

UNIVERSIDADE FEDERAL DO PARANÁ – UFPR
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE GRADUAÇÃO EM BACHARELADO EM ESTATÍSTICA

JEFERSON LUIS POVOROZNEK
JULIO ALEXANDRE MELLO GARCIA

MODELOS MARKOVIANOS
MULTI-ESTADOS

ORIENTADORA:
PROF^a.DR^a. SILVIA EMIKO SHIMAKURA

CURITIBA – PR
2008

JEFERSON LUIS POVOROZNEK
JULIO ALEXANDRE MELLO GARCIA

MODELOS MARKOVIANOS
MULTI-ESTADOS

Trabalho de Conclusão de Curso apresentado
ao Curso de Graduação de Bacharelado em
Estatística da Universidade Federal do Paraná
como requisito parcial para aprovação na disciplina
de Laboratório de Estatística II.

Orientadora: Prof^a. Dr^a Silvia Emiko Shimakura

CURITIBA – PR
2008

JEFERSON LUIS POVOROZNEK
JULIO ALEXANDRE MELLO GARCIA

MODELOS MARKOVIANOS
MULTI-ESTADOS

Aprovados em ____/____/____.

Profª Silvia Emiko Shimakura
Depto. Estatística/ UFPR

Prof. Benito Olivares Aguilera
Depto. Estatística/ UFPR

CONCEITO FINAL: _____

Sumário

1	INTRODUÇÃO	1
1.1	Comentários Iniciais	1
1.2	Objetivos	2
1.2.1	Objetivo Geral	2
1.2.2	Objetivos Específicos	2
1.3	Material e Métodos	2
1.4	Delimitações do Estudo	3
2	CONCEITOS	3
2.1	Revisão de Processos Estocásticos	3
2.1.1	Processo Markoviano	4
2.1.2	Cadeias de Markov	4
2.1.3	Classificação de Estados em Cadeias de Markov	5
2.2	Modelos Multi-estado	6
2.3	Modelos de progressão de doença	8
2.4	Observação em tempo arbitrário	9
2.5	Informativo sobre amostragem no tempo	10
2.6	Tempo médio de permanência (<i>Mean Sojourn Time</i>)	11
2.7	Função de Verossimilhança	12
3	MSM	12
3.1	O conjunto de dados do CAV	13
3.2	Instalando e carregando o MSM	13
3.3	Obtendo os dados	13
3.4	Conhecendo as variáveis	14
3.5	Sumarizando os Dados	15
3.6	Ajustando o Modelo MSM	17
3.7	Funções de Extração	21
3.7.1	Média de tempo em cada estado	21
3.7.2	Probabilidades de transição	22
3.7.3	Gráfico de Sobrevivência	22
4	CONSIDERAÇÕES FINAIS	24
5	REFERÊNCIAS BIBLIOGRÁFICAS	25

Lista de Figuras

1	Modelo Multi-estado Geral.	6
2	Modelo Geral para progressão de doença.	8
3	Modelo de doença-morte.	9
4	Evolução de um Modelo Multi-estado.	10
5	Gráfico de Sobrevivência vs Tempo sem o estado absorvente.	23

RESUMO

Modelos multi-estados de Markov são utilizados na área médica para estimar as probabilidades de transição entre vários estágios de uma doença, podendo o paciente recuperar-se ou morrer, por exemplo. Outros modelos mais usuais como modelos de sobrevida e regressão logística, não levam em consideração as informações entre os estados adjacentes. Por isso, modelos multi-estados tornam-se mais atraentes do que outro tipo de análise, por aproveitar as características e informações dos estados relacionados ao fenômeno estudado. O principal interesse desse trabalho é apresentar a aplicação dessa metodologia de análise através do *software* estatístico *R*, servindo assim, de suporte básico a quem tenha interesse na utilização dessa técnica.

Palavras-chave: Processos Estocásticos, Modelos Markovianos, MSM

1 INTRODUÇÃO

1.1 Comentários Iniciais

Quando tem-se um evento que varia ao longo do tempo, assumindo diferentes características, vem o interesse em saber a probabilidade da transição entre esses estágios. Uma doença por exemplo, na primeira visita ao médico o paciente não apresenta sintomas, na segunda os sintomas já estão exageradamente avançados. Ou seja, ele transitou de um estágio inicial para um mais avançado sem ser apercebido a mudança, mas e os outros estágios intermediários da doença? Quando se deu a transição para os estágios iniciais? Como prever essas transições sendo que o tempo exato de ocorrência é desconhecido?

A mudança de um estágio da doença para outro possui uma probabilidade de ocorrência. Uma pessoa com características x pode ter probabilidade diferente de uma pessoa de características y para transitar de uma fase para outra. Se forem analisadas as características dessas pessoas e como determinada doença se manifesta e evolui probabilisticamente, conclusões poderão ser tiradas. A um processo como esse, dá-se o nome de processo estocástico ou processo probabilístico. Existem várias técnicas estatísticas usadas em processos estocásticos que poderiam ser usadas para resolver um problema assim, dependendo é claro, do ponto de vista do problema. A proposta desse trabalho é justamente o estudo e apresentação de uma técnica capaz de mapear probabilisticamente essas transições.

Andrei Andreyevich Markov nasceu no dia 14 de junho de 1856 em Ryazan, na Rússia e morreu no dia 20 de julho de 1922 em St Petersburg, Rússia. Seu trabalho teve início com estudos em teoria dos números, frações contínuas, limites de integrais, teoria da aproximação e a convergência de séries. Markov, provou o teorema central do limite e é lembrado pelo estudo de cadeias que levam seu nome. As Cadeias de Markov, como assim são chamadas, são processos estocásticos em que as variáveis estão definidas em espaço de estados discreto. Além disso possuem uma característica peculiar, a de não possuir memória, isto é, a probabilidade condicional de ocorrência de qualquer evento futuro independe dos eventos passados e sim depende, somente do evento presente. Não obstante existirem as cadeias de markov, os processos por ele tratados podem também apresentar estados em espaço infinito, além de variarem também em tempo, contínuo ou discreto. Por não tratar-se de somente espaço de estados finitos, tais processos passam a se chamar simplesmente Processos Markovianos ou ainda Processos Markovianos Multi-estados.

1.2 Objetivos

1.2.1 Objetivo Geral

O trabalho tem como objetivo, o estudo dos Modelos Markovianos Multi-estados aplicando-os via *software* estatístico R, pacote MSM.

1.2.2 Objetivos Específicos

Elaborar de forma clara e didática, material para descrever os Modelos Markovianos Multi-estados, de forma que sirva de suporte ao *MSM*, utilizando-se para isso, exemplos contidos no próprio pacote estatístico.

1.3 Material e Métodos

Para dar seguimento ao estudo, é usado conjunto de dados contido no próprio pacote MSM. O conjunto *CAV* da versão 0.8, mais recente, ou *HEART*, como era chamado em versões anteriores do pacote. Nele estão contidas informações sobre pacientes transplantados do coração, onde cada estado representa o quão deterioradas estão as paredes arteriais. Esse conjunto de dados contempla um total de 622 pacientes e possui 9 variáveis. São observados 4 estágios de transição onde o estágio 4 é um estado absorvente, pois entrando nesse estágio, caracteriza-se a morte do indivíduo. O objetivo da análise dos dados será encontrar a probabilidade de transição entre os estados, estágios, da *CAV* - alografia vascular-coronária - do inglês *coronary allograft vasculopathy*, levando em consideração as demais variáveis expostas no conjunto, porém, sem perder o enfoque didático do estudo.

