Modelos multi-estados markovianos com erros de classificação: aplicação ao problema da morte súbita do citrus

Silvia Shimakura
UFPR

Caxambu
25 Julho 2006

Agradecimentos

- FUNDECITRUS
- Renato Bassanezzi (FUNDECITRUS)
- Paulo Justiniano Ribeiro Jr (UFPR)
- Elias Krainski (UFMG)

Conteúdo

- Suposições em análise de sobrevivência
- Motivação: Problema da MSC
- Modelos multi-estados
- Modelos multi-estados markovianos
- Aplicação
- Conclusões

Suposições em análise de sobrevivência clássica

- Sujeitos são "saudáveis" ou "doentes", sem estágios intermediários

- A doença é *irreversível*, ou requer intervenção para cura

- O tempo de incidência da doença é conhecido exatamente

- A doença é diagnosticada sem erro

Estas suposições são verdadeiras para *morte* e muitas *doenças crônicas*.

A doença é dicotômica?

- A doença pode ser precedida de uma fase *sub-clínica* antes de mostrar seus sintomas.

  AIDS declínio nas contagens de CD4
  Câncer lesões pré-cancerígenas

- Ou a doença pode ser classificada em *graus de severidade* (leve, moderada, severa)

Exemplo: Modelo para Adenocarcinoma gástrica

- O desenvolvimento de câncer gástrico demora muitas décadas.

$$\text{Normal} \leftrightarrow \text{GS} \leftrightarrow \text{GC} \leftrightarrow \text{GCA} \leftrightarrow \text{MI} \leftrightarrow \text{Displasia} \rightarrow \text{Câncer}$$

- Uma gastrocopia é necessária para diagnose de lesões pré-câncer.

Exemplo: Modelo para câncer cervical

- Câncer cervical é precedido por neoplasia cervical intraepitelial (CIN)

$$\text{Normal} \leftrightarrow \text{CIN I} \leftrightarrow \text{CIN II} \leftrightarrow \text{CIN III} \rightarrow \text{Câncer}$$

- O propósito de um programa de varredura é detectar e tratar CIN.

Quando a doença ocorre?

- Pode ser necessária uma *visita clínica* para diagnose da doença.

  ⋆ exame médico, ou

  ⋆ testes laboratoriais em amostras de sangue, ou

  ⋆ biópsia feita por patologista

- Não se sabe o que acontece entre visitas consecutivas (censura intervalar)

Conhecemos o verdadeiro estatus da doença?

- Biomarcadores podem ter um *erro aleatório*
- Diagnóstico correto pode requerer procedimento *invasivo*
- Diagnóstico preciso pode requerer um painel de consenso

Motivação: Morte Súbita dos Citrus

Da epidemiologia de gente para a epidemiologia de plantas!

MSC: tem todas as características mencionados anteriormente!

Terminologia para doenças multi-estado

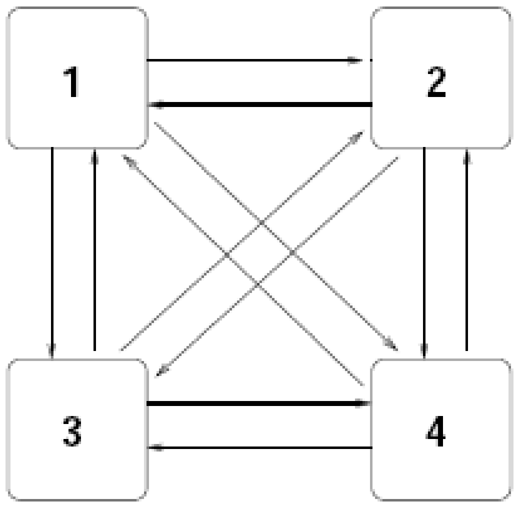
Terminologia de análise de sobrevivência pode ser convertida:

Estatus caso/controle	$\longrightarrow$	Estatus da doença
Incidência da doença	$\longrightarrow$	Transições entre estados
Taxa de incidência	$\longrightarrow$	Intensidades de transição

Modelo multi-estados

- Indivíduo que se move através de uma série de estados $S(t)$ em tempo contínuo.

- **Exemplo**: Modelo multi-estado em tempo contínuo com 4 estados



- As flechas apontam quais transições são possíveis entre estados.

Intensidades de transição

- $q_{rs}(t, z(t))$ governam a *transição* entre os estados r e s e o *tempo* de mudança.

