pode ser estimado observando o efeito dele no máximo do logaritmo da verossimilhança.

Segundo Kalbfleisch & Lawless, o enfoque para estimação da variância e intervalos de confiança é considerar λ como sendo fixo, geralmente o valor $\hat{\lambda}$ que maximiza o logaritmo da verossimilhança maximizado e usá-lo para estimar a matriz de covariância $M(\hat{\theta})^{-1}$ para a correspondente estimativa $\hat{\theta}$ de θ .

Uma outra maneira de se trabalhar com modelos dependentes do tempo é permitir que a matriz de intensidades de transição mude para cada tempo de observação, porém seja constante entre os tempos de observação. Por exemplo,

$$Q(t) = \begin{cases} Q_1 & \text{se } 0 \leq t < t_{m_1} \\ Q_2 & \text{se } t_{m_1} \leq t < t_m, \end{cases}$$

para algum específico m_1 menor que m . A estimação de Q_1 e Q_2 é feita considerando os primeiros m_1 intervalos de tempo e os $m - m_1$ subsequentes respectivamente, em separado.

4.3 A incorporação de variáveis exógenas

Na grande maioria das aplicações práticas, as intensidades de transição são funções de variáveis exógenas que têm influência na resposta. Por exemplo, o fato de um paciente de uma doença pulmonar passar de um estágio pior para um melhor dessa doença pode estar diretamente ligado aos níveis de poluição atmosférica do lugar onde reside, a quantidade de cigarros que ele fuma diariamente, a aplicação de um determinado tipo de tratamento, e assim por diante. Uma característica importante dos métodos desenvolvidos por Kalbfleisch & Lawless é que eles permitem a incorporação dessa informação na estimação das intensidades de transição, o que até então nenhum outro método havia feito. O que se faz é modelar a matriz de intensidades de transição Q através de regressões onde as intensidades são as variáveis dependentes e as variáveis exógenas são as variáveis regressoras. Desse modo estamos quantificando a relação entre as variáveis exógenas e as intensidades de transição no modelo de Markov. Uma restrição deve ser observada: muitos valores distintos para as variáveis regressoras torna a tarefa computacional quase que inviável. A implementação pode exigir que as variáveis regressoras sejam grupadas.

Suponhamos que cada indivíduo tenha associado a ele um vetor de c variáveis exógenas, que são as variáveis regressoras $\mathbf{w} = (w_1, w_2, \dots, w_c)'$, onde $w_1 = 1$. Note que na verdade temos $c - 1$ variáveis regressoras, já que a primeira é sempre igual a 1, por isso deve ficar claro que quando dizemos que temos c variáveis exógenas, estamos dizendo $c - 1$ variáveis mais uma com valor 1. Para um dado \mathbf{w} , vamos supor que o processo é de Markov estacionário com matriz de intensidades de transição

$$Q(\mathbf{w}) = (q_{ij}(\mathbf{w})),$$

onde

$$\begin{aligned} q_{ij}(\mathbf{w}) &= \exp(\mathbf{w}'\boldsymbol{\beta}_{ij}) \quad \text{se } i \neq j \\ q_{ii}(\mathbf{w}) &= -\sum_{j \neq i} q_{ij}(\mathbf{w}) \quad \text{para } \forall i \in \mathcal{S} \end{aligned} \tag{4.9}$$

Em (4.9), $\boldsymbol{\beta}_{ij} = (\beta_{1ij}, \dots, \beta_{cij})'$ é um vetor composto pelos c parâmetros que relacionam as intensidades de transição do estado i para o estado j às variáveis regressoras \mathbf{w} . Note

que β_{1ij} corresponde ao “intercept” da regressão. Conforme a aplicação, outros tipos de parametrização podem ser mais apropriados. Esse modelo tem a característica interessante de produzir intensidades de transição não negativas quaisquer que sejam w e β_{ij} , e tem sido sugerido por diversos autores (ver Tuma & Robins (1980)).

O algoritmo da seção 4.2 requer uma decomposição de $Q(w)$ para cada um dos r vetores distintos compostos pelos valores assumidos pelas variáveis regressoras w na amostra. Vamos denotá-los por $w_h = (w_{1h}, \dots, w_{ch})'$ com $w_{1h} = 1$, e seja

$$Q_h = Q(w_h) = (q_{ij}(w_h)), \quad h = 1, \dots, r,$$

conforme definido em (4.9). Por exemplo, no caso de 2 variáveis regressoras com valores 0 e 1 teríamos $w_1 = (1, 0, 0)'$, $w_2 = (1, 0, 1)'$, $w_3 = (1, 1, 0)'$ e $w_4 = (1, 1, 1)'$ de modo que seriam necessárias no máximo 4 decomposições diferentes, uma para cada Q_h com $h = 1, \dots, 4$. Obviamente se algum dos vetores não estiver presente na amostra, não será computado. Seja $n_{ij}^{(h)}(l)$ o número de indivíduos que estavam em i no tempo t_{l-1} e em j no tempo t_l para o vetor w_h . A verossimilhança é então um produto de termos como em (4.2), onde o h -ésimo termo surge dos dados coletados no modelo estacionário com matriz de intensidades Q_h . Portanto o logaritmo da verossimilhança é

$$\log \mathcal{L}(\theta) = \sum_{h=1}^r \sum_{l=1}^T \sum_{i,j=1}^k n_{ij}^{(h)}(l) \log p_{ij}(y_l; w_h), \quad (4.10)$$

onde

$$P_h(s, t) = \exp(Q_h(t - s)) = (p_{ij}(t - s; w_h)).$$

O parâmetro θ será utilizado para representar o vetor de parâmetros β_{ij} ($i \neq j$) a ser estimado. As equações de máxima verossimilhança envolvem a soma de r termos, um para cada w_h . Temos então

$$s(\theta) = \sum_{h=1}^r s^{(h)}(\theta), \quad (4.11)$$

onde $s^{(h)}(\theta)$ é calculado como na seção 4.2 (ver equação (4.6)). Da mesma forma,

$$M(\theta) = \sum_{h=1}^r M^{(h)}(\theta), \quad (4.12)$$

é calculada com $M^{(h)}(\theta)$ para cada r usando (4.7) e as fórmulas (4.4) e (4.5). O algoritmo necessita das derivadas de $P_h(s, t)$ com respeito aos elementos de θ . Note que $G^{(u)}$ em (4.5) exige o cálculo das derivadas de $Q(w_h)$ com respeito a θ_u . Nesse caso, como $\theta_u = \beta_{pij}$ para algum $p = 1, \dots, c, i, j = 1, \dots, k$, essas derivadas são facilmente calculadas:

$$\frac{\partial Q(w_h)}{\partial \beta_{pij}} = w_{ph} q_{ij}(w_h) L, \quad (4.13)$$

onde L é uma matriz $k \times k$ com todos os elementos iguais a 0 exceto $L_{ij} = 1$ e $L_{ii} = -1$. A especificação e ajuste de modelos utilizando variáveis exógenas serão exemplificados no Exemplo 4.3 da próxima seção.

4.4 Exemplos

Vamos nesta seção, através de alguns exemplos, discutir alguns aspectos relacionados aos métodos descritos nas seções anteriores. O objetivo é usá-los como um auxílio para uma melhor compreensão das idéias expostas anteriormente sem a preocupação de desenvolver análises detalhadas dos dados, o que é feito no capítulo 5 com outro conjunto de dados.

EXEMPLO 4.1: Vamos considerar um experimento sobre evolução da renda no estudo das mudanças no “status” conjugal discutido em Tuma *et al.* (1979).

O objetivo desse exemplo é mostrar como as probabilidades de transição escritas explicitamente como função dos parâmetros de θ tornaria o trabalho de estimação extremamente mais árduo. O espaço de estados para esse estudo é composto por: 1- solteiro, 2- casado e 3- saiu do estudo. O estado 3 é um estado absorvente. A matriz de intensidades de transição Q tem a seguinte forma:

$$Q = \begin{pmatrix} -(\theta_1 + \theta_2) & \theta_1 & \theta_2 \\ \theta_3 & -(\theta_3 + \theta_4) & \theta_4 \\ 0 & 0 & 0 \end{pmatrix}.$$

Sendo assim, $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)'$, onde $(\theta_i > 0, i = 1, \dots, 4)$. É importante observar que $\theta_1, \theta_2, \theta_3$ e θ_4 não têm qualquer relação funcional entre si, ou seja, são funcionalmente independentes.

Sem utilizar a metodologia descrita anteriormente, podemos escrever a matriz $P(s, t)$ como função de θ . Através de técnicas para solução de sistemas de equações diferenciais, temos:

$$p_{jj}(s, t) = \frac{1}{(\delta_1 - \delta_2)} [(\alpha_k + \delta_1)e^{\delta_1(t-s)} - (\alpha_k + \delta_2)e^{\delta_2(t-s)}]$$

$$p_{jk}(s, t) = \frac{q_{jk}}{(\delta_1 - \delta_2)} [e^{\delta_1(t-s)} - e^{\delta_2(t-s)}]$$

$$p_{j3}(s, t) = 1 + \frac{1}{(\delta_1 - \delta_2)} [(q_{j3} + \delta_2)e^{\delta_1(t-s)} - (q_{j3} + \delta_1)e^{\delta_2(t-s)}]$$

$$p_{31}(s, t) = p_{32}(s, t) = 0$$

$$p_{33}(s, t) = 1$$

onde $j = 1$ ou 2 , $k = 3 - j$, $\delta_1 \neq \delta_2$, $\alpha_k = \lambda_k + \mu_k$, $\lambda_1 = \theta_1$, $\lambda_2 = \theta_3$, $\mu_1 = \theta_2$, $\mu_2 = \theta_4$,

$$\delta_1 = -[\alpha_1 + \alpha_2 + \sqrt{(\alpha_1 - \alpha_2)^2 + 4\lambda_1\lambda_2}]/2$$

e

$$\delta_2 = -[\alpha_1 + \alpha_2 - \sqrt{(\alpha_1 - \alpha_2)^2 + 4\lambda_1\lambda_2}]/2.$$

É fácil perceber que o cálculo das derivadas de $P(s, t)$ em relação a θ seria bastante trabalhoso. Por outro lado, através da metodologia descrita, daríamos valores iniciais para θ e obteríamos a matriz Q , faríamos a decomposição e encontraríamos A e D . Posteriormente substituiríamos na equação (4.4). As derivadas seriam facilmente encontradas através de (4.5).