Modelos Markovianos Multi-estados são frequentemente utilizados para modelar o curso de doenças. No caso do conjunto *CAV* temos 3 estágios adjacentes e um estágio absorvente, que seria a morte do indivíduo. Um modelo multi-estado é governado por uma matriz de intensidade de transição, que será chamada de *Q* e pode ser ilustrada conforme se segue para o caso do *CAV*.

$$Q = \begin{pmatrix} -(q_{12} + q_{14}) & q_{12} & 0 & q_{14} \\ q_{21} & -(q_{21} + q_{23} + q_{24}) & q_{23} & q_{24} \\ 0 & q_{32} & -(q_{32} + q_{34}) & q_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Um fato importante é que a transição pode ocorrer em qualquer tempo, mas os dados coletados, serão fotografias(*snapshots*) do processo, portanto um paciente pode ser observado no estado um e posteriormente no estado 3, mas isso não quer dizer que ele não passou pelo estado 2, e sim que o tempo de transição para o estado 2 não foi observado, contudo deve ser considerado, mesmo que estimando-o.

1.4 Delimitações do Estudo

Este estudo tem motivação em escrever material em português para servir de suporte ao MSM, já que a maioria da bibliografia é composta por artigos e escritos em inglês. Tendo em vista que esses modelos são muito vastos e abrangentes, com várias ramificações e derivações, o trabalho terá limitações quanto ao conteúdo desse estudo e inferência dos dados. Como os dados utilizados não são representativos e não há fidedignidade de coleta da amostra, não há como inferir ou generalizar os resultados apresentados, ou seja, são meramente didáticos. No que se refere ao conteúdo, não serão abordados dados com erros de classificação e nem adição de covariáveis, que existem, o MSM possui tratamento específico e é sempre muito interessante seu conhecimento.

2 CONCEITOS

O Modelo Markoviano Multi-estado é um modo útil para descrever um processo no qual um indivíduo movimenta-se por uma série de estados em tempo contínuo. Os dados podem ser, observações de determinado processo em tempo arbitrário de maneira que o tempo exato de quando as mudanças ocorrem não é conhecido, podem ser processos completos, quando tem-se exatamente o tempo de transição e processos com estados censurados, quando a transição pula estados intermediários, porém o paciente em determinado momento passou por esses estados. Por exemplo, a progressão de doenças crônicas é descrita freqüentemente por fases de severidade, e o estado do paciente só pode ser conhecido através da observação de um médico, portanto tem-se transições em tempo arbitrários e muitas censuras, inobservação de estados. Já um exemplo de processos completos seria quando determinado processo é acompanhado instantaneamente e o momento exato de transição é observado. Imagine o acompanhamento de deterioração de determinado produto. Considere que existam graus de deterioração e que um desses graus seja o de desgaste total do produto, conseqüentemente ao chegar nesse determinado grau o produto se estraga e torna-se inconsumível. Se esse processo for observado em tempo real e se o tempo de transição de um grau de deterioração para outro for observado no instante da transição, esse será um processo completo.

A seguir serão vistos conceitos teóricos importantes para completo entendimento e aplicação do pacote MSM, como método de estimação e interpretações das estatísticas extraídas com o pacote.

2.1 Revisão de Processos Estocásticos

Um Processo Estocástico é um conjunto de variáveis aleatórias $X(t)$ que variam ao longo de um período de tempo T pertencente, geralmente, ao conjunto dos inteiros não-

negativos. A variável $X(t)$ representa o estado que o processo está em um tempo t . Processos estocásticos podem ser classificados quanto ao estado e quanto ao tempo. Quanto ao estado podem haver:

- Estado Discreto: também chamado de cadeia, onde $X(t)$ é definido sobre um conjunto enumerável e finito.
- Estado Contínuo: $X(t)$ é definido sobre um conjunto infinito de estados.

E podem ser definidos em relação ao parâmetro tempo:

- Tempo Discreto: t é definido sobre um conjunto finito e enumerável.
- Tempo Contínuo: t está definido sobre um conjunto infinito.

2.1.1 Processo Markoviano

Dentre os vários tipos de Processos Estocásticos existentes, existem os Processos Markovianos. Esses processos tem a característica de terem a probabilidade condicional de ocorrência de qualquer evento futuro independente dos eventos passados e sim dependente, somente do evento presente. São chamados de processos sem memória. Em outras palavras tem-se que:

$$P[X(t_{k+1}) \leq x_{k+1} \mid X(t_k) = x_k, X(t_{k-1}) = x_{k-1}, \dots, X(t_1) = x_1, X(t_0) = x_0] = P[X(t_{k+1}) \leq x_{k+1} \mid X(t_k) = x_k]$$

para $t_0 \leq t_1 \leq \dots \leq t_k \leq t_{k+1}$ e toda sequência $k_0, k_1, \dots, k_{t-1}, k_t, k_{t+1}$

2.1.2 Cadeias de Markov

Tem-se uma cadeia de Markov, quando um processo markoviano tem seu espaço de estados definido em um conjunto discreto de estados, seja ele finito ou infinito. Quando se tem espaço de estados discreto e o tempo também é discreto, a Cadeia de Markov é dita ser uma Cadeia de Markov em Tempo Discreto. Neste caso, tem-se:

$$P[X(k+1) = x_{k+1} \mid X(k) = x_k, X(k-1) = x_{k-1}, \dots, X(1) = x_1, X(0) = x_0] = P[X(k+1) = x_{k+1} \mid X(k) = x_k]$$

\forall sequência $1, 2, \dots, k-1, k, k+1$

Caso tenha-se:

$$P[X(k+1) = x_{k+1} \mid X(k) = x_k = PX(1) = x_1 \mid X(0) = x_0]$$

\forall sequência $1, 2, \dots, k - 1, k, k + 1$

então as probabilidades de transição são ditas estacionárias.

Assim, tendo-se Probabilidades de Transição Estacionárias implica que as Probabilidades de Transição não mudam em relação ao tempo, ou seja, dado que se tenha uma matriz de transição P , as transições a n passos serão multiplicações entre matrizes, ou melhor dizendo, a matriz P elevada a n -ésima potência, P^n , sendo n o número de transições requeridas.

2.1.3 Classificação de Estados em Cadeias de Markov

Os estados em uma cadeia de Markov podem ser classificados quanto as suas características de transição. Um estado em uma cadeia pode ser chamado de:

- **Alcançável:** Quando a probabilidade de um estado i chegar a um outro determinado estado j , em qualquer tempo, ou seja, independente do número de transições, seja diferente de zero. Matematicamente tem-se que $p_{ij}^{(n)} > 0$, para algum $n \geq 0$.
- **Comunicante :** Um estado é dito comunicante com determinado estado se esses dois estados são alcançáveis entre si. Se dois estados são comunicantes entre si, então eles pertencem a mesma classe. Se todos os estados de uma cadeia são comunicantes entre si, então eles pertencem a mesma classe e assim, como a cadeia toda é comunicante, ela é dita ser uma cadeia irredutível. Ainda pode-se citar a seguinte regra: "Se um estado i é comunicante com um estado k e o estado k é comunicante com um estado j , então o estado i é comunicante com o estado j ".
- **Transiente:** Um estado é dito ser transiente se, entrando neste estado, o processo pode nunca retornar novamente para este estado, ou seja, partindo de um estado i e chegando a um estado j , alcançável por i mas não reciprocamente, esse processo nunca mais retorne ao estado i . Ele então é visitado um número finito de vezes.
- **Recorrente:** Um estado é dito recorrente quando ele não é transiente, pois diz-se ser um estado recorrente aquele que visitado um número infinito de vezes. A probabilidade de retorno a esse estado é 1.
- **Absorvente:** Um estado absorvente é um estado que, entrando-se nele nunca mais o processo saíra dele. A morte, em um processo onde está sendo analisado uma doença, é um exemplo de estado absorvente, ou seja, uma vez nesse estado, a probabilidade de se continuar nele é 1, $p_{ii} = 1$.
- **Periódico:** Um estado i é periódico com período t se um retorno a este estado é possível somente em $t, 2t, 3t, \dots$ passos para $t > 1$.