- Pode depender do tempo do processo t, ou

- de um conjunto de covariáveis (tempo-dependentes) $z(t)$

- Representa o risco instantâneo de transição do estado r para s

$$q_{rs}(t, z(t)) = \lim_{\delta t \to 0} Pr(S(t + \delta t) = s | S(t) = r) / \delta t$$

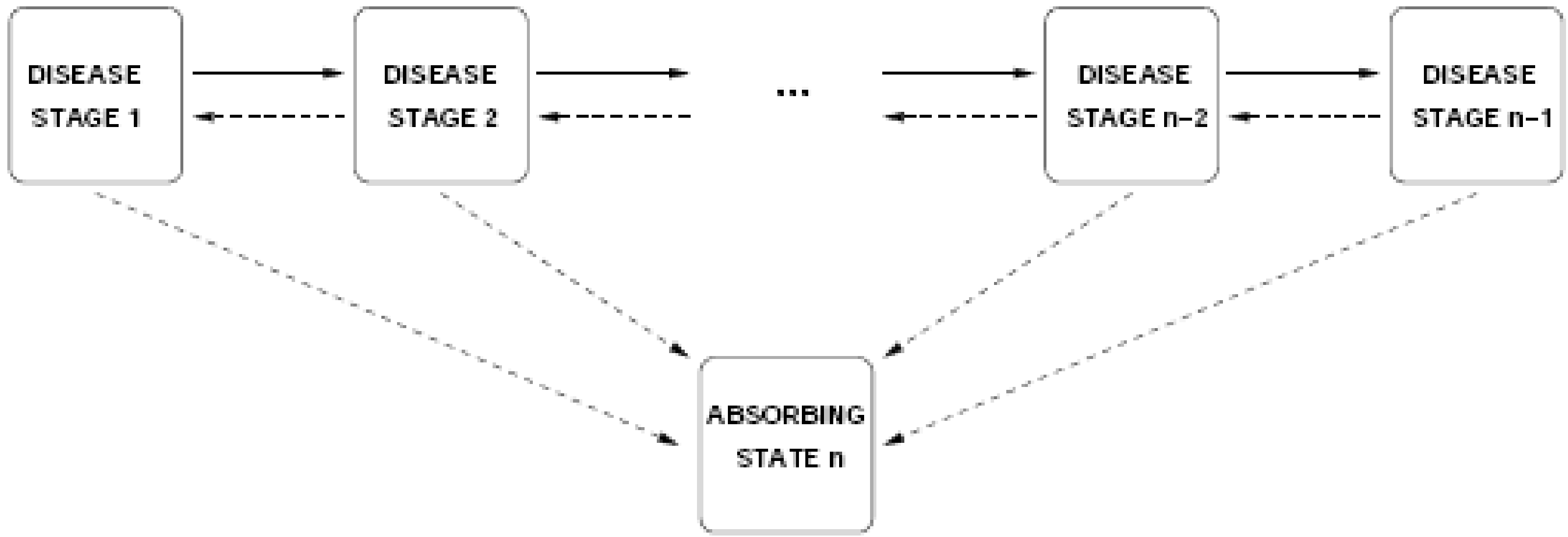
Intensidades de transição

- As intensidades formam uma matriz Q com elementos da diagonal $q_{rr} = -\sum_{s \neq r} q_{rs}$

- Ajuste do modelo multi-estado consiste em *estimar* a matriz de transição.

- **Modelo multi-estado markoviano** assume que evoluções futuras dependem somente do estado atual.

Modelo geral para progressão de doenças



- Representa uma série de estados sucessivamente mais graves e um estado "absorvente"

- O paciente pode avançar para ou regredir de estados adjacentes ou morrer a partir de qualquer estado.

- As observações do estado $S(t)$ são feitas em qualquer tempo arbitrário t (que pode variar entre indivíduos).

Matriz de transição para modelo de progressão de doenças

- Os estágios da doença podem ser modelados como um processo markoviano homogêneo em tempo contínuo, com matriz de transição Q

$$Q = \begin{pmatrix} q_{11} & q_{12} & 0 & 0 & \cdots & q_{1n} \\ q_{21} & q_{22} & q_{23} & 0 & \cdots & q_{2n} \\ 0 & q_{32} & q_{33} & q_{34} & \ddots & q_{3n} \\ 0 & 0 & q_{43} & q_{44} & \ddots & q_{4n} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

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

Probablidade de transição

- Verossimilhança é calculada da matriz de *probabilidade de transição* $P(t)$ (Kalbfleish & Lawless, 1985; Kay, 1986).