EXEMPLO 4.2: Vamos ver agora um outro exemplo, extraído de Kalbfleisch & Lawless (1985), o qual será utilizado para aplicar a metodologia descrita. Nesse exemplo foi estudado o comportamento com relação ao hábito de fumar em crianças de duas cidades do Canadá (Waterloo e Oxford). Foi considerado como o início dos estudos a entrada no 6º ano colegial e denotou-se $t_0 = 0$. As ocasiões de observação (em anos) foram $t_1 = 0.15$, $t_2 = 0.75$, $t_3 = 1.10$ e $t_4 = 1.90$. O objetivo do estudo foi comparar um grupo controle de cada cidade com um grupo que recebeu material didático sobre o fumo durante os dois primeiros meses do estudo nas duas cidades. A resposta obtida foi classificada em três categorias: 1- nunca fumou, 2- é fumante ou 3- fumou, mas parou. Assumindo um processo

de Markov estacionário, teríamos a seguinte matriz de intensidades de transição:

$$Q = \begin{pmatrix} -\theta_1 & \theta_1 & 0 \\ 0 & -\theta_2 & \theta_2 \\ 0 & \theta_3 & -\theta_3 \end{pmatrix},$$

onde $\theta_i > 0$ ($i = 1, 2, 3$). Observe que $p_{13}(s, t) = p_{21}(s, t) = p_{31}(s, t) = 0$ e portanto q_{13}, q_{21} e q_{31} têm valor 0. Queremos estimar $\theta = (\theta_1, \theta_2, \theta_3)'$ para cada uma das combinações de tratamento e cidade.

A tabela 4.1 mostra os números de transições $n_{ij}(l)$ para o grupo tratamento na cidade de Oxford, e as frequências esperadas $e_{ijl} = n_{i.} \hat{p}_{ij}(y_l)$, onde $n_{i.} = \sum_{j=1}^k n_{ijl}$, que são usadas para responder questões relacionadas ao ajuste do modelo (ver seção 2.4.3).

Tabela 4.1: Número de transições (Frequência esperada)

| | (t_0, t_1) | | | | (t_1, t_2) | | | |
|---|--------------|---------|----------|-------|--------------|--------|----------|-------|
| | 1 | 2 | 3 | Total | 1 | 2 | 3 | Total |
| 1 | 93(96.0) | 3(1.7) | 2(0.3) | 98 | 89(85.7) | 2(4.2) | 2(3.1) | 93 |
| 2 | 0(0.0) | 8(12.9) | 10(5.1) | 18 | 0(0.0) | 7(4.0) | 5(8.0) | 12 |
| 3 | 0(0.0) | 1(0.5) | 8(8.50) | 9 | 0(0.0) | 5(2.8) | 15(17.2) | 20 |
| | (t_2, t_3) | | | | (t_3, t_4) | | | |
| | 1 | 2 | 3 | Total | 1 | 2 | 3 | Total |
| 1 | 83(84.9) | 3(2.9) | 3(1.2) | 89 | 76(74.5) | 3(4.3) | 4(4.3) | 83 |
| 2 | 0(0.0) | 9(6.8) | 5(7.2) | 14 | 0(0.0) | 6(3.7) | 8(10.3) | 14 |
| 3 | 0(0.0) | 2(2.3) | 20(19.7) | 22 | 0(0.0) | 0(4.3) | 28(23.7) | 28 |

As equações (4.4) e (4.5) foram utilizadas para obter (4.6) e (4.7) que por sua vez, são necessárias para o uso de (4.8) e isso é feito até que seja alcançada a convergência desejada, no nosso caso, até que $(\|\theta_1 - \theta_0\|) / \|\theta_1\| \leq 0.001$. Foi implementado um programa computacional para aplicação do algoritmo, através do módulo CM do "software" SOC, desenvolvido pela EMBRAPA. Maiores detalhes relacionados à utilização do programa são descritos no capítulo 5.

Dessa forma para esse conjunto de dados encontramos

$$\hat{Q} = \begin{pmatrix} -0.136 & 0.136 & 0 \\ 0 & -2.280 & 2.280 \\ 0 & 0.470 & -0.470 \end{pmatrix}.$$

A convergência foi bem rápida (6 iterações) e utilizamos a sugestão dada pelos autores para o “chute” inicial, ou seja, escolheu-se um dos intervalos, no nosso caso escolhemos o intervalo (t_2, t_3) e assumiu-se que cada indivíduo fez no máximo uma transição nesse intervalo. Um valor inicial para q_{ii} é obtido por $\exp(q_{ii}(t_3 - t_2)) = n_{ii}(3)/n_i(3)$. Os valores 0.20, 1.26 e 0.27 para θ_i ($i = 1, 2, 3$) obtidos dessa forma foram adequados. A matriz de covariância assintótica para $\hat{\theta}$ pode ser obtida por (4.7) e usada para dar intervalos de confiança aproximados para os θ 's. Hipóteses de interesse podem ser testadas de acordo com a seção 2.4.3 do capítulo 2, utilizando as estimativas das probabilidades de transição e da matriz de covariância assintótica.

Geralmente encontrar-se-á que modelos de Markov estacionários não são estritamente apropriados. O ajuste desses modelos, contudo, é um primeiro passo útil e frequentemente necessário para se desenvolver um modelo adequado.

EXEMPLO 4.3: Para ilustrar a utilização da metodologia para incorporar variáveis exógenas no modelo, vamos considerar os dados gerados pela simulação feita por Kalbfleisch & Lawless de um processo de Markov com três estados. A matriz de intensidades de transição é dada por:

$$Q = \begin{pmatrix} q_{11}(w) & e^{\beta_{112} + w_2 \beta_{212} + w_3 \beta_{312}} & e^{\beta_{113}} \\ e^{\beta_{121} + w_2 \beta_{221} + w_3 \beta_{321}} & q_{22}(w) & e^{\beta_{123}} \\ 0 & 0 & 0 \end{pmatrix}$$

que é uma extensão do Exemplo 4.1 desta seção. Note que assumiu-se que o fato do indivíduo sair do estudo, representado na 3ª coluna da matriz Q , não depende das variáveis exógenas envolvidas na simulação, o que parece ser bastante razoável. Para a simulação, os parâmetros foram $\beta_{112} = \beta_{113} = \beta_{123} = -2.30$, $\beta_{121} = -1.90$, $\beta_{212} = \beta_{221} = \beta_{321} = 0.50$ e $\beta_{312} = -0.50$. Foi assumido que 15 indivíduos iniciaram nos estados 1 e 2 no tempo $t_0 = 0$, para cada um dos quatro grupos com vetores de regressão $(w_1, w_2, w_3) = (1, 0, 0)$, $(1, 0, 1)$, $(1, 1, 0)$ e $(1, 1, 1)$ respectivamente. Assume-se ainda que as observações foram feitas em 6 ocasiões t_0, \dots, t_5 . Os dados simulados estão apresentados na tabela 4.2 a seguir:

Tabela 4.2: Dados gerados pela simulação

| w_h | | (t_0, t_1) | | | (t_1, t_2) | | | (t_2, t_3) | | | (t_3, t_4) | | | (t_4, t_5) | | |
|---------|---|--------------|----|---|--------------|---|---|--------------|---|---|--------------|---|---|--------------|----|----|
| | | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| (1,0,0) | 1 | 13 | 2 | 0 | 14 | 1 | 2 | 12 | 2 | 1 | 11 | 0 | 2 | 11 | 12 | 0 |
| | 2 | 4 | 8 | 3 | 1 | 9 | 0 | 1 | 6 | 3 | 1 | 7 | 0 | 0 | 6 | 1 |
| | 3 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 5 | 0 | 0 | 9 | 0 | 0 | 11 |
| (1,0,1) | 1 | 14 | 1 | 0 | 16 | 0 | 1 | 19 | 0 | 0 | 17 | 2 | 1 | 16 | 1 | 2 |
| | 2 | 3 | 11 | 1 | 3 | 8 | 1 | 1 | 5 | 2 | 2 | 3 | 0 | 3 | 2 | 0 |
| | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 3 | 0 | 0 | 5 | 0 | 0 | 6 |
| (1,1,0) | 1 | 11 | 4 | 0 | 10 | 2 | 2 | 10 | 2 | 3 | 10 | 2 | 2 | 10 | 1 | 0 |
| | 2 | 3 | 11 | 1 | 5 | 9 | 1 | 4 | 7 | 0 | 1 | 6 | 2 | 3 | 5 | 0 |
| | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 4 | 0 | 0 | 7 | 0 | 0 | 11 |
| (1,1,1) | 1 | 11 | 3 | 1 | 12 | 1 | 1 | 17 | 1 | 0 | 17 | 0 | 1 | 13 | 1 | 3 |
| | 2 | 3 | 11 | 1 | 6 | 6 | 2 | 1 | 5 | 1 | 0 | 6 | 0 | 1 | 5 | 0 |
| | 3 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 5 | 0 | 0 | 6 | 0 | 0 | 7 |

O parâmetro a ser estimado é $\theta = (\beta_{112}, \beta_{212}, \beta_{312}, \beta_{113}, \beta_{121}, \beta_{221}, \beta_{321}, \beta_{123})'$. Por exemplo, para $h = 2$ temos $w_2 = (1, 0, 1)$ e a matriz de intensidades correspondente é:

$$Q(w_2) = \begin{pmatrix} q_{11}(w_2) & e^{\beta_{112} + \beta_{312}} & e^{\beta_{113}} \\ e^{\beta_{121} + \beta_{321}} & q_{22}(w_2) & e^{\beta_{123}} \\ 0 & 0 & 0 \end{pmatrix},$$

onde $q_{11}(w_2) = -(e^{\beta_{112} + \beta_{312}} + e^{\beta_{113}})$ e $q_{22}(w_2) = -(e^{\beta_{121} + \beta_{321}} + e^{\beta_{123}})$. Temos ainda que a derivada

$$\frac{\partial Q(w_2)}{\partial \beta_{121}} = \begin{pmatrix} 0 & 0 & 0 \\ e^{\beta_{121} + \beta_{321}} & -e^{\beta_{121} + \beta_{321}} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

é calculada da mesma maneira, ou seja, conforme a equação (4.13), para todos os outros parâmetros. Para cada $Q(w_h)$ é obtido $s^{(h)}(\theta)$ e $M^{(h)}(\theta)$. Temos então $s(\theta) = \sum_{h=1}^4 s^{(h)}(\theta)$ e $M(\theta) = \sum_{h=1}^4 M^{(h)}(\theta)$. Utilizando (4.8) obtemos estimativas para θ . Para o “chute” inicial utilizamos um critério semelhante ao utilizado no Exemplo 4.2. Partindo do vetor inicial $\theta_0 = (0.17, 1.24, -2.08, -2.82, -1.27, 2.03, -0.24, 0.52)'$ foi alcançada a convergência após 8 iterações e as estimativas de máxima verossimilhança e seus respectivos erros padrão são dados na tabela 4.3 a seguir:

Tabela 4.3: Estimativas dos parâmetros e respectivos erros padrão

| Parâmetro | Valor Verdadeiro | Valor Estimado | Erro Padrão Estimado |
|---------------|------------------|----------------|----------------------|
| β_{112} | -2.30 | -2.18 | 0.356 |
| β_{212} | 0.50 | 0.70 | 0.406 |
| β_{312} | -0.50 | -0.77 | 0.406 |
| β_{113} | -2.30 | -2.66 | 0.235 |
| β_{121} | -1.90 | -1.39 | 0.278 |
| β_{221} | 0.50 | 0.28 | 0.307 |
| β_{321} | 0.50 | 0.11 | 0.304 |
| β_{123} | -2.30 | -2.25 | 0.254 |

Pode-se observar que o resultado foi bastante satisfatório. Esses resultados permitem expressar de forma simples, intervalos de confiança para os parâmetros. Teste estatísticos para adequabilidade do ajuste bem como outros testes de interesse podem ser construídos com base na razão de verossimilhança ou na estatística qui-quadrado, conforme descrito na seção 2.4.3 do capítulo 2.