- Ergódicos: São estados recorrentes e aperiódicos, $t=1$. Uma cadeia com estados todos recorrentes e aperiódicos é dita uma cadeia ergódica.

2.2 Modelos Multi-estado

Como dito anteriormente, será usado como exemplo, conjunto de dados contido no próprio pacote MSM. O conjunto *CAV* é descrito por um Modelo Multi-estado em tempo contínuo, assim como no caso da Figura 1. Seus quatro estados são chamados 1, 2, 3, 4. Em cada tempo t o indivíduo está em um estado $S(t)$. A Figura 1 mostra que transições são possíveis entre estados e que entrando no estado 4, não há retorno. Estando o indivíduo em um estado r , o próximo estado para o qual o indivíduo se move, e o tempo da mudança, são indicados pela intensidade de transição $q_{rs}(t, z(t))$, para cada par de estados r e s . As intensidades também podem depender do tempo t do processo, ou de forma mais geral, de um conjunto de variáveis explicativas $Z(t)$ específicas para indivíduos ou dependentes do tempo.

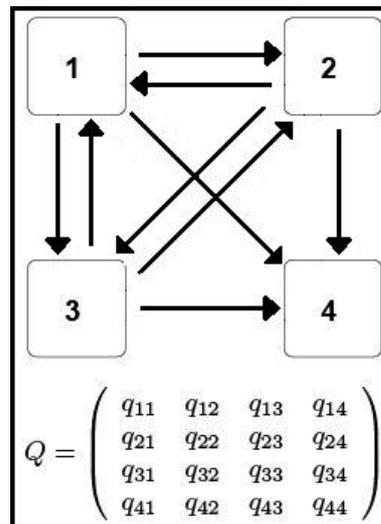


Figura 1: Modelo Multi-estado Geral.

A intensidade de transição representa o risco instantâneo do indivíduo migrar para o estado s a partir do estado r .

$$q_{rs}(t, z(t)) = \lim_{\delta t \rightarrow 0} P(S_i(t + \delta t) = s | S_i(t) = r) / \delta t$$

As intensidades formam uma matriz Q cujas linhas devem ter soma igual a zero, de forma que as entradas diagonais são definidas por:

$$q_{rr} = - \sum_{s \neq r} q_{rs}$$

Para ajustar um Modelo de Multi-estado aos dados, deve-se primeiro estimar essa matriz de intensidade de transição Q . Na suposição de Markov, a evolução de futuro só depende do estado presente. Ou seja, $q_{rs}(t, z(t), F_t)$, é independente de F_t , onde F_t é o histórico observado do processo até o tempo t precedente. Veja, por exemplo, Cox e Miller (1965), uma introdução completa para a teoria de Cadeias de Markov em tempo contínuo. No modelo de tempo contínuo de Markov, um único período de ocupação (ou tempo médio estimado) em um estado r , tem uma distribuição exponencial, com taxa dada por q_{rr} , o r -ésimo elemento da diagonal de Q (ou media $-1/q_{rr}$). Os demais elementos da diagonal de Q são proporcionais às probabilidades que definem o próximo estado r para o qual o indivíduo faz a transição. A probabilidade de que um indivíduo ocupe o estado s imediatamente após o estado r é q_{rs}/q_{rr} .

Através da matriz intensidade de transição $Q(t)$ pode-se achar a matriz probabilidade de transição $P(t)$, que é escrita como uma função de $Q(t)$.

$$P(t) = \exp(Qt)$$

Há algumas maneiras de se calcular a exponencial de uma matriz. Uma maneira simples é através da série de Taylor. Segue abaixo a fórmula e seu desdobramento que tende à convergência aproximadamente quando se eleva matriz à sua ordem quadrática. Seja Q uma matriz $n \times n$ então tem-se que:

$$P(t) = \exp(Qt) = \sum_{n=0}^{+\infty} \frac{(Qt)^n}{n!} = I + Qt + \frac{(Qt)^2}{2!} + \frac{(Qt)^3}{3!} + \dots$$

Sendo I um matriz indentidade de ordem n . Obtem-se então a matriz $P(t)$.

$$P(t) = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix}$$

2.3 Modelos de progressão de doença

O desenvolvimento do pacote msm, foi motivado para aplicação em modelagem de doenças. Como será utilizado um conjunto de dados de doença do coração é interessante algumas considerações.

Muitas doenças crônicas têm uma interpretação natural em termos de progressão organizada. Os Modelos Markovianos Multi-estado em tempo contínuo são usados frequentemente para modelar o curso delas. Um modelo geral usado é ilustrado na Figura 2. O modelo representa uma série sucessiva de estados de doenças severas e um estado chamado absorvente, frequentemente caracterizado pela morte do paciente. O paciente pode transitar entre os estados da doença, pode recuperar-se de estados adjacentes ou ainda, morrer em qualquer um dos estados. As observações do estado $S_i(t)$ para os indivíduos i são medidos em tempos arbitrários t que podem variar entre os indivíduos ou até sofrer censura de estados. As fases de doença podem ser modeladas como tempo contínuo homogêneo de Markov através de uma matriz de transição Q , ilustrada abaixo do modelo geral, na Figura2.

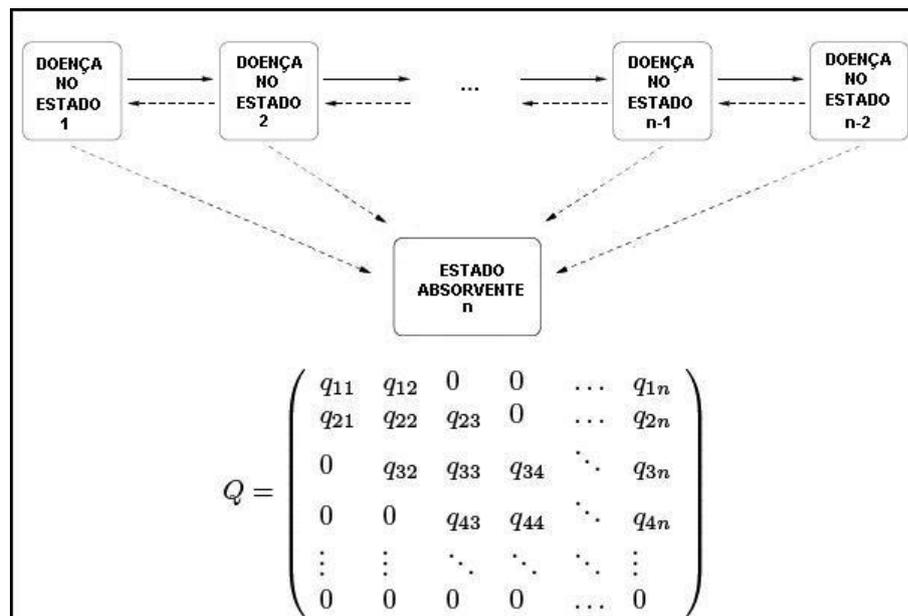


Figura 2: Modelo Geral para progressão de doença.