- Para processos homogêneos no tempo, o elemento (r, s) de $P(t)$

$$p_{rs}(t) = Pr\{S_i(t+u) = s | S_i(u) = r\}$$

  é a probabilidade de atingir o estado s no tempo $t + u$ dado que se está no estado r no tempo u.

- Depende dos parâmetros desconhecidos em Q através da relação $P(t) = \exp(tQ)$

- Expressões analíticas podem ser obtidas para modelos de até 5 estados.

Modelos markovianos latentes

- Para casos em que o processo da doença $S(t)$ não é diretamente observado ou é observado com erro. Por exemplo,

  - $\star$ diagnóstico passível de classificação incorreta, ou

  - $\star$ é tomada um medida indireta do estado da doença

- Adiciona-se um componente estocástico para descrever a relação entre o estado observado e o verdadeiro.

Matriz de erro de classificação

- Para um indivíduo i, tempo de observação t_{ij}, os estados observados O_{ij} são gerados condicionalmente aos estados verdadeiros S_{ij} de acordo com uma *matriz de erros de classificação* E, com elemento (r, s)

$$e_{rs} = Pr(O(t_{ij}) = s | S(t_{ij}) = r)$$

  (assumido independente do tempo t).

- alguns e_{rs} fixados: reflete conhecimento do processo de diagnose. Ex. e_{rs} pequeno para estados não adjacentes.

Efeitos de covariáveis

- Na intensidade de transição:

  ⋆ Riscos proporcionais para relacionar as intensidades de transição $q_{rs}(t)$ no tempo t a covariáveis $z(t)$

  $$q_{rs}\{t, z(t)\} = q_{rs} \exp\{\beta_{rs}^T z(t)\}$$

  ⋆ Cada transição pode ter uma intensidade de base diferente

  ⋆ Se as covariáveis são tempo-dependentes

  $$p_{rs}(t_2 - t_1) = p_{rs}\{t_2 - t_1, z(t_1)\}$$

- Na probabilidade de erro de classificação:

  ⋆ Modelo logístico para relacionar as probabilidades de erro de classificação e_{rs} a covariáveis $w(t)$

  $$\log \frac{e_{rs}(t)}{1 - e_{rs}(t)} = \gamma_{rs}^T w(t)$$

  .

Verossimilhança

- Método direto para cálculo de verossimilhança em tempo discreto ou contínuo baseado em produtos de matrizes (Macdonald e Zucchini, 1997).

- Contribuição do indivíduo i para a verossimilhança é

$$\begin{aligned} L_i &= Pr(O_{i1}, \cdots, O_{im_i}) \\ &= \sum Pr(O_{i1}, \cdots, O_{im_i} | S_{i1}, \cdots, S_{im_i}) Pr(S_{i1}, \cdots, S_{im_i}) \end{aligned}$$

  a soma é tomada sobre todos os possíveis caminhos dos estados latentes $S_{i1}, \cdots, S_{im_i}$.

Verossimilhança

- Assumindo independência condicional entre estados observados dado S; e

- Assumindo a propriedade markoviana

$$Pr(S_{ij}|S_{i(j-1)},\cdots,S_{i1}) = Pr(S_{ij}|S_{i(j-1)})$$

- $$\begin{aligned} L_i \ = \ & \sum_{S_{i1}} Pr(O_{i1}|S_{i1})Pr(S_{i1}) \sum_{S_{i2}} Pr(O_{i2}|S_{i2})Pr(S_{i2}|S_{i1}) \sum_{S_{i3}} Pr(O_{i3}|S_{i3})Pr(S_{i3}|S_{i2}) \\ & \cdots \sum Pr(O_{im_i}|S_{im_i})Pr(S_{im_i}|S_{i(m_i-1)}) \end{aligned}$$

- $Pr(O_{ij}|S_{ij}) = e_{S_{ij}O_{ij}}$

- $Pr(S_{i(j+1)}|S_{ij})$ é o elemento $(S_{ij}, S_{i(j+1)})$ de $P(t)$ avaliada em $t = t_{i(j+1)} - t_{ij}$

Verossimilhança

- L_i pode ser escrita como produto de matrizes

$$L_i = \mathsf{f}\, T_{i2} T_{i3} \cdots T_{im_i} \mathsf{1}$$

- $\mathbf{f}$: vetor de probabilidades de ocupação no estágio inicial $Pr(S_{i1})$.

- T_{ij} $(j = 2, \cdots, m_i)$: matriz $(n \times n)$ com elemento (r, s)

$$T_{ij} = e_{sO_{ij}} p_{rs}(t_{ij} - t_{i(j-1)})$$

- Se $S(t_j) = D$ é um estado absorvente (ex morte) medido sem erro e com tempo de entrada conhecido, a contribuição para verossimilhança é somada sobre os estados desconhecidos no instante anterior à morte.