Para o caso de incorporação de variáveis exógenas, também foi implementado um programa computacional cuja utilização é discutida mais detalhadamente no capítulo 5 .

Capítulo 5

Apresentação dos Programas Computacionais Desenvolvidos

5.1 Introdução

Neste capítulo apresentamos os programas computacionais desenvolvidos para análise de dados longitudinais com respostas categorizadas seguindo a metodologia apresentada no capítulo anterior. Para esta apresentação foi utilizado um conjunto de dados extraído de Landis *et al.* (1988) o qual é mostrado na tabela do apêndice C. Esses dados são provenientes de um experimento realizado com o objetivo de testar a eficácia de uma droga na condição da pele do indivíduo. Os pacientes foram aleatoriamente escolhidos de modo que alguns receberam o medicamento e outros não tiveram qualquer tipo de tratamento especial. O grupo que recebeu o medicamento foi denominado grupo tratamento e o outro, grupo placebo. Antes da aplicação do tratamento foi determinada a condição inicial da pele de cada indivíduo em estudo. Subsequentemente, o estado da pele foi avaliado em 3 ocasiões diferentes. No estudo original as respostas são compostas por 5 categorias. No nosso caso, reduzimos para 3 categorias a fim de facilitar a compreensão dos resultados obtidos e elas foram definidas da seguinte forma: 1- melhorando, 2- sem alteração e 3- piorando.

O número de transições ocorridas em cada um dos dois intervalos de tempo está na tabela 5.1 a seguir:

Tabela 5.1: Número de transições entre categorias para cada intervalo de tempo

| | | 1 | 2 | 3 |
|---------------------|----------|----------|----------|----------|
| 1º Intervalo | 1 | 64 | 2 | 0 |
| | 2 | 11 | 18 | 4 |
| | 3 | 2 | 16 | 18 |
| 2º Intervalo | 1 | 73 | 3 | 1 |
| | 2 | 1 | 29 | 6 |
| | 3 | 0 | 2 | 20 |

Podemos ver nessa tabela, por exemplo, que 11 pacientes que estavam na categoria 2 no tempo 1, passaram para a categoria 1 no tempo 2, enquanto que apenas 1 paciente foi da categoria 2 para a categoria 1 no segundo intervalo de tempo.

Na seção 5.2 vamos descrever os programas e mostrar como os dados citados podem ser analisados através deles e na seção 5.3 faremos algumas considerações finais a respeito dos programas bem como propostas para trabalhos futuros.

5.2 Programas Computacionais Desenvolvidos

Foram desenvolvidos dois programas, um que denominamos KALBLAW1, que estima as intensidades de transição baseado na metodologia descrita na seção 4.2 e um outro denominado KALBLAW2 que modela as intensidades de transição pelas variáveis exógenas envolvidas no estudo, e foi baseado na metodologia descrita na seção 4.3. Ambos os programas fornecem as matrizes de probabilidades de transição estacionárias e respectivas variâncias assintóticas, obtidas a partir da matriz de intensidades de transição, bem como as probabilidades marginais para cada ocasião. Esses programas foram implementados na forma de “macros” utilizando o módulo CM do SOC, “software” estatístico desenvolvido pelo NTIA/EMBRAPA. Para maiores detalhes sobre o SOC ver NTIA/EMBRAPA (1989).

Nos dois programas a entrada dos dados pode ser feita de duas maneiras. Ou na forma individual, isto é, cada indivíduo com as suas respostas em cada ocasião (veja exemplo na tabela 3.1), ou na forma resumida, isto é, como contagens do número de transições ocorridas entre os tempos de observação (ver tabela 5.1). No exemplo que discutiremos neste capítulo, os dados estão na forma individual conforme mostrado na tabela do apêndice C.

Vamos primeiramente descrever a utilização de KALBLAW1. A matriz de dados para o caso individual deve ser inicializada com a denominação Y. Neste exemplo Y consiste das colunas denotadas por 1, 2 e 3 da tabela do apêndice C. Esse programa é utilizado para estudar as intensidades de transição para cada grupo em separado, neste exemplo os grupos placebo e tratamento. Assim devemos usá-lo duas vezes, uma para cada um dos grupos. Dessa forma, para cada conjunto é necessário uma matriz Y. Para o grupo placebo, Y tem dimensão 61×3 que corresponde aos indivíduos que têm valor 0 na coluna denotada por T no apêndice C. Para o grupo tratamento, a dimensão de Y é 74×3 . Para a forma resumida a matriz Y não é necessária. Nessa situação, a matriz que tem que ser inicializada é denominada N e representa as contagens do número de transições ocorridas em cada intervalo de tempo. Quando os dados estão na forma individual essa matriz é gerada automaticamente pelo programa. Para este exemplo, os valores que formam a matriz N têm a mesma disposição da tabela 5.1, ou seja, 2 matrizes 3×3 dispostas uma embaixo da outra gerando no final uma matriz 6×3 . Outros valores devem ser inicializados, porém a forma de fazê-lo é direta bastando para isso digitá-los no local apropriado. Esses locais estão indicados no programa. Um exemplo é o vetor de parâmetros iniciais denominado TETA0.

Para a execução correta do programa é fundamental a definição da matriz de intensidades de transição que se quer estimar. No nosso exemplo ela foi definida da seguinte forma:

$$Q = \begin{pmatrix} -\theta_1 & \theta_1 & 0 \\ \theta_2 & -(\theta_2 + \theta_3) & \theta_3 \\ 0 & \theta_4 & -\theta_4 \end{pmatrix},$$

onde $\theta_1 = q_{12}$, $\theta_2 = q_{21}$, $\theta_3 = q_{23}$ e $\theta_4 = q_{32}$. A suposição básica é que um paciente que estava melhorando não vai em um espaço de tempo muito pequeno passar para o estado piorando e vice-versa, conseqüentemente $q_{13} = q_{31} = 0$. Observe porém que, conforme citado na seção 2.3, isso não implica que as probabilidades de transição da categoria 1 para 3 e de 3 para 1 sejam nulas. Em outras palavras estamos assumindo que apesar de ser possível que um indivíduo melhorando (piorando) em uma determinada ocasião, esteja piorando (melhorando) na ocasião imediatamente posterior, isso não ocorrerá instantaneamente. Considerando a matriz de intensidades de transição exposta anteriormente temos $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)'$ para cada grupo, ou seja, um total de 8 parâmetros.

Não existe uma forma padrão para a matriz de intensidades de transição, ou seja, cada exemplo tem um tipo particular de matriz de intensidades. Neste exemplo foi feita a suposição que $q_{13} = q_{31} = 0$. Se essa suposição não fosse razoável, teríamos uma outra matriz

de intensidades. Portanto é essencial que para cada problema seja especificada adequadamente a forma definida para essa matriz. Em KALBLAW1 a matriz de intensidades é denotada por Q . No programa existe uma indicação de onde e como especificá-la corretamente utilizando o vetor dos parâmetros. Para que o programa possa ser executado adequadamente, a matriz $DQ = (\partial Q / \partial \theta_u)$, utilizada para calcular $(\partial P(t) / \partial \theta_u)$ conforme descrito na equação (4.5), deve ser inicializada. A matriz DQ tem bk linhas e k colunas onde b é o número de parâmetros em cada execução do programa e k o número de categorias, que corresponde a b matrizes $k \times k$ dispostas uma embaixo da outra. Para o exemplo temos, $b = 4$, $k = 3$,

$$\frac{\partial Q}{\partial \theta_1} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \frac{\partial Q}{\partial \theta_2} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\frac{\partial Q}{\partial \theta_3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{e} \quad \frac{\partial Q}{\partial \theta_4} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

Consequentemente

$$DQ = \begin{pmatrix} -1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

Após a inicialização de todos os valores necessários, estamos em condição de executar o programa e obter os resultados de interesse. Com a execução de KALBLAW1 obtivemos as seguintes estimativas para as matrizes de intensidades de transição:

$$\hat{Q}_p = \begin{pmatrix} -0.193 & 0.193 & 0 \\ 0.022 & -0.273 & 0.251 \\ 0 & 0.519 & -0.519 \end{pmatrix} \quad \text{e} \quad \hat{Q}_t = \begin{pmatrix} -0.012 & 0.012 & 0 \\ 0.717 & -0.896 & 0.179 \\ 0 & 0.163 & -0.163 \end{pmatrix},$$

para os grupos placebo e tratamento respectivamente. A partir desses valores, algumas características do processo podem ser estudadas. Vimos por exemplo no capítulo 2, que $-q_{ii}(t)^{-1}$