Um modelo comum e muito usado é o modelo de doença-morte, com três estados bem definidos que representam saúde, doença e morte conforme vê-se na Figura 3. Neste modelo são permitidas transições de saúde para doença, doença para morte e saúde para morte. A recuperação de doença para saúde também é considerada. Uma gama extensiva de situações médicas foi modelada usando métodos de multi-estado, por exemplo em rastreamento de aneurismas da aorta abdominal, problemas consecutivos de transplante de pulmão e de coração todos apresentados por Jackson et al.(2002), infecção por *HIV* e *AIDS* em Longini et al.(1989), dentre muitos outros.

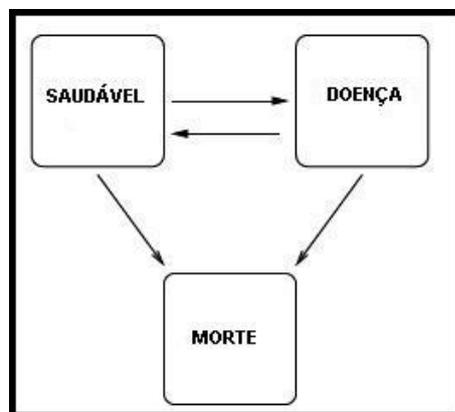


Figura 3: Modelo de doença-morte.

2.4 Observação em tempo arbitrário

Dados longitudinais para monitorar progressão de doença estão freqüentemente incompletos. Normalmente os pacientes são observados em visitas médicas intermitentes de monitoramento onde as informações são coletadas, mas as informações dos períodos entre visitas não estão disponíveis. Com isso o tempo exato de início da doença é desconhecido. Assim, as mudanças de estado em um modelo de multi-estado normalmente acontecem em tempo desconhecido. Um paciente só pode ser observado rigorosamente em parte da história de sua doença. Um horário de observação fixo pode ser especificado com antecedência, mas na prática as visitas podem variar devido às necessidades do paciente ou das pressões impostas pelo hospital, dentre outras causas. Os estados de modelos de progressão de doença incluem freqüentemente a morte. Estados de morte são normalmente registrados dentro de um dia. Observações podem também ser censuradas. Por exemplo, ao término de um estudo, um indivíduo só pode estar em um único estado. Uma situação típica desta situação é ilustrada na Figura 4. O indivíduo é observado em quatro ocasiões durante 10 meses. O evento final é a data de morte do indivíduo que é

registrada dentro de um dia no nono mês. As outras informações disponíveis, são as observações do indivíduo nos de estados 2 aos 1,5 meses do estudo, novamente no estado 2 aos 3,5 meses do estudo e no estado 1 aos 5 meses do estudo. Geralmente a transição entre estados e a permanência dentro dos mesmos, entre os tempos de observação, são desconhecidos. Veja por exemplo, que quando o paciente esta no estado 3 entre os 7 e 9 meses e não houve nenhuma observação sobre transição entre estados. Note também que o espaçamento entre as observações não são iguais, isso influencia na estimação da função de verossimilhança que leva em consideração a diferença $(t_{(j+1)} - t_j)$.

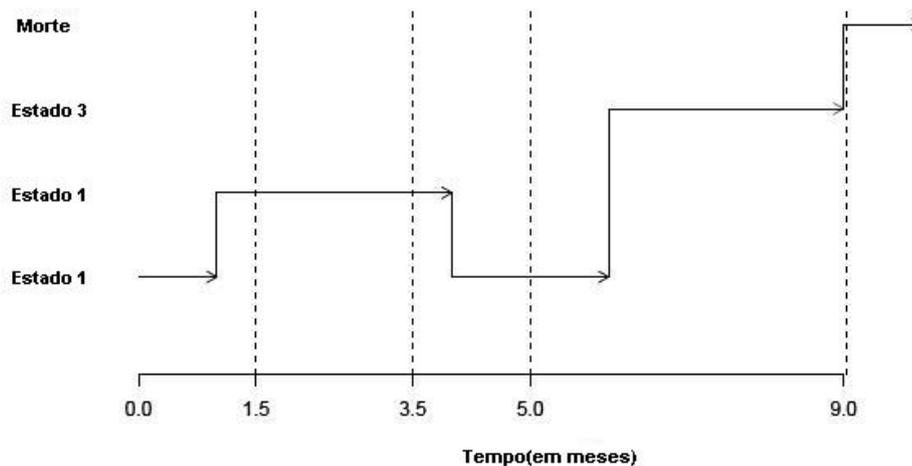


Figura 4: Evolução de um Modelo Multi-estado.

2.5 Informativo sobre amostragem no tempo

A característica dos dados longitudinais é a existência de observações repetidas para os mesmos indivíduos. Também chamados por muitos de dados de painel (do inglês *panel data*), são dados quando se têm observações de séries temporais sobre uma amostra de unidades individuais. Para ajustar um modelo de dados longitudinais com amostragem em tempos arbitrários é necessário considerar as razões pelas quais foram feitas observações nesses determinados tempos. Isto é análogo para o problema de dados faltantes onde o fato de uma particular observação estar faltando pode, implicitamente, dar informação sobre o valor daquela observação. Possíveis esquemas de observação incluem:

- observação fixa: Cada paciente é observado a intervalos fixos especificados com antecedência.

- observação aleatória: Os tempos reais variam aleatoriamente, independentes do estado atual da doença.
- cuidado médico: Os pacientes em estado severo têm acompanhamento mais rigoroso pelo médico. O próximo tempo real de observação é determinado com base no estado atual da doença.
- interesse do paciente: O paciente pode decidir pela necessidade de visitar o médico quando ele próprio avaliar que suas condições requerem acompanhamento.

Grüger et al.(1991) discutiu as condições sob as quais as informações sobre tempos arbitrários são necessárias. Se um modelo multi-estado é ajustado ignorando a informação disponível dos tempos arbitrários, então as conclusões podem não ser verídicas ou terem sofrido influências pela falta destas informações. Matematicamente, devido aos tempos arbitrários serem aleatórios, eles devem ser modelados junto com o processo observado X_t . Porém a situação ideal é onde a união da verossimilhança para o tempo e o processo deve ser proporcional à verossimilhança obtida se o tempo arbitrário foi fixado com antecedência. Então os parâmetros do processo podem ser calculados independentemente dos parâmetros do modelo ajustado. Em particular, isso ficou demonstrado que para esquemas de observação fixo, aleatório e de cuidado médico não são informativos e, o esquema de observação do interesse do paciente é informativo.

2.6 Tempo médio de permanência (*Mean Sojourn Time*)

O cálculo do tempo de permanência em um processo Markoviano é de grande importância. Como têm-se probabilidades de transição entre os estados, é também de interesse saber quanto tempo o indivíduo demora para passar de um estado para outro. Esse cálculo por exemplo, pode determinar o tempo que um paciente ainda tem de vida, considerando uma transição de um estado adjacente para outro absorvente, no caso, a morte.