- Métodos numéricos usados para maximizar a verossimilhança e obter erros padrão aproximados da inversa do Hessiano, estimada por diferenças finitas.

O processo de observação é informativo?

- Inferências podem ser viciadas se ignorado, análogo ao problema de dados perdidos/censurados informativos. (Gruger *et al.*(1991))

- Esquemas de observação possíveis:

  ⋆ Intervalos fixos (OK)

  ⋆ Intervalos aleatórios (OK)

  ⋆ Próximo tempo de observação determinado pelo estado atual (OK)

  ⋆ Tempo de observação determinado pelo paciente! (X)

Aplicação: Modelo com 4 estados e erros de classificação para MSC

- O risco de transição depende do estágio atual da doença e de características ambientais?

- Matriz intensidade de transição

$$Q = \begin{pmatrix} -q_{12} & q_{12} & 0 & 0 \\ 0 & -q_{23} & q_{23} & 0 \\ 0 & 0 & -q_{34} & q_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- Covariáveis: idade, estação do ano

Aplicação: Modelo com 4 estados e erros de classificação para MSC

- Matriz de erros de classificação

$$E = \begin{pmatrix} 1-e_{12} & e_{12} & 0 & 0 \\ e_{21} & 1-e_{21}-e_{23} & e_{23} & 0 \\ 0 & e_{32} & 1-e_{32}-e_{34} & e_{34} \\ 0 & 0 & e_{43} & 1-e_{43} \end{pmatrix}$$

- Covariáveis: estação do ano

- Vetor de probabilidades de ocupar os estados 1, 2, 3, 4:

$$Pr(S_{i1}) = (0,95; 0,02; 0,02; 0,01)$$

Resultados (pacote *msm* do R)

- Estimativas de parâmetros e I.C. 95%: Variedade Valência

Parâmetros		Transição
Outono	$\widehat{\beta}_{12}$	1.611 (1.469,1.753)
	$\widehat{\beta}_{23}$	0.275 (0.109,0.442)
	$\widehat{\beta}_{34}$	0.962 (0.770,1.154)
Inverno	$\widehat{\beta}_{12}$	1.746 (1.616,1.876)
	$\widehat{\beta}_{23}$	0.924 (0.782,1.067)
	$\widehat{\beta}_{34}$	0.862 (0.686,1.037)
Primavera	$\widehat{\beta}_{12}$	1.321 (1.182,1.459)
	$\widehat{\beta}_{23}$	1.120 (0.960,1.280)
	$\widehat{\beta}_{34}$	0.516 (0.305,0.727)
Idade	$\widehat{\beta}_{12}$	0.269 (0.257,0.281)
	$\widehat{\beta}_{23}$	-0.063 (-0.080,-0.045)
	$\widehat{\beta}_{34}$	0.168 (0.140,0.197)

Resultados (pacote *msm* do R)

- Estimativas de parâmetros e I.C. 95%: Variedade Valência

Parâmetros	Erros de classificação	
Outono	$\widehat{\gamma}_{12}$	-2.18 (-3.4,-0.96)*
	$\widehat{\gamma}_{21}$	-1.04 (-1.25,-0.82)*
	$\widehat{\gamma}_{23}$	-2.18 (-2.64,-1.71)*
	$\widehat{\gamma}_{32}$	-0.49 (-1.14,0.15)
Inverno	$\widehat{\gamma}_{12}$	-1.20 (-1.91,-0.49)*
	$\widehat{\gamma}_{21}$	-0.20 (-0.38,-0.01)*
	$\widehat{\gamma}_{23}$	-1.44 (-1.84,-1.03)*
	$\widehat{\gamma}_{32}$	-0.47 (-1.19,0.24)
Primavera	$\widehat{\gamma}_{12}$	0.02 (-0.43,0.48)
	$\widehat{\gamma}_{21}$	0.10 (-0.09,0.29)
	$\widehat{\gamma}_{23}$	0.02 (-0.26,0.30)
	$\widehat{\gamma}_{32}$	-1.60 (-2.50,-0.69)*

Comentários finais

- Modelos multi-estados com erros de classificação markovianos em tempos contínuos abrange uma grande classe de problemas

- História de doenças progressivas (ex MSC) podem ser facilmente explicadas usando os modelo descrito

- Implementação disponível: pacote msm no sistema R

- Interesse em abordagem Bayesiana