é o tempo médio de permanência na categoria i . Podemos observar neste exemplo, que o tempo médio estimado de permanência na categoria 1 (melhorando) no grupo tratamento é 83, enquanto que no grupo placebo é 5, ou seja, em termos relativos o paciente vai permanecer melhorando em um período de tempo aproximadamente 16 vezes maior se receber o medicamento. Outras características baseadas nas intensidades de transição podem ser estudadas mais detalhadamente conforme o interesse do pesquisador. O programa ainda fornece a estimativa da matriz de covariância assintótica dos estimadores. Não temos evidência de que o modelo de Markov estacionário de primeira ordem não seja adequado para essa situação. O teste para a adequabilidade do ajuste baseado na subseção que trata de teste de hipóteses sobre os parâmetros descrita no capítulo 2, é fornecido automaticamente pelo programa. Neste caso, obtivemos como estatísticas de teste e respectivos níveis descritivos $\phi = 16.041$ (0.040) e $\lambda = 13.871$ (0.085) para o grupo placebo e $\phi = 8.086$ (0.424) e $\lambda = 8.895$ (0.351) para o grupo tratamento, ambos com 8 graus de liberdade. Pode-se perceber que o ajuste para o grupo tratamento foi um pouco melhor. A partir das estimativas obtidas para as intensidades de transição o programa calculou as seguintes matrizes de probabilidades de transição:

$$\hat{P}_p(t) = \begin{pmatrix} 0.826 & 0.156 & 0.018 \\ 0.018 & 0.809 & 0.173 \\ 0.004 & 0.358 & 0.638 \end{pmatrix} \text{ e } \hat{P}_t(t) = \begin{pmatrix} 0.991 & 0.008 & 0.001 \\ 0.473 & 0.418 & 0.109 \\ 0.042 & 0.099 & 0.859 \end{pmatrix},$$

para os grupos placebo e tratamento respectivamente. Devemos notar que neste exemplo consideramos os tempos de observação como sendo igualmente espaçados e portanto, todos os intervalos de tempo têm o mesmo tamanho. Como consequência foi gerada apenas uma matriz de probabilidades de transição para cada grupo, caso contrário, para cada intervalo de tempo de tamanho diferente, uma matriz de probabilidades de transição seria gerada. Pode-se observar através dessas matrizes que a probabilidade de estar melhorando no tempo seguinte dado que o paciente estava melhorando no tempo anterior é bastante alta, quase 1, no grupo tratamento, dando um indicativo de que o medicamento está sendo eficiente na melhoria da condição da pele do indivíduo. Porém, curiosamente a probabilidade de continuar piorando dado que estava piorando é mais alta no grupo tratamento. Esse tipo de questão pode ser melhor interpretada com o auxílio das probabilidades marginais. O programa fornece essas probabilidades diretamente utilizando a equação (2.3) dada no capítulo 2. As estimativas das probabilidades marginais calculadas pelo programa estão na tabela 5.2 a seguir:

Tabela 5.2: Estimativas para as probabilidades marginais

| | PLACEBO | | | TRATAMENTO | | |
|-------------------|---------|---------|---------|------------|---------|---------|
| | $t = 1$ | $t = 2$ | $t = 3$ | $t = 1$ | $t = 2$ | $t = 3$ |
| $\Pr\{X(t) = 1\}$ | 0.246 | 0.209 | 0.182 | 0.689 | 0.812 | 0.866 |
| $\Pr\{X(t) = 2\}$ | 0.229 | 0.412 | 0.501 | 0.270 | 0.123 | 0.064 |
| $\Pr\{X(t) = 3\}$ | 0.525 | 0.379 | 0.317 | 0.041 | 0.065 | 0.070 |

Podemos notar nessa tabela que a probabilidade de estar piorando ($\Pr\{X(t) = 3\}$) é muito pequena no grupo tratamento o que reforça a idéia da eficácia do tratamento. Pode ser visto ainda que as probabilidades de estar melhorando são sempre maiores no grupo tratamento e aumentam com o tempo. Essas observações parecem confirmar que realmente a aplicação do medicamento é satisfatória na melhoria da condição da pele.

Vamos considerar o mesmo exemplo, só que utilizando o programa KALBLAW2 desenvolvido com o objetivo de modelar as intensidades de transição conforme as variáveis exógenas de interesse. Sua utilização é muito semelhante à de KALBLAW1. Na forma individual a matriz Y deve ter suas primeiras colunas reservadas para as variáveis exógenas. Vamos trabalhar por enquanto apenas com a variável exógena tratamento que corresponde à primeira coluna de Y . Desse modo a matriz Y completa é formada por todas as linhas e as colunas denotadas por **T**, **1**, **2** e **3** da tabela do apêndice C. Para a forma resumida, as contagens devem ser feitas em cada grupo separadamente. A matriz Y nesse caso não é necessária. A matriz a ser inicializada é a matriz de contagens denotada por $N_{.h}$. Para o exemplo, considerando apenas a variável exógena tratamento, a matriz a ser inicializada para utilização de KALBLAW2 com entrada de dados na forma resumida seria:

$$N_{.h} = \begin{pmatrix} 51 & 0 & 0 \\ 11 & 6 & 2 \\ 1 & 1 & 2 \\ 62 & 1 & 0 \\ 1 & 5 & 1 \\ 0 & 0 & 4 \\ 13 & 2 & 0 \\ 0 & 12 & 2 \\ 1 & 15 & 16 \\ 11 & 2 & 1 \\ 0 & 24 & 5 \\ 0 & 2 & 16 \end{pmatrix}$$

Essa matriz é composta por 4 matrizes 3×3 onde a primeira matriz corresponde às contagens no primeiro grupo no primeiro intervalo, a segunda às contagens no primeiro grupo no segundo intervalo e assim sucessivamente até a quarta matriz.

As matrizes de intensidades de transição são modeladas conforme as equações (4.9) de modo que

$$\begin{aligned}\theta_1(\mathbf{w}) &= e^{w_1\beta_{112}+w_2\beta_{212}}, & \theta_2(\mathbf{w}) &= e^{w_1\beta_{121}+w_2\beta_{221}}, \\ \theta_3(\mathbf{w}) &= e^{w_1\beta_{123}+w_2\beta_{223}} & \text{e } \theta_4(\mathbf{w}) &= e^{w_1\beta_{132}+w_2\beta_{232}},\end{aligned}$$

onde $w_1 = 1$ e w_2 assume valor 0 para o grupo placebo e 1 para o grupo tratamento. Temos então

$$\beta = (\beta_{112}, \beta_{121}, \beta_{123}, \beta_{132}, \beta_{212}, \beta_{221}, \beta_{223}, \beta_{232})'.$$

Observe que temos o mesmo número de parâmetros do caso anterior, entretanto o programa só precisa ser executado uma vez com todos os dados juntos. É de se esperar que os resultados obtidos com a execução de KALBLAW1 para cada grupo e KALBLAW2 sejam os mesmos, uma vez que estamos estimando exatamente os mesmos parâmetros utilizando as variáveis exógenas de forma diferente em cada um deles.

Considerando todas as combinações dos valores da variável exógena (0 e 1), categoria anterior (1, 2 e 3) e categoria posterior (1, 2 e 3), o vetor completo dos parâmetros a serem estimados neste exemplo seria composto por 18 elementos:

$$(\beta_{111}, \beta_{112}, \beta_{113}, \beta_{121}, \beta_{122}, \beta_{123}, \beta_{131}, \beta_{132}, \beta_{133}, \beta_{211}, \beta_{212}, \beta_{213}, \beta_{221}, \beta_{222}, \beta_{223}, \beta_{231}, \beta_{232}, \beta_{233})'.$$

Entretanto, como citado anteriormente, apenas 8 parâmetros são necessários para a modelagem sugerida. O programa KALBLAW2 exige a criação de um vetor denominado AUX que indica quais parâmetros estão envolvidos no modelo. Esse vetor é composto por 0's e 1's de modo que os parâmetros a serem estimados recebem o valor 1 e os outros recebem valor 0. O número de elementos nesse vetor, que tem a mesma dimensão do vetor com todos os parâmetros, é sk^2 , onde s é o número de variáveis exógenas e k o número de categorias. A ordem desses valores em AUX é determinada pelas combinações dos índices dos parâmetros. Para o exemplo que estamos discutindo, o vetor AUX apropriado para indicar β é

$$\text{AUX} = (0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0),$$

indicando que os parâmetros a serem estimados são: $\beta_{112}, \beta_{121}, \beta_{123}, \beta_{132}, \beta_{212}, \beta_{221}, \beta_{223}$ e β_{232} (observe os índices dos elementos no vetor completo dos parâmetros).

Em KALBLAW2 a matriz de intensidades de transição é denotada por Q_h . Deve ser observado que para cada grupo formado conforme os valores assumidos pelas variáveis exógenas, uma diferente matriz Q_h é gerada. No programa é mostrada a forma de inicializá-la corretamente. Em KALBLAW2, a matriz das derivadas $(\partial Q/\partial \theta_u)$ não precisa ser fornecida pelo usuário pois ela é calculada automaticamente para cada vetor de variáveis exógenas.

As estimativas para os parâmetros e seus respectivos erros padrão, obtidos com a execução do programa são dadas na tabela 5.3 a seguir:

Tabela 5.3: Estimativas para os parâmetros e erros padrão

| Parâmetro | Estimativa | Erro Padrão |
|---------------|------------|-------------|
| β_{112} | -1.645 | 0.447 |
| β_{212} | -2.755 | 1.094 |
| β_{121} | -3.794 | 1.005 |
| β_{221} | 3.462 | 1.045 |
| β_{123} | -1.382 | 0.379 |
| β_{223} | -0.334 | 0.704 |
| β_{132} | -0.657 | 0.257 |
| β_{232} | -1.149 | 1.048 |

Substituindo os valores correspondentes em w_2 , isto é, 0 para o grupo placebo e 1 para o grupo tratamento, em β obtemos os mesmos resultados para Q e P obtidos anteriormente, que como foi mencionado, era de se esperar. O teste de adequabilidade do ajuste fornecido pelo programa não acusou evidência muito forte de que o modelo de Markov estacionário de primeira ordem não seja adequado. As estatísticas para o teste são $\phi = 24.098$ (0.09) e $\lambda = 22.766$ (0.12) com 16 graus de liberdade. Note que esses valores são nada mais do que a soma dos valores obtidos para cada grupo em separado da análise anterior.