Imagine clientes em uma fila de espera, ou qualquer tipo de atividade que envolva espera para se transitar de uma posição para outra e que seja importante e necessário o cálculo de permanência nessas posições, tanto quanto das probabilidades de transição. Sem dúvida que o tempo exato de espera não é passível de ser obtido, mas um tempo aproximado, uma média de tempo de trânsito de um estado para outro. Portanto *mean sojourn time* é o tempo médio de permanência em cada estado transiente.

Em um modelo de progressão unidirecional ela é estimada por $-\frac{1}{q_{rr}}$.

2.7 Função de Verossimilhança

A função de Máxima Verossimilhança é a função de densidade de probabilidade conjunta de uma amostra, vista como função do vetor de parâmetros desconhecidos θ .

Dada uma amostra, a intenção que se tem, é obter uma estimativa dos valores dos parâmetros desconhecidos, um vetor desses parâmetros estimados de θ que maximize a probabilidade de obtenção da amostra específica (y_1, y_2, \dots, y_n), ou seja, deseja-se encontrar o valor de θ que maximize a função de verossimilhança.

Para um modelo multi-estado em tempo contínuo a verossimilhança é calculada a partir da matriz de probabilidades de transição, $P(t)$. Para um processo homogêneo no tempo, o elemento (r, s) de $P(t)$ é a probabilidade de estar no estado s no tempo $t+u$ no futuro, dado que o estado no tempo t estava no estado r . A série de tempos $(t_{i1}, t_{i2}, \dots, t_{ij})$ e os correspondentes estados $S_i(t_{i1}), S_i(t_{i2}), \dots, S_i(t_{ij})$ são os dados para o indivíduo i . Considere um modelo multi-estado geral, com um par de sucessivos estados observados da doença $S(t_j), S(t_{j+1})$ nos tempos t_j, t_{j+1} . A contribuição deste par de estados para a verossimilhança é:

$$L_{ij} = P_{s(t_i)s(t_{j+i})}(t_{j+i} - t_j)$$

em que $L_{i,j}$ é o elemento da matriz de transição $P(t)$ na linha $S(t_j)$ e na coluna $S(t_{j+1})$, calculados em $t = t_{j+1} - t_j$. A verossimilhança completa $L(Q)$ é o produto de todos os termos $L_{i,j}$ sobre todos os indivíduos e todas as transições, e depende da matriz desconhecida de transição Q , que é usada para determinar $P(t)$.

$$L(Q) = \prod_{i=1}^n \prod_{j=1}^{m-1} P_{S(t_{j+i})|S(t_i)}(t_{j+i} - t_j); \quad 1 \leq i \leq n; \quad 1 \leq j \leq m$$

Fazendo-se uma analogia, a função de máxima verossimilhança determina os θ desconhecidos maximizando a probabilidade daqueles indivíduos fazerem parte daquela amostra. Para o MSM a função de máxima verossimilhança vai determinar a probabilidades dos indivíduos transitarem de um estado ao outro, assim determinando a matriz de transição desconhecida.

3 MSM

O msm, pacote desenvolvido por Christopher Jackson, presente no *software* livre *R*, permite ajustar Modelos Markovianos Multi-estado. Esse pacote contém um conjunto de dados chamado na versão mais atual, de *CAV*, em versões anteriores, tinha o nome de *HEART*, o qual será usado para apresentar a técnica.

3.1 O conjunto de dados do CAV

Os dados descrevem a história de 622 pacientes que efetuaram um transplante de coração, enquanto monitoram o desenvolvimento de uma complicação de pós-transplante chamado alografia vascular-coronária(CAV).

3.2 Instalando e carregando o MSM

Para poder utilizar o MSM, primeiramente tem-se de tê-lo instalado no software R. Trata-se de um pacote adicional, que não vem na base do software. Para instalá-lo pode-se utilizar o comando `> install.packages("msm")`, com o nome do pacote entre aspas. Ao executarmos esse comando sera pedido para que seja escolhido um *mirror*, que trata-se de um servidor de onde é feito o download.

Após instalado o pacote, tem-se que torna-lo disponível para o uso no software R. Os comandos mais usuais são `> library(msm)` e `> require(msm)`

O pacote *msm* utiliza simultaneamente o pacote *mvtnorm*, para computar técnicas multivariadas de cálculos entre matrizes, portanto é necessário também a instalação do *mvtnorm* para correto funcionamento do *msm*. Não é necessário carregá-lo porém é preciso que esteja instalado.

3.3 Obtendo os dados

Os dados de um processo markoviano assim como o conjunto *CAV* são especificados como uma série de observações, agrupados por um indivíduo, objeto ou no caso, por paciente, e devem, impreterivelmente, ter as seguintes informações:

- O tempo da observação
- O estado observado no processo
- O identificador do indivíduo

Como vê-se-á a seguir, o conjunto *CAV* possui essas informações.

Para carregarmos os dados do conjunto *CAV* podemos usar o comando `> data(cav)`, este comando deixa disponível o conjunto de dados para que seja usado. Caso não haja outros conjuntos de dados carregados, pode-se digitar simplesmente o nome de uma variável do conjunto que essa será reconhecida, desde que o conjunto esteja anexado ao software. Para anexar um conjunto de dados usa-se o comando `> attachcav` e para desanexar, o comando `> detach(cav)`. Caso haja mais de um conjunto carregado mas com nomes de variáveis diferentes, também é possível chamar uma variável somente com o nome dela. Caso haja variáveis homônimas é necessário discriminar a qual conjunto de

dados pertence. Por exemplo, o conjunto *CAV* possui 9 variáveis. Um comando para poder visualizar os nomes das variáveis é o *names(cav)* que tem a saída abaixo.

```
> names(cav)
[1] "PTNUM"    "age"      "years"    "dage"     "sex"      "pdiag"
[7] "cumrej"   "state"    "firstobs"
```

Caso deseje-se visualizar a variável *years* por exemplo, pode-se digitar no *prompt* de comando a palavra *years*.

```
> years
 [1] 0.000000000  1.002739726  2.002739726  3.093150685  4.000000000
 [6] 4.997260274  5.854794521  0.000000000  1.189041096  2.008219178
[11] 2.991780822  0.000000000  1.002739726  2.010958904  3.010958904
[16] 4.008219178  4.991780822  6.021917808  7.019178082  8.016438356
[21] 8.901369863  0.000000000  1.013698630  2.394520548  3.030136986
[26] 4.068493151  4.953424658  6.002739726  8.002739726  8.065753425
[31] 0.000000000  1.013698630  2.123287671  3.041095890  0.000000000
[36] 1.967123288  3.005479452  4.123287671  5.005479452  5.983561644
```

Caso houvesse mais de uma variável chamada *years* poderia-se usar o comando *> cav\$years*, que aí estaria discriminando a qual conjunto de dados pertence a variável.

3.4 Conhecendo as variáveis

O conjunto *CAV* contempla 622 pacientes. *PTNUM* é o Identificador. Aproximadamente à cada ano após o transplante, cada paciente tem um angiograma, no qual *CAV* - alografia vascular coronária - pode ser diagnosticada. Na dimensão do conjunto de dados, vê-se que são 2846 linhas, porém há repetições no indentificador já que os paciente são observados várias vezes e transitam entre estados. Desconsiderando as repetições vê-se que são 622 pacientes no estudo . O resultado do teste de deteriorização das paredes coronárias está na variável *state*, com valores possíveis 1, 2, 3 representando *CAV*-livre, *CAV*-leve e *CAV*-moderada ou grave, respectivamente. O valor 4 é registrado como falecimento. A variável *years* dá o momento do teste, em anos desde o transplante cardíaco. Outras variáveis incluem *age*(idade do transplantado), *dage*(idade do doador), *sexo*(0 = masculino, 1 = feminino), *pdiag* (diagnóstico primário, ou motivo para o transplante - *ihd* representando a doença isquêmica cardíaca e *idc* cardiomiopatia dilatada idiopática), *cumrej*(número acumulado de episódios de rejeição), e *firstobs* que assume o valor 1 quando a observação corresponde ao transplante do paciente (a primeira observação), e 0 quando a observação corresponde a um posterior angiograma.

3.5 Sumarizando os Dados

Toda análise estatística de qualquer conjunto de dados deve começar pela descrição dos dados. A estatística descritiva ajuda a ter-se uma visão geral do conjunto de dados e até a pressupor resultados. Costuma-se dizer que quando a inferência não diz nada, uma boa análise descritiva pode-se trazer as respostas.

Abaixo, são apresentados comandos que sumarizam os dados. Alguns deles repetem o que já é conhecido e outros dão indicativos do que esperar dos resultados inferenciais.

```
> dim(cav)
[1] 2846    9
```

```
> summary(cav$firstobs==1)
  Mode  FALSE   TRUE
logical 2224   622
```

Ao utilizar o comando *dim*, se o indivíduo não souber nada a respeito dos dados, ele pode ser levado a pensar que são 2846 pacientes e 9 variáveis. Isso porque, como é sabido, há repetições desse indivíduos. Com o comando logo abaixo do *dim* vê-se que na realidade o conjunto tem 622 pacientes. Como relatado anteriormente, ao chamar a variável *firstobs* do conjunto *cav*, o que espera-se é que apareçam indicadores 0 e 1, com 1 designando quando se deu a primeira observação e 0 quando a observação ocorreu devido a um angiograma. Obviamente o número 1 não se repete para o mesmo indivíduo, sendo assim, tem-se que a quantidade de repetições do número 1, é o número total de pacientes do estudo e o 0 pode se repetir enquanto o paciente continuar a ter acompanhamento por angiogramas, até o final do estudo. Executando o comando *summary* e determinando que a sumarização da variável *firstobs* será para que seja determinado os valores iguais a 1, pode-se verificar pela resposta do comando que o conjunto de dados possui 622 observações da variável *firstobs* iguais a 1 e portanto 622 indivíduos. Pode-se obter o mesmo resultado utilizando o comando *table* que reproduz uma tabela com argumentos x e y informados, dois vetores ou uma variável nominal.

```
> table(cav$firstobs==1,cav$state==1)

      FALSE TRUE
FALSE   807 1417
TRUE     0   622
```

Ainda sumarizando os dados, quando estes se tratarem de dados com muitas colunas e linhas, pode-se visualizar somente parte deles, seja pelas últimas ou primeiras obser-

vações. Os comandos *head* e *tail* mostram respectivamente as primeiras e últimas 6 linhas e todas as colunas. Pode-se também determinar quantas linhas serão mostradas, pela adição do argumento *n*, assim como no comando abaixo.

```
> head(cav, n=10)
  PTNUM   age   years dage sex pdiag cumrej state firstobs
1 100002 52.49589 0.000000  21  0  IHD      0     1      1
2 100002 53.49863 1.002740  21  0  IHD      2     1     0
3 100002 54.49863 2.002740  21  0  IHD      2     2     0
4 100002 55.58904 3.093151  21  0  IHD      2     2     0
5 100002 56.49589 4.000000  21  0  IHD      3     2     0
6 100002 57.49315 4.997260  21  0  IHD      3     3     0
7 100002 58.35068 5.854795  21  0  IHD      3     4     0
8 100003 29.50685 0.000000  17  0  IHD      0     1     1
9 100003 30.69589 1.189041  17  0  IHD      1     1     0
10 100003 31.51507 2.008219  17  0  IHD      1     3     0
```

As primeiras 10 linhas do conjunto *cav* podem passar muita informação. Nota-se que *PTNUM* se repete várias vezes, como já dito essa variável é o indicador do indivíduo. Vê-se que na primeira linha, *firstobs* possui o valor 1, e só vai apresentar novamente esse valor na oitava linha, quando *PTNUM* muda, ou seja, trata-se de outro indivíduo. O que pode-se perceber também, apesar de que só seja apresentado dados de dois pacientes, é que na primeira observação da variável *state* para cada paciente, o estado é 1. Isso indica que todos os pacientes começaram com o transplante e teoricamente estavam saudáveis novamente.

O conjunto *cav* possui um estado absorvente, como relatado antes esse estado é a morte do indivíduo. O estado 4 é um estado que, ao indivíduo entrar nesse estado não há mais transição para outro estado, portanto obviamente esse estado não se repete para o mesmo paciente no conjunto de dados. Com o comando abaixo pode-se saber quantos pacientes morreram no final do estudo, que totalizam 251 indivíduos.

```
> summary(cav$state==4)
  Mode  FALSE   TRUE
logical 2595   251
```

Uma forma útil para resumir os dados multi-estado é em uma tabela de frequências de transição entre os estados. O *msm* possui uma função para cálculo desse tipo de tabela com o comando *statetable.msm*. Essa função calcula a contagem de todos os indivíduos, para cada estado *r* e *s*, o número de vezes que um indivíduo tinha uma observação do estado *r* seguido por uma observação do estado *s*. O comando possui três argumentos. O primeiro é o vetor de estados dos indivíduos envolvidos, no caso do *cav*, pacientes. O

segundo é o vetor que indexa o vetor de estados, nesse caso o PTNUM que é o indentificador do indivíduo e o terceiro argumento destina-se a informar de qual conjunto de dados serão coletados os vetores, variáveis. Por *default* a função está como *NULL* no terceiro argumento. Isso quer dizer que a função pegará os vetores do conjunto de dados que estiver carregado, caso haja dois conjuntos e com nomes homônimos de variáveis é necessário informar qual conjunto está sendo usado.

```
> statetable.msm(state, PTNUM, data = cav)
      to
from   1    2    3    4
  1 1367  204   44  148
  2   46  134   54   48
  3    4   13  107   55
```

Todos os pacientes entraram no estudo começaram no estado 1, que condizia com o transplante. A partir de então houve transições entre os estados. portanto o somatório das observações da tabela acima somam 2846 transições menos 622 pois todos começaram no estado 1.