Como foi mencionado no início desta seção, cada paciente teve o estado da pele classificado antes da aplicação do tratamento. Os valores estão apresentados na coluna denotada por **0** na tabela do apêndice C. Como os pacientes tinham diferentes estados de pele no início, esta informação adicional foi incorporada à análise na forma de blocos. Bloco 1 formado pelos pacientes com condição inicial 2 (sem alteração) e bloco 2 formado pelos pacientes com condição inicial 3 (piorando). A matriz Y é composta por todas as linhas e as colunas denotadas por **T**, **0**, **1**, **2** e **3** da tabela do apêndice C, com os valores da coluna denotada por **0** subtraídos de 2,

Tabela 5.4: Estimativas para os parâmetros e erros padrão levando-se em consideração os blocos

| Parâmetro | Estimativa | Erro Padrão |
|---------------|------------|-------------|
| β_{112} | -1.443 | 0.519 |
| β_{212} | -2.681 | 1.106 |
| β_{121} | -3.513 | 0.999 |
| β_{221} | 3.416 | 1.041 |
| β_{123} | -1.024 | 0.375 |
| β_{223} | -0.443 | 0.708 |
| β_{132} | -0.344 | 0.261 |
| β_{232} | -1.181 | 1.019 |
| β_3 | -0.551 | 0.874 |

Substituindo os valores para as variáveis exógenas w_2 e w_3 , tratamento e bloco respectivamente, e utilizando as estimativas da tabela 5.4, obtivemos as matrizes de intensidade e probabilidade de transição estimadas para cada grupo formado que são as seguintes:

- 1) Grupo placebo, inicialmente sem alteração

$$\hat{Q} = \begin{pmatrix} -0.236 & 0.236 & 0 \\ 0.030 & -0.389 & 0.359 \\ 0 & 0.709 & -0.709 \end{pmatrix} \text{ e } \hat{P}(t) = \begin{pmatrix} 0.792 & 0.180 & 0.028 \\ 0.023 & 0.759 & 0.218 \\ 0.007 & 0.430 & 0.562 \end{pmatrix}.$$

- 2) Grupo placebo, inicialmente piorando

$$\hat{Q} = \begin{pmatrix} -0.136 & 0.136 & 0 \\ 0.017 & -0.224 & 0.207 \\ 0 & 0.409 & -0.409 \end{pmatrix} \text{ e } \hat{P}(t) = \begin{pmatrix} 0.873 & 0.115 & 0.011 \\ 0.015 & 0.831 & 0.154 \\ 0.003 & 0.304 & 0.693 \end{pmatrix}.$$

- 3) Grupo tratamento, inicialmente sem alteração

$$\hat{Q} = \begin{pmatrix} -0.016 & 0.016 & 0 \\ 0.907 & -1.137 & 0.230 \\ 0 & 0.218 & -0.218 \end{pmatrix} \text{ e } \hat{P}(t) = \begin{pmatrix} 0.989 & 0.010 & 0.001 \\ 0.544 & 0.333 & 0.123 \\ 0.066 & 0.116 & 0.818 \end{pmatrix}.$$

- 4) Grupo tratamento, inicialmente piorando

$$\hat{Q} = \begin{pmatrix} -0.009 & 0.009 & 0 \\ 0.523 & -0.656 & 0.133 \\ 0 & 0.125 & -0.125 \end{pmatrix} \text{ e } \hat{P}(t) = \begin{pmatrix} 0.993 & 0.007 & 0.000 \\ 0.385 & 0.523 & 0.092 \\ 0.026 & 0.087 & 0.887 \end{pmatrix}.$$

Comparações entre os grupos são feitas baseando-se nessas matrizes e nas probabilidades marginais obtidas conforme citado anteriormente, cujos valores são fornecidos pelo programa automaticamente e estão mostrados na tabela 5.5 a seguir:

Tabela 5.5: Estimativas para as probabilidades marginais levando-se em consideração os blocos

| GRUPO | CONDIÇÃO INICIAL | TEMPO | $\Pr\{X(t) = 1\}$ | $\Pr\{X(t) = 2\}$ | $\Pr\{X(t) = 3\}$ |
|------------|------------------|-------|-------------------|-------------------|-------------------|
| Placebo | 2 | 1 | 0.310 | 0.207 | 0.483 |
| | 2 | 2 | 0.254 | 0.421 | 0.325 |
| | 2 | 3 | 0.213 | 0.505 | 0.282 |
| Placebo | 3 | 1 | 0.188 | 0.250 | 0.562 |
| | 3 | 2 | 0.169 | 0.401 | 0.430 |
| | 3 | 3 | 0.155 | 0.483 | 0.362 |
| Tratamento | 2 | 1 | 0.647 | 0.294 | 0.059 |
| | 2 | 2 | 0.804 | 0.111 | 0.085 |
| | 2 | 3 | 0.861 | 0.055 | 0.084 |
| Tratamento | 3 | 1 | 0.725 | 0.250 | 0.025 |
| | 3 | 2 | 0.817 | 0.138 | 0.045 |
| | 3 | 3 | 0.865 | 0.082 | 0.053 |

Podemos observar que a probabilidade de estar melhorando na primeira ocasião após o início do estudo é bem alta e a probabilidade de estar piorando na mesma ocasião é bem baixa, ambas para o grupo tratamento. Analisando as probabilidades marginais em conjunto com as probabilidades de transição, podemos concluir que o medicamento é bastante eficaz, uma vez que pacientes que o receberam têm uma probabilidade alta de melhorar ainda no primeiro intervalo de tempo e probabilidades muito altas (praticamente 1) de se manterem melhorando. Isso não ocorre no grupo placebo. Sem a aplicação do medicamento é mais provável que no primeiro intervalo haja uma piora na condição da pele do paciente. Note ainda, que para o grupo tratamento a probabilidade de estar piorando dado que estava piorando no tempo anterior é alta, indicando que apesar de pouco provável, se o paciente piorar após a aplicação do medicamento, muito provavelmente ele continuará piorando, dando um indicativo de que é melhor suspender o tratamento, pois ele não adianta para esse caso.

A incorporação da condição inicial melhorou o ajuste do modelo. Como estatísticas para o teste de adequabilidade do ajuste fornecido pelo programa foram obtidas $\phi = 42.632$ (0.318) e $\lambda = 38.308$ (0.501) com 39 graus de liberdade.

Deve-se observar que nesse caso se considerarmos cada grupo em separado vamos obter resultados diferentes devido a suposição de igualdade do efeito de blocos para todas as intensidades de transição.

Deve-se destacar que para a implementação da metodologia de Kalbfleisch & Lawless é imprescindível a decomposição da matriz de intensidades de transição para a obtenção de seus autovalores e autovetores. Em geral, os “softwares” estatísticos disponíveis não fazem essa decomposição para matrizes não simétricas, que é o nosso caso em particular. A partir de algoritmos descritos em Smith *et al.* (1976), foi desenvolvida uma função com essa finalidade no módulo CM do SOC, de modo que através dessa função qualquer matriz real pode ser decomposta em seus autovalores e respectivos autovetores. Outra função criada especificamente é a função “contagem”, que como o próprio nome diz, conta o número de transições ocorridas no intervalo de tempo indicado. No exemplo, as colunas de Y indicam os tempos de observação. Caso sejam indicadas as colunas 1 e 2, será feita a contagem para o primeiro intervalo e se as colunas indicadas são as colunas 2 e 3 faz-se a contagem para o segundo intervalo. Essas funções ainda não estão implementadas na última versão do CM (1989).

Outro fato que merece destaque está relacionado ao algoritmo utilizado. O método “Scoring” de Fisher não garante que a função de verossimilhança seja aumentada em cada iteração do processo. Jennrich & Schluchter (1986) sugerem que em cada iteração seja verificado se houve efetivamente um aumento da função. Caso isso não tenha ocorrido, os autores recomendam a divisão do incremento por potências de 2 até que ocorra um aumento na função. Passa-se então para a próxima iteração. Utilizamos em KALBLAW1 e KALBLAW2 o incremento dividido por 4 e obtivemos bons resultados, apesar do número de iterações ter crescido consideravelmente. Os interessados em obter uma cópia dos programas e respectivos manuais devem enviar um disquete de 5 $\frac{1}{4}$ ” para:

Antônio de Queiroz Noletto
Rod. D. Pedro I (SP65) - Km 143,6
NTIA/EMBRAPA
CEP: 13081
Campinas, SP

5.3 Conclusões Finais e Sugestões para Trabalhos Futuros

Como pudemos constatar através da utilização de um exemplo para descrever os programas desenvolvidos, a metodologia de Kalbfleisch & Lawless pode ser utilizada satisfatoriamente no estudo das transições quando temos dados obtidos longitudinalmente e respostas

categorizadas. Uma análise mais detalhada deve ser feita em conjunto com o pesquisador da área da qual os dados foram obtidos.

Foram utilizados nesse exemplo observações completas, isto é, aquelas onde todos os indivíduos foram observados nas 3 ocasiões determinadas para o estudo. No estado atual os programas não estão aptos a trabalhar com valores perdidos, porém já estão sendo feitas adaptações para tornar possível a inclusão desse tipo de dados. Ainda com respeito aos programas desenvolvidos, a preocupação inicial foi ter os resultados de uma forma simples de se trabalhar, porém a sua execução de uma forma geral está muito lenta. Uma idéia é tentar torná-los mais eficientes.

Temos ainda a intenção de dar prosseguimento a esse trabalho de duas formas: implementação de programas para o caso de tempo discreto e desenvolvimento de metodologias para respostas contínuas. No caso em tempo discreto, um programa computacional desenvolvido por Dall'Agnol(1990) para o modelo marginal pode ser adaptado e utilizado para modelos de transição. No entanto para respostas contínuas, a teoria de processos estocásticos é muito complexa e requer um nível de conhecimento teórico que está além dos objetivos deste trabalho.

Bibliografia

- [1] Albert, A. (1962) - "Estimating the Infinitesimal Generator of a Continuous Time, Finite State Markov Process", *Annals of Mathematical Statistics*, 33, 727-753
- [2] Anderson, T. W. & Goodman, L. A. (1957) - "Statistical Inference about Markov Chains", *Annals of Mathematical Statistics*, 28, 89-110
- [3] Andrade, D. F. & Singer, J. M. (1986) - *Análise de Dados Longitudinais*, VII SINAPE, UNICAMP, Campinas, SP.
- [4] Bartholomew, D. J. (1982) - *Stochastic Models for Social Processes*, Jonh Wiley: London
- [5] Basawa, I. V. & Prakasa Rao, B. L. S. (1980) - *Statistical Inference for Stochastic Processes*, Academic Press
- [6] Billingsley, P. (1961a) - *Statistical Inference for Markov Processes*, The University of Chicago Press
- [7] Billingsley, P. (1961b) - "Statistical Methods in Markov Chains", *Annals of Mathematical Statistics*, 32, 12-40
- [8] Bishop, Y. M. M.; Fienberg, S. E. & Holland, P. W. (1980) - *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press
- [9] Coddington, E. A. & Levinson, N. (1955) - *Theory of Ordinary Differential Equations*, Mc Graw-Hill: New York