3.6 Ajustando o Modelo MSM

Para ajustar um modelo multi-estado é necessário utilizar a função `textitmsm`. Tal função possui alguns argumentos imprescindíveis. A fórmula pela qual será estimada o modelo, obtida com o vetor estado em função do vetor tempo. No caso do exemplo abaixo ela é feita com o vetor de estados (*state*) em função do vetor de anos, ou seja, a quantidade de tempo medida em anos que o paciente levou para ser observado. O segundo argumento é o vetor de indentificação, caso esse não seja informado a função irá assumir que todas as informações são do mesmo indivíduo, o que no caso do *cav* não é verdade pois há várias repetições das visitas dos pacientes. É interessante informar de qual conjunto de dados está sendo extraída as variáveis, faz-se isso no terceiro argumento. A *qmatrix*, o quarto argumento da fórmula, é uma matriz intensidade de transição inicial, calculada e apresentada abaixo pela função atribuída à variável *qm* criada. Repare que uma vez o processo no estado 4, não há mais transições, devido ao estado 4 ser um estado absorvente, morte. Esse estado, se houver, deve ser informado justamente no parâmetro *death*, mesmo que o estado absorvente não seja propriamente a morte do indivíduo. Esses são os argumentos básicos para se ajustar um modelo *msm*.

Primeiramente, para ajustar o modelo, é necessário criar uma matriz intensidade de transição inicial.

Conforme a matriz Q abaixo, vê-se os estados possíveis de transição. Como a diagonal q_{rr} é inversamente proporcional à soma das intensidades q_{rs} e a soma das linhas deve ser igual a zero chega-se a uma matriz intensidade de transição inicial. Vale ressaltar que independente do valor colocado nos estados transitáveis, o resultado será o mesmo.

```
qm <- rbind(c(0, 0.25, 0, 0.25), c(0.166, 0, 0.166, 0.166),
+ c(0, 0.25, 0, 0.25), c(0, 0, 0, 0))
```

```
> qm
      [,1] [,2] [,3] [,4]
[1,] 0.000 0.25 0.000 0.250
[2,] 0.166 0.00 0.166 0.166
[3,] 0.000 0.25 0.000 0.250
[4,] 0.000 0.00 0.000 0.000
```

ou

```
qm <- rbind(c(0, 1, 0, 1), c(1, 0, 1, 1),
+ c(0, 1, 0, 1), c(0, 0, 0, 0))
```

```
> qm
      [,1] [,2] [,3] [,4]
[1,]    0    1    0    1
[2,]    1    0    1    1
[3,]    0    1    0    1
[4,]    0    0    0    0
```

Uma melhor estimação para essa matriz intensidade de transição pode ser obtida através da função *crudeinits.msm*. Essa função possui os argumentos :

- formula: deve ser informado os vetores de estados e tempo de observação, por exemplo estado tempo.
- subject: informa-se o vetor que identifica cada indivíduo e seus estados.
- data: informa-se o nome do conjunto de dados a que pertencem o vetor de estado e tempo.
- qmatrix: uma matriz intensidade de transição inicial.
- censor: um estado, vetor que indica censura.

- `sensor.states`: informa-se os estados adjacentes que representam censuras

```
> qm0 <- crudeinits.msm(state ~ years, PTNUM, data=cav, qmatrix=qm)
```

```
> qm0
```

	[,1]	[,2]	[,3]	[,4]
[1,]	-0.1173149	0.06798932	0.0000000	0.04932559
[2,]	0.1168179	-0.37584883	0.1371340	0.12189692
[3,]	0.0000000	0.04908401	-0.2567471	0.20766310
[4,]	0.0000000	0.00000000	0.0000000	0.00000000

A partir dessa estimativa da matriz Q , pode-se ajustar um modelo para os dados de *cav*. O comando *msm* para o ajuste do modelo possui vários argumentos, porém serão citados os principais.

- `formula`: deve ser informado os vetores de estados e tempo de observação, por exemplo estado tempo.
- `subject`: informa-se o vetor que identifica cada indivíduo e seus estados.
- `data`: informa-se o nome do conjunto de dados a que pertencem o vetor de estado e tempo.
- `qmatrix`: uma matriz intensidade de transição inicial.
- `death`: estado ou vetor de estados absorventes.

```
> heart.msm <- msm(state ~ years, subject = PTNUM, data = cav,
+ qmatrix = qm0, death = 4)
```

```
> heart.msm
```

Call:

```
msm(formula = state ~ years, subject = PTNUM, data = cav,
qmatrix = qm0, death = 4)
```

Maximum likelihood estimates:

Transition intensity matrix

	State 1	State 2
State 1	-0.1692 (-0.1889,-0.1515)	0.1267 (0.1104,0.1455)
State 2	0.2188 (0.1622,0.2952)	-0.6037 (-0.704,-0.5177)
State 3	0	0.1282 (0.0776,0.2118)
State 4	0	0

	State 3	State 4
State 1	0	0.04244 (0.03402,0.05294)
State 2	0.3496 (0.2769,0.4413)	0.03528 (0.007199,0.1729)
State 3	-0.432 (-0.5501,-0.3392)	0.3038 (0.2335,0.3952)
State 4	0	0

-2 * log-likelihood: 3968.95

O resultado do modelo apresenta uma matriz intensidade de transição de máxima verossimilhança, onde se vê que as maiores intensidades são as transições do estado 2 para o estado 3 e do estado 3 para o estado 4. Além disso, o modelo informa as estimativas das diagonais q_{rr} que serão úteis para cálculo do tempo médio esperado.

```
> {heart.msm$qmodel}$qmatrix
```

	State 1	State 2	State 3	State 4
State 1	-0.1173149	0.06798932	0.0000000	0.04932559
State 2	0.1168179	-0.37584883	0.1371340	0.12189692
State 3	0.0000000	0.04908401	-0.2567471	0.20766310
State 4	0.0000000	0.00000000	0.0000000	0.00000000

3.7 Funções de Extração

Informações podem ser extraídas do MSM após processado o ajuste do modelo, por exemplo, plotagem de gráficos ou tabelas em geral. Algumas dessas funções nos mostram interessantes informações do modelo ajustado. Normalmente o que se faz após obtenção do modelo, é executar a função `> summary(heart.msm)` e analisar seus resultados, porém nesse caso de modelos multi-estados a análise torna-se um pouco rebuscada. Dado isso o melhor a se fazer é utilizar funções que extraíam as informações necessárias.

3.7.1 Média de tempo em cada estado

A função que calcula o tempo médio estimado é `sojourn.msm`. Ela extrai o tempo médio estimado do paciente em cada estado transiente, para um dado valor da covariável ajustada. A equação para seu cálculo segue abaixo segue abaixo:

$$s = -\frac{1}{q_{rr}}$$

em que, q_{rr} é a r -ésima entrada na diagonal da matriz de intensidade de transição. O método por *default* do R para cálculo, é o *deltamethod* para encontrar o erro padrão aproximado. O intervalo de confiança é estimado assumindo normalidade na escala log.

Computacionalmente temos algumas maneiras de se calcular. Extrair diretamente do modelo: ajustado:

```
> heart.msm$sojourn
      estimates      SE      L      U
State 1  5.911427 0.3324019 5.294548 6.600179
State 2  1.656461 0.1298807 1.420495 1.931624
State 3  2.314949 0.2855606 1.817780 2.948094
```

ou ainda, utilizar a função do pacote *MSM*.

```
> sojourn.msm(heart.msm)
      estimates      SE      L      U
State 1  5.911427 0.3324019 5.294548 6.600179
State 2  1.656461 0.1298807 1.420495 1.931624
State 3  2.314949 0.2855606 1.817780 2.948094
```

As estimativas do indivíduo permanecer em cada estado, medida em anos, podem ser observadas na segunda coluna, *estimates*. As colunas subsequentes apresentam o desvio padrão, limite inferior e limite superior .

3.7.2 Probabilidades de transição

Após ajustar o modelo e obter a matriz $Q(t)$, então é possível obter a matriz $P(t)$. A obtenção de probabilidades de determinado indivíduo transitar de um estado para outro é quase que o intuito do estudo em si. Por exemplo na área médica, quando um paciente está com uma doença grave e sabe que suas chances de recuperação são limitadas, os questionamentos são sempre do tipo: - Qual a probabilidade de recuperação? Com isso o cálculo dessas probabilidades de transição em um determinado tempo tornam-se importantíssimas. O comando `pmatrix.msm` fornece a probabilidade de transição para um dado intervalo de tempo. Pode-se usar por exemplo, as matrizes probabilidade de transição a 5 anos e a 15 anos respectivamente mostradas abaixo

```
> pmatrix.msm(heart.msm, t=5)
      State 1   State 2   State 3   State 4
State 1 0.51956378 0.13840544 0.09352276 0.2485080
State 2 0.23899466 0.13958581 0.18785684 0.4335627
State 3 0.05921954 0.06888738 0.17332264 0.6985704
State 4 0.00000000 0.00000000 0.00000000 1.0000000
```

```
> pmatrix.msm(heart.msm, t=15)
      State 1   State 2   State 3   State 4
State 1 0.18903857 0.06259554 0.06294398 0.6854219
State 2 0.10808823 0.03807559 0.04214087 0.8116953
State 3 0.03985675 0.01545312 0.01937589 0.9253142
State 4 0.00000000 0.00000000 0.00000000 1.0000000
```

Observa-se que a probabilidade de um transplantado que está no estado 3 morrer em 15 anos é aproximadamente 33% maior do que a probabilidade de um transplantado morrer em 5 anos.

3.7.3 Gráfico de Sobrevivência

Em estudos de doenças crônicas uma importante utilidade do MSM é a predição da probabilidade de sobrevivência de doentes em estados cada vez mais grave da doença, para o mesmo tempo t no futuro. Isto pode ser obtido diretamente da matriz de probabilidade de transição $P(t)$.

O comando `> plot.msm()` aplicado a um modelo `msm` ajustado retorna curvas de sobrevivência para os estados transientes do processo. Pode-se pré-determinar quais estados

serão plotados, caso seja omitido a informação ocorre conforme supracitado. Quando plota-se o modelo ajustado ao conjunto *cav* tem -se o seguinte gráfico:

```
plot.msm(heart.msm, legend.pos = c(8, 1), xlab = "Tempo(anos)",  
ylab = "Prob. de Sobrevivência")
```

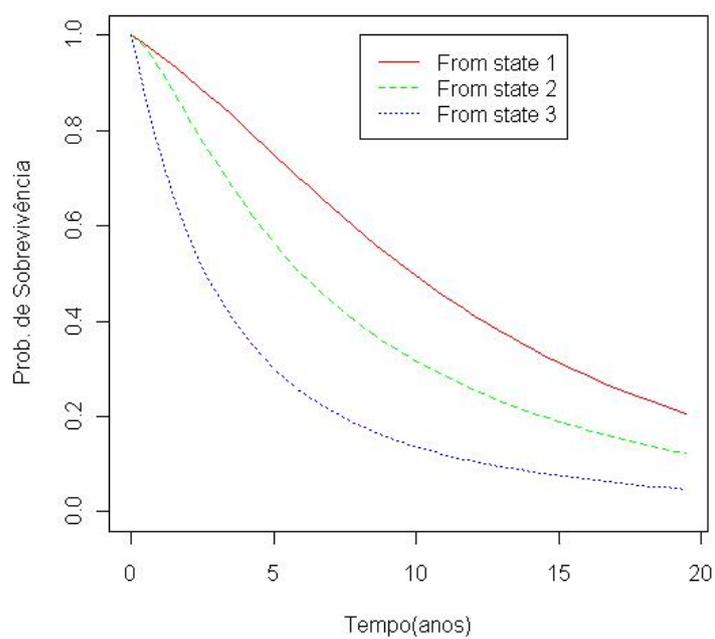


Figura 5: Gráfico de Sobrevivência vs Tempo sem o estado absorvente.

O plot mostra que a sobrevivência para quem, após o transplante, continua no estado 1 é maior do que a de quem inicia em outro estado transiente.

4 CONSIDERAÇÕES FINAIS

O MSM é um modo alternativo e muitas vezes o mais indicado em se usar, se tratando de problemas de doenças onde se quer extrair informações sobre as transições entre os estados. Há muitas técnicas mais conhecidas que o msm, mas que não utilizam todas as transições que o indivíduo faz durante o estudo. Isso se deve a essas técnicas terem interesse somente na falha desse indivíduo e não no que ocorre com ele no decorrer do estudo, como Análise de Sobrevivência e Regressão Logística.

Apesar de ser uma teoria com bibliografia composta basicamente de artigos, muitos autores procuram escrever e pesquisar sobre modelos markovianos multi-estados devido a sua facilidade em se tratar dados longitudinais.

O objetivo desse estudo era estudar uma técnica estatística não vista durante o curso de graduação e além de tudo formar material para servir de suporte ao pacote estatístico msm e suas funções básicas.

Foram apresentadas neste estudo uma revisão sobre a teoria de acordo com as delimitações, que não previam a inclusão de erros de classificação e nem adição de covariáveis, e também foi apresentada a aboradagem prática dessa técnica via *software R*, que nesse trabalho tem os comandos explicados de forma que o leitor possa seguir os passos da análise e ajuste do modelo multi-estado de Markov.

5 REFERÊNCIAS BIBLIOGRÁFICAS

- C. H. Jackson et al. Multistate Markov models for disease progression with classification error. *Journal of the Royal Statistical Society, Series D: The Statistician*, 52(2):1–17, 2003.
- TARTU. Practical: Multi-State Markov Models. SPE(2006).
- NOGUEIRA, Fernando. Modelagem e Simulação - Cadeias de Markov. Disponível em <<http://www.engprod.ufjf.br/fernando/epd042/cadeiaMarkov.pdf>> Acesso em 13/03/2008.
- JACKSON, Christopher. The msm Package. Disponível em < <http://cran-r.c3sl.ufpr.br/web/packages/msm/msm.pdf> > Acesso em 13/03/2008.
- JACKSON, Christopher. Multi–state Markov modelling with R. Disponível em <<http://www.ci.tuwien.ac.at/Conferences/useR-2004/abstracts/Jackson.pdf>> Acesso em 13/03/2008.
- R.C. Gentleman et al. Multi-State Markov Models for Anlysing Incomplete Disease History Data with Illustrations for Disease. , vol 13, 805-821 (1994).
- Jackson, C. H. and Sharples, L. D. (2002) Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. *Statist. Med.*, 21, 113?128.
- Longini, I. M., Clark, W. S., Byers, R. H., Ward, J. W., Darrow, W. W., Lemp, G. F. and Hethcote, H. W. (1989) Statistical analysis of the stages of HIV infection using a Markov model. *Statist. Med.*, 8, 831?843.
- Grüger, J., Kay, R. and Schumacher, M. (1991) The validity of inferences based on incomplete observations in disease state models. *Biometrics*, 47, 595?605.
- Cox, D. R. and Miller, H. D. (1965) *The Theory of Stochastic Processes*. London:Chapman and Hall.