- [10] Cook, N. R. & Ware, J. H. (1983) - "Design and Analysis Methods for Longitudinal Research", *Annual Reviews of Public Health*, 4, 1-23
- [11] Cox, D. R. (1970) - *Analysis of Binary Data*, Methuen and Co.: London
- [12] Cox, D. R. & Miller, H. D. (1977) - *The Theory of Stochastic Processes*, Chapman and Hall: London
- [13] Dall'Agnol, I. (1990) - *Alguns Aspectos da Análise em Estudos Longitudinais com Respostas Categorizadas*, Tese de Mestrado, UNICAMP, Campinas, SP.
- [14] Goodman, L. A. (1962) - "Statistical Methods for Analysing Processes of Change", *American Journal of Sociology*, 68, 57-78
- [15] Hoel, P. G.; Port, S. C. & Stone, C. J. (1972) - *Introduction to Stochastic Processes*, Houghton Mifflin Company: Boston
- [16] Kalbfleisch, J. D. & Lawless, J. F. (1985) - "The Analysis of Panel Data under a Markov Assumption", *Journal of American Statistical Association*, 80, 863-871
- [17] Jennrich, R. I. & Schluchter, M. D. (1986) - "Unbalanced Repeated-Measures Models with Structured Covariance Matrices", *Biometrics*, 42, 805-820
- [18] Karlin, S. & Taylor, H. M. (1975) - *A First Course in Stochastic Processes*, Academic Press: New York
- [19] Koch, G. G.; Landis, J. R.; Freeman, J. L.; Freeman, D. H. & Lehnen, R. G. (1977) - "A General Methodology for the Analysis of Experiments with Repeated Measurements of Categorical Data", *Biometrics*, 33, 133-158
- [20] Koch, G. G.; Singer, J. M.; Carr, G. J.; Cohen, S. B. & Stokes, M. E. (1986) - "Some Aspects of Weighted Least Squares Analysis for Longitudinal Categorical Data", *Proceedings of the Workshop on Longitudinal Methods in Health Research*, Forthcoming

- [21] Korn, E. L. & Whittemore, A. S. (1979) - "Methods for Analysing Panel Studies of Acute Health Effects of Air Pollution", *Biometrics*, 35, 795-802
- [22] Kowalski, C. J. & Guire, K. E. (1974) - "Longitudinal Data Analysis", *Growth*, 38, 131-169
- [23] Landis, J. R.; Miller, M. E.; Davis, C. S. & Koch, G. G. (1988) - "Some General Methods for the Analysis of Categorical Data in Longitudinal Studies", *Statistics in Medicine*, 7, 109-137
- [24] Leite, J. G. & Singer, J. M. (1990) - *Métodos Assintóticos em Estatística - Fundamentos e Aplicações*, IX SINAPE, USP, São Paulo, SP.
- [25] McCullagh, P. (1980) - "Regression Methods for Ordinal Data", *Journal of the Royal Statistical Society, Series B*, 42, 109-142
- [26] Muenz, L. R. & Rubinstein, L. V. (1985) - "Markov Models for Covariate Dependence of Binary Sequences", *Biometrics*, 41, 91-101
- [27] NTIA/EMBRAPA (1989) - *SOC, Manual do Usuário*, Campinas, SP.
- [28] Phelan, M. J. (1988) - "Inference from Censored Markov Chains with Applications to Multiwave Panel Data", *Stochastic Processes and their Applications*, 29, 85-102
- [29] Plewis, I. (1985) - *Analysing Change: Measurement and Explanation Using Longitudinal Data*, John Wiley & Sons
- [30] Singer, B. & Cohen, S. B. (1980) - "Estimating Malaria Incidence and Recovery Rates from Panel Surveys", *Mathematical Biosciences*, 49, 273-305
- [31] Singer, B. & Spilerman, S. (1976) - "The Representation of Social Processes by Markov Models", *American Journal of Sociology*, 82, 1-54
- [32] Smith, B.T.; Boyle, J. M.; Dongarra, J. J.; Garbow, B. S.; Ikebe, Y.; Klema, V. C. & Moler, C. B. (1976) - *Matrix Eigen-System Routines: Eispack Guide*, Springer-Verlag

- [33] Stram, D. O.; Wei, L. J. & Ware, J. H. (1988) - "Analysis of Repeated Ordered Categorical Outcomes with Possibly Missing Observations and Time-Dependent Covariates", *Journal of American Statistical Association*, 83, 631-637
- [34] Strang, G. (1980) - *Linear Algebra and Its Applications*, Harcourt Brace Jovanovich, Publishers
- [35] Tuma, N. B.; Hannan, M. T. & Groeneveld, L. P. (1979) - "Dynamic Analysis of Event Histories", *American Journal of Sociology*, 84, 820-854
- [36] Tuma, N. B. & Robins, P. K. (1980) - "A Dynamic Model of Employment Behavior: An Application to the Seattle and Denver Income Maintenance Experiments", *Econometrica*, 48, 1031-1052
- [37] Ware, J. H. (1985) - "Linear Models for the Analysis of Longitudinal Studies", *The American Statistician*, 39, 95-101
- [38] Ware, J. H.; Dockery, D. W.; Spiro, A.; Speizer, F. E. & Ferris, B. G. J. (1984) - "Passive Smoking, Gas Cooking and Respiratory Health of Children Living in Six Cities", *American Review of Respiratory Diseases*, 129, 366-374
- [39] Ware, J. H.; Lipsitz, S. & Speizer, F. E. (1988) - "Issues in the Analysis of Repeated Categorical Outcomes", *Statistics in Medicine*, 7, 95-107
- [40] Wasserman, S. (1980) - "Analysing Social Networks as Stochastic Processes", *Journal of American Statistical Association*, 75, 280-294
- [41] Zeger, S. L. & Liang, K. Y. (1986) - "Longitudinal Data Analysis for Discrete and Continuous Outcomes", *Biometrics*, 42, 121-130

Apêndice A

A-1 Obtenção de (4.4)

Vamos primeiramente expor um resultado importante na teoria de equações diferenciais ordinárias (ver Coddington & Levinson (1955)):

Se

$$\frac{\partial X(t)}{\partial t} = AX(t) \quad \text{e} \quad X(0) = C,$$

Então

$$X(t) = e^{At}C \tag{A.1}$$

De acordo com as equações retrospectivas (2.18), vistas na seção 2.3 do capítulo 2, temos

$$\frac{\partial P(t)}{\partial t} = QP(t) \quad \text{e} \quad P(0) = I, \tag{A.2}$$

onde I denota a matriz identidade.

Qualquer matriz quadrada pode ser decomposta da seguinte forma: $Q = ADA^{-1}$, onde $D = \text{diag}(d_1, \dots, d_k)$ ($d_i \neq d_j, \forall i, j = 1, \dots, k$). Vamos fazer $P(t) = AY(t)$. Logo, $Y(t) = A^{-1}P(t)$. Então

$$\frac{\partial Y(t)}{\partial t} = A^{-1} \frac{\partial P(t)}{\partial t} = A^{-1}QP(t) = \underbrace{A^{-1}QA}_{D}Y(t)$$

$$\Rightarrow \frac{\partial Y(t)}{\partial t} = DY(t) \quad \text{e} \quad Y(0) = A^{-1}.$$

Portanto, por (A.1) temos:

$$Y(t) = e^{Dt}A^{-1}$$

Retornando à variável original, temos:

$$A^{-1}P(t) = e^{Dt}A^{-1} \Rightarrow P(t) = Ae^{Dt}A^{-1}.$$

$$e^{Dt} = \sum_{m=0}^{\infty} \frac{(Dt)^m}{m!}$$

$$\frac{(Dt)^m}{m!} = \text{diag} \left(\frac{d_1^m t^m}{m!}, \dots, \frac{d_k^m t^m}{m!} \right)$$

$$\sum_{m=0}^{\infty} \frac{(Dt)^m}{m!} = \text{diag} \left(\sum_{m=0}^{\infty} \frac{d_1^m t^m}{m!}, \dots, \sum_{m=0}^{\infty} \frac{d_k^m t^m}{m!} \right) = \text{diag}(e^{d_1 t}, \dots, e^{d_k t}) = e^{Dt}.$$

Portanto,

$$P(t) = A \text{diag}(e^{d_1 t}, \dots, e^{d_k t}) A^{-1}.$$

Dessa forma é obtida a equação (4.4) do capítulo 4.

A-2 Obtenção de (4.5)

Suponhamos que Q tem autovalores distintos d_1, \dots, d_k para todo θ . Por (4.4) e pelo fato de $Q = ADA^{-1}$, temos

$$\frac{\partial P(t)}{\partial \theta_u} = \frac{\partial e^{Qt}}{\partial \theta_u} = \sum_{s=1}^{\infty} \frac{\partial}{\partial \theta_u} \left(\frac{Q^s t^s}{s!} \right) = \sum_{s=1}^{\infty} \frac{t^s}{s!} \frac{\partial Q^s}{\partial \theta_u}$$

$$\frac{\partial Q^s}{\partial \theta_u} = \frac{\partial Q Q^{s-1}}{\partial \theta_u} = \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-1} + Q \times \frac{\partial Q^{s-1}}{\partial \theta_u}$$

$$\frac{\partial Q^{s-1}}{\partial \theta_u} = \frac{\partial Q Q^{s-2}}{\partial \theta_u} = \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-2} + Q \times \frac{\partial Q^{s-2}}{\partial \theta_u}$$

$$\frac{\partial Q^{s-2}}{\partial \theta_u} = \frac{\partial Q Q^{s-3}}{\partial \theta_u} = \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-3} + Q \times \frac{\partial Q^{s-3}}{\partial \theta_u}$$

⋮

$$\frac{\partial Q^{s-(s-1)}}{\partial \theta_u} = \left(\frac{\partial Q}{\partial \theta_u} \right)$$

E portanto, substituindo adequadamente, temos:

$$\begin{aligned}
\frac{\partial Q^s}{\partial \theta_u} &= \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-1} + Q \left[\left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-2} + Q \frac{\partial Q^{s-2}}{\partial \theta_u} \right] \\
&= \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-1} + Q \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-2} + Q^2 \frac{\partial Q^{s-2}}{\partial \theta_u} \\
&= \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-1} + Q \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-2} + Q^2 \left[\left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-3} + Q \frac{\partial Q^{s-3}}{\partial \theta_u} \right] \\
&= \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-1} + Q \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-2} + Q^2 \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-3} + Q^3 \frac{\partial Q^{s-3}}{\partial \theta_u} \\
&= \sum_{l=0}^{s-1} Q^l \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-1-l}.
\end{aligned}$$

Temos então,

$$\frac{\partial P(t)}{\partial \theta_u} = \sum_{s=1}^{\infty} \frac{t^s}{s!} \sum_{l=0}^{s-1} Q^l \left(\frac{\partial Q}{\partial \theta_u} \right) Q^{s-1-l} = \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} \mathbf{A} D^l \mathbf{G}_u D^{s-1-l} \mathbf{A}^{-1} \frac{t^s}{s!}$$

onde $\mathbf{G}_u = \mathbf{A}^{-1}(\partial Q/\partial \theta_u)\mathbf{A}$. Portanto, temos a equação (4.5), isto é,

$$\frac{\partial P(t)}{\partial \theta_u} = \mathbf{A} \left(\sum_{s=1}^{\infty} \sum_{l=0}^{s-1} D^l \mathbf{G}_u D^{s-1-l} \frac{t^s}{s!} \right) \mathbf{A}^{-1} = \mathbf{A} \mathbf{V}_u \mathbf{A}^{-1}.$$

\mathbf{V}_u é uma matriz $k \times k$, cujo (i, j) -ésimo elemento é dado por:

$$g_{ij}^{(u)} \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} d_i^{s-1-l} d_j^l \frac{t^s}{s!} = g_{ij}^{(u)} \frac{e^{d_i t} - e^{d_j t}}{d_i - d_j}, \quad \text{se } i \neq j,$$

$$g_{ii}^{(u)} \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} d_i^{s-1} \frac{t^s}{s!} = g_{ii}^{(u)} t e^{d_i t}, \quad \text{se } i = j,$$

onde $g_{ij}^{(u)}$ é o (i, j) -ésimo elemento em \mathbf{G}_u .

Apêndice B

B-1 Método δ Multivariado para Distribuições Assintóticas

Seja $\hat{\theta}$, um vetor b -dimensional de estimativas dos parâmetros e seja θ um vetor b -dimensional de parâmetros, isto é, $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_b)$ e $\theta = (\theta_1, \dots, \theta_b)$. Assumimos que

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}_b(\mathbf{0}, \Sigma), \quad (\text{B.1})$$

onde Σ é a matriz $b \times b$ de covariância assintótica de $\hat{\theta}$.

Suponhamos f um vetor definido em um intervalo aberto do espaço b -dimensional assumindo valores no espaço l -dimensional, isto é,

$$f(\theta) = (f_1(\theta), \dots, f_l(\theta)).$$

Assumimos que f tem uma diferencial em θ , ou seja, a função f tem a seguinte expansão à medida que $x \rightarrow \theta$:

$$f(x) = f(\theta) + (x - \theta) \left(\frac{\partial f}{\partial \theta} \right)' + o(\|x - \theta\|), \quad (\text{B.2})$$

onde $(\partial f / \partial \theta)$ denota a matriz $l \times b$ cujo (i, j) -ésimo elemento é a derivada parcial de f_i com respeito à j -ésima coordenada de $x = (x_1, \dots, x_b)$ calculada para $x = \theta$, ou seja,

$$\left(\frac{\partial f}{\partial \theta} \right)_{ij} = \left. \frac{\partial f_i}{\partial x_j} \right|_{x=\theta},$$

quando $x \rightarrow \theta$.

Se (B.1) e (B.2) ocorrem, a distribuição assintótica de $f(\hat{\theta})$ é dada por:

$$\sqrt{n}(f(\hat{\theta}) - f(\theta)) \xrightarrow{D} \mathcal{N}_l\left(\mathbf{0}, \left(\frac{\partial f}{\partial \theta} \right) \Sigma \left(\frac{\partial f}{\partial \theta} \right)'\right). \quad (\text{B.3})$$

Em termos de coordenadas individuais de $f(\hat{\theta})$, temos que, baseado em (B.3) e com σ_{uv} representando o (u, v) -ésimo componente da matriz Σ , a variância assintótica de $f_i(\hat{\theta})$ é

$$\sum_{u=1}^b \sum_{v=1}^b \sigma_{uv}(\theta) \left(\frac{\partial f_i}{\partial \theta_u} \right) \left(\frac{\partial f_i}{\partial \theta_v} \right)$$

e a covariância assintótica de $f_i(\hat{\theta})$ e $f_j(\hat{\theta})$ é

$$\sum_{u=1}^b \sum_{v=1}^b \sigma_{uv}(\theta) \left(\frac{\partial f_i}{\partial \theta_u} \right) \left(\frac{\partial f_j}{\partial \theta_v} \right).$$

Para maiores detalhes ver Bishop *et al.* (1980) - Cap. 14.

B-2 Utilização do Método δ Multivariado

Como foi mencionado na seção 4.2, características específicas do modelo podem ser de interesse e foi mostrado que elas podem ser facilmente estimadas utilizando-se a estimativa $\hat{\theta}$ de θ e variâncias e covariâncias assintóticas são obtidas através do método δ multivariado. No capítulo 4 foi dado como exemplo o tempo médio de permanência no estado i , $-q_{ii}(\theta)^{-1}$. Vamos agora mostrar outros exemplos de utilização do método para encontrar a matriz de covariância assintótica que é utilizada para construção de testes de hipótese.

Podemos utilizá-lo, por exemplo, a fim de estimar a variância assintótica da probabilidade de transição $p_{ij}(t; \theta)$ para i, j e t fixos. Através de (4.4) tem-se as probabilidades de transição como função de θ . A variância assintótica de $p_{ij}(t; \theta)$ é

$$\sum_{u=1}^b \sum_{v=1}^b \frac{\partial p_{ij}(t; \theta)}{\partial \theta_u} \times \frac{\partial p_{ij}(t; \theta)}{\partial \theta_v} \times M^{uv}(\theta) \Big|_{\theta=\hat{\theta}},$$

onde $M^{uv}(\theta)$ é o (u, v) -ésimo elemento de $M^{-1}(\theta)$ e as derivadas são obtidas utilizando-se (4.5).

Uma característica importante desses modelos é a distribuição estacionária $\pi' = (\pi_1, \dots, \pi_k)$, definida como a solução única da equação $Q'\pi = 0$ satisfazendo $\sum_{i=1}^k \pi_i = 1$. O vetor π não é uma função simples de θ , porém para a utilização do método δ multivariado são necessárias as derivadas $\partial \pi_i / \partial \theta_j$, as quais mostraremos como podem ser encontradas.

O vetor $\pi' = (\pi_1, \dots, \pi_{k-1}, 1 - \pi_1 - \dots - \pi_{k-1})$ pode ser obtido como a solução única do sistema de equações

$$Q_1' \pi = 0 \tag{B.4}$$

onde $Q_1 = Q_1(\theta)$ é a matriz $k \times k - 1$ obtida eliminando-se a última coluna de $Q(\theta)$. Dessa forma,

$$Q_1' \pi = \begin{pmatrix} q_{11}\pi_1 + q_{21}\pi_2 + \cdots + q_{k-1,1}\pi_{k-1} + q_{k1}(1 - \pi_1 - \cdots - \pi_{k-1}) \\ q_{12}\pi_1 + q_{22}\pi_2 + \cdots + q_{k-1,2}\pi_{k-1} + q_{k2}(1 - \pi_1 - \cdots - \pi_{k-1}) \\ \vdots \\ q_{1,k-1}\pi_1 + q_{2,k-1}\pi_2 + \cdots + q_{k-1,k-1}\pi_{k-1} + q_{k,k-1}(1 - \pi_1 - \cdots - \pi_{k-1}) \end{pmatrix}$$

$$= B(\theta)\pi^* + Q^*,$$

onde $B(\theta)$ é uma matriz $(k - 1) \times (k - 1)$ cujo (i, j) -ésimo elemento é

$$B_{ij}(\theta) = q_{ji}(\theta) - q_{ki}(\theta), \quad i, j = 1, \dots, k - 1,$$

$$\pi^* = (\pi_1, \dots, \pi_{k-1})' \quad \text{e} \quad Q^* = (q_{k1}, \dots, q_{k,k-1})'.$$

Portanto,

$$Q_1' \pi = B(\theta)\pi^* + Q^*. \quad (\text{B.5})$$

Denotando $Q_1' \pi$ por $F(\theta, \pi)$, isto é, a função definindo π_1, \dots, π_{k-1} implicitamente em termos de $\theta_1, \dots, \theta_b$, temos

$$\frac{\partial F(\theta, \pi)}{\partial \theta_j} = \left(\frac{\partial B(\theta)}{\partial \theta_j} \right) \pi^* + B(\theta) \left(\frac{\partial \pi^*}{\partial \theta_j} \right) + \frac{\partial Q^*}{\partial \theta_j} \quad (\text{B.6})$$

Mas como

$$\left(\frac{\partial B(\theta)}{\partial \theta_j} \right) \pi^* + \frac{\partial Q^*}{\partial \theta_j} = \begin{pmatrix} \frac{\partial q_{11}}{\partial \theta_j} \pi_1 + \cdots + \frac{\partial q_{k-1,1}}{\partial \theta_j} \pi_{k-1} + \frac{\partial q_{k1}}{\partial \theta_j} (1 - \pi_1 - \cdots - \pi_{k-1}) \\ \vdots \\ \frac{\partial q_{1,k-1}}{\partial \theta_j} \pi_1 + \cdots + \frac{\partial q_{k-1,k-1}}{\partial \theta_j} \pi_{k-1} + \frac{\partial q_{k,k-1}}{\partial \theta_j} (1 - \pi_1 - \cdots - \pi_{k-1}) \end{pmatrix}$$

$$= \left(\frac{\partial Q_1}{\partial \theta_j} \right)' \pi,$$

temos de (B.6) e (B.4) que:

$$\frac{\partial F(\theta, \pi)}{\partial \theta_j} = B(\theta) \left(\frac{\partial \pi^*}{\partial \theta_j} \right) + \left(\frac{\partial Q_1}{\partial \theta_j} \right)' \pi = B(\theta) \begin{pmatrix} \frac{\partial \pi_1}{\partial \theta_j} \\ \vdots \\ \frac{\partial \pi_{k-1}}{\partial \theta_j} \end{pmatrix} + C_j(\theta) = 0, \quad (\text{B.7})$$

onde $C_j(\theta)$ é o vetor $(k-1) \times 1$ dado por

$$C_j(\theta) = \left(\frac{\partial Q_1}{\partial \theta_j} \right)' \pi$$

De (B.7) temos que as derivadas $\partial \pi_i / \partial \theta_j$ ($i = 1, \dots, k-1$) são obtidas através de

$$\left(\frac{\partial \pi_1}{\partial \theta_j}, \dots, \frac{\partial \pi_{k-1}}{\partial \theta_j} \right)' = -B(\theta)^{-1} C_j(\theta). \quad (\text{B.8})$$

A matriz com todas as derivadas de π_1, \dots, π_{k-1} com respeito a $\theta_1, \dots, \theta_b$ pode ser encontrada através de

$$W(\theta) = -B(\theta)^{-1} C(\theta), \quad (\text{B.9})$$

onde $C(\theta)$ é a matriz $(k-1) \times b$ com j -ésima coluna $C_j(\theta)$.

Uma estimativa da matriz de covariância de $\hat{\pi}^* = (\hat{\pi}_1, \dots, \hat{\pi}_{k-1})'$ é

$$\hat{\Sigma}_{\pi^*} = W(\hat{\theta}) \hat{\Sigma} W(\hat{\theta})', \quad (\text{B.10})$$

onde $\hat{\Sigma} = M(\hat{\theta})^{-1}$ é a estimativa da matriz de covariância assintótica de $\hat{\theta}$.

Podemos agora, partindo desse resultado obter a matriz de covariância assintótica de $\hat{\pi}' = (\hat{\pi}_1, \dots, \hat{\pi}_k)$. Sabemos que

$$\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_k \end{pmatrix} = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_{k-1} \\ 1 - \pi_1 - \dots - \pi_{k-1} \end{pmatrix}.$$

Temos então que:

$$\pi = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ -1 & -1 & \dots & -1 & -1 \end{pmatrix} \times \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_{k-1} \end{pmatrix}. \quad (\text{B.11})$$

Portanto, de (B.10) e (B.11) temos que:

$$\hat{\Sigma}_\pi = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ -1 & -1 & \cdots & -1 & -1 \end{pmatrix} \times \hat{\Sigma}_\pi \times \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ -1 & -1 & \cdots & -1 & -1 \end{pmatrix}', \quad (\text{B.12})$$

onde $\hat{\Sigma}_\pi$ é a estimativa da matriz de covariância assintótica de $\hat{\pi}$.

Apêndice C

Dados para o exemplo do capítulo 5

| P | T | 0 | 1 | 2 | 3 | P | T | 0 | 1 | 2 | 3 | P | T | 0 | 1 | 2 | 3 |
|----|---|---|---|---|---|----|---|---|---|---|---|----|---|---|---|---|---|
| 1 | 0 | 2 | 3 | 2 | 2 | 31 | 0 | 3 | 3 | 2 | 2 | 61 | 0 | 3 | 3 | 2 | 3 |
| 2 | 0 | 2 | 3 | 3 | 3 | 32 | 0 | 3 | 1 | 1 | 1 | 62 | 1 | 3 | 2 | 1 | 1 |
| 3 | 0 | 3 | 3 | 3 | 3 | 33 | 0 | 3 | 3 | 2 | 2 | 63 | 1 | 3 | 1 | 1 | 1 |
| 4 | 0 | 2 | 3 | 3 | 3 | 34 | 0 | 3 | 3 | 2 | 2 | 64 | 1 | 3 | 1 | 1 | 1 |
| 5 | 0 | 2 | 3 | 3 | 3 | 35 | 0 | 3 | 2 | 2 | 2 | 65 | 1 | 3 | 1 | 1 | 1 |
| 6 | 0 | 3 | 3 | 3 | 3 | 36 | 0 | 3 | 1 | 1 | 1 | 66 | 1 | 2 | 1 | 1 | 1 |
| 7 | 0 | 3 | 2 | 2 | 3 | 37 | 0 | 2 | 3 | 2 | 2 | 67 | 1 | 2 | 1 | 1 | 1 |
| 8 | 0 | 2 | 3 | 3 | 3 | 38 | 0 | 3 | 3 | 3 | 3 | 68 | 1 | 2 | 1 | 1 | 1 |
| 9 | 0 | 2 | 3 | 3 | 3 | 39 | 0 | 2 | 3 | 3 | 2 | 69 | 1 | 2 | 1 | 1 | 1 |
| 10 | 0 | 3 | 2 | 2 | 2 | 40 | 0 | 3 | 3 | 2 | 2 | 70 | 1 | 2 | 1 | 1 | 1 |
| 11 | 0 | 3 | 3 | 3 | 3 | 41 | 0 | 2 | 3 | 2 | 2 | 71 | 1 | 2 | 2 | 1 | 1 |
| 12 | 0 | 3 | 1 | 1 | 1 | 42 | 0 | 3 | 2 | 2 | 2 | 72 | 1 | 2 | 1 | 1 | 1 |
| 13 | 0 | 3 | 1 | 1 | 1 | 43 | 0 | 3 | 3 | 2 | 2 | 73 | 1 | 2 | 1 | 1 | 1 |
| 14 | 0 | 3 | 3 | 2 | 2 | 44 | 0 | 2 | 2 | 2 | 2 | 74 | 1 | 2 | 2 | 1 | 1 |
| 15 | 0 | 2 | 3 | 1 | 3 | 45 | 0 | 2 | 3 | 3 | 3 | 75 | 1 | 2 | 1 | 1 | 1 |
| 16 | 0 | 3 | 3 | 3 | 2 | 46 | 0 | 2 | 1 | 1 | 1 | 76 | 1 | 2 | 2 | 1 | 1 |
| 17 | 0 | 3 | 2 | 3 | 3 | 47 | 0 | 2 | 1 | 1 | 1 | 77 | 1 | 2 | 2 | 2 | 1 |
| 18 | 0 | 3 | 3 | 2 | 2 | 48 | 0 | 2 | 1 | 1 | 1 | 78 | 1 | 2 | 1 | 1 | 1 |
| 19 | 0 | 2 | 1 | 1 | 1 | 49 | 0 | 2 | 1 | 2 | 2 | 79 | 1 | 2 | 1 | 1 | 1 |
| 20 | 0 | 3 | 1 | 1 | 2 | 50 | 0 | 2 | 3 | 2 | 2 | 80 | 1 | 2 | 2 | 2 | 2 |
| 21 | 0 | 2 | 1 | 1 | 2 | 51 | 0 | 2 | 2 | 2 | 2 | 81 | 1 | 2 | 1 | 1 | 1 |
| 22 | 0 | 2 | 2 | 3 | 3 | 52 | 0 | 3 | 2 | 2 | 3 | 82 | 1 | 2 | 1 | 1 | 1 |
| 23 | 0 | 2 | 1 | 1 | 1 | 53 | 0 | 2 | 2 | 2 | 3 | 83 | 1 | 2 | 1 | 1 | 1 |
| 24 | 0 | 3 | 2 | 2 | 2 | 54 | 0 | 2 | 2 | 2 | 2 | 84 | 1 | 3 | 1 | 1 | 1 |
| 25 | 0 | 2 | 2 | 2 | 2 | 55 | 0 | 2 | 1 | 1 | 1 | 85 | 1 | 2 | 3 | 2 | 2 |
| 26 | 0 | 3 | 3 | 2 | 2 | 56 | 0 | 2 | 3 | 3 | 3 | 86 | 1 | 2 | 1 | 1 | 1 |
| 27 | 0 | 3 | 3 | 3 | 3 | 57 | 0 | 2 | 3 | 2 | 2 | 87 | 1 | 3 | 1 | 1 | 1 |
| 28 | 0 | 3 | 3 | 2 | 3 | 58 | 0 | 3 | 1 | 1 | 1 | 88 | 1 | 3 | 1 | 1 | 1 |
| 29 | 0 | 2 | 1 | 2 | 2 | 59 | 0 | 3 | 3 | 3 | 3 | 89 | 1 | 3 | 1 | 1 | 1 |
| 30 | 0 | 3 | 2 | 2 | 2 | 60 | 0 | 3 | 3 | 3 | 3 | 90 | 1 | 2 | 1 | 1 | 1 |

| P | T | 0 | 1 | 2 | 3 | P | T | 0 | 1 | 2 | 3 | P | T | 0 | 1 | 2 | 3 |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 91 | 1 | 2 | 2 | 1 | 1 | 106 | 1 | 3 | 2 | 2 | 2 | 121 | 1 | 3 | 1 | 1 | 1 |
| 92 | 1 | 3 | 2 | 1 | 1 | 107 | 1 | 2 | 1 | 1 | 1 | 122 | 1 | 2 | 2 | 1 | 1 |
| 93 | 1 | 2 | 1 | 1 | 1 | 108 | 1 | 3 | 2 | 1 | 1 | 123 | 1 | 3 | 1 | 1 | 1 |
| 94 | 1 | 2 | 2 | 1 | 1 | 109 | 1 | 3 | 1 | 1 | 1 | 124 | 1 | 3 | 1 | 1 | 1 |
| 95 | 1 | 3 | 1 | 1 | 2 | 110 | 1 | 2 | 2 | 3 | 3 | 125 | 1 | 3 | 1 | 1 | 1 |
| 96 | 1 | 3 | 1 | 1 | 1 | 111 | 1 | 3 | 1 | 1 | 1 | 126 | 1 | 3 | 1 | 1 | 1 |
| 97 | 1 | 3 | 1 | 1 | 1 | 112 | 1 | 3 | 1 | 1 | 1 | 127 | 1 | 2 | 1 | 1 | 1 |
| 98 | 1 | 2 | 1 | 1 | 1 | 113 | 1 | 3 | 2 | 3 | 3 | 128 | 1 | 3 | 2 | 1 | 1 |
| 99 | 1 | 3 | 3 | 3 | 3 | 114 | 1 | 3 | 1 | 1 | 1 | 129 | 1 | 3 | 2 | 2 | 3 |
| 100 | 1 | 3 | 2 | 1 | 1 | 115 | 1 | 3 | 1 | 1 | 1 | 130 | 1 | 3 | 1 | 1 | 1 |
| 101 | 1 | 3 | 1 | 1 | 1 | 116 | 1 | 3 | 2 | 1 | 1 | 131 | 1 | 2 | 1 | 1 | 1 |
| 102 | 1 | 3 | 2 | 2 | 2 | 117 | 1 | 3 | 1 | 1 | 1 | 132 | 1 | 3 | 1 | 1 | 1 |
| 103 | 1 | 3 | 1 | 1 | 1 | 118 | 1 | 3 | 1 | 1 | 1 | 133 | 1 | 2 | 1 | 1 | 1 |
| 104 | 1 | 2 | 3 | 3 | 3 | 119 | 1 | 2 | 1 | 1 | 1 | 134 | 1 | 3 | 1 | 1 | 1 |
| 105 | 1 | 2 | 2 | 2 | 2 | 120 | 1 | 3 | 1 | 1 | 1 | 135 | 1 | 3 | 1 | 1 | 1 |

NOTAÇÃO

P - Paciente

T - Tratamento

0 - Condição Inicial

1 - Condição na 1ª ocasião

2 - Condição na 2ª ocasião

3 - Condição na 3ª ocasião