

Applications of continuous time hidden Markov models to the study of misclassified disease outcomes

Henrique Aparecido Laureano

est171 - Aprendizado de Máquina
Pós-Graduação em Estatística
UFMG - Universidade Federal de Minas Gerais

21 de novembro de 2016

Roteiro

Contextualizando

Modelo markoviano oculto

Aplicações

Leucoplasia pilosa

Infecção por HPV

Considerações finais

Contextualizando

Autores

Alexander Bureau

*Group in Biostatistics, School of Public Health,
University of California, Berkeley*

Stephen Shiboski

*Department of Epidemiology and Biostatistics,
University of California, San Francisco*

James P. Hughes

*Department of Biostatistics,
School of Public Health and Community Medicine,
University of Washington*



Publicação

Statistics in Medicine. 2003; **22**:441-462 (DOI: 10.1002/sim.1270)

Artigo

Aplicação de modelos markovianos ocultos de tempo contínuo em medidas longitudinais de doenças com resposta binária

A resposta da doença é representada por um processo markoviano homogêneo de dois estados em tempo contínuo

Artigo

Aplicação de modelos markovianos ocultos de tempo contínuo em medidas longitudinais de doenças com resposta binária

A resposta da doença é representada por um processo markoviano homogêneo de dois estados em tempo contínuo

Modelo markoviano oculto

Extensão de modelos markovianos que fornece uma maneira de lidar com possíveis erros de classificação devido ao processo de avaliação

Modelo markoviano oculto

Variáveis

T_{ij} → tempo cronológico das avaliações

Y_{ij} → medida dicotômica da resposta da doença

\mathbf{Z}_{ij} → vetor $1 \times p$ de covariáveis

em que i representa o indivíduo e j a avaliação

Variáveis

T_{ij} → tempo cronológico das avaliações

Y_{ij} → medida dicotômica da resposta da doença

Z_{ij} → vetor $1 \times p$ de covariáveis

em que i representa o indivíduo e j a avaliação

$X(t), t > 0$ → status atual subjacente da doença
(Modelado como um processo markoviano em tempo contínuo de dois estados)

Variáveis

$T_{ij} \rightarrow$ tempo cronológico das avaliações

$Y_{ij} \rightarrow$ medida dicotômica da resposta da doença

$Z_{ij} \rightarrow$ vetor $1 \times p$ de covariáveis

em que i representa o indivíduo e j a avaliação

$X(t), t > 0 \rightarrow$ status atual subjacente da doença
(Modelado como um processo markoviano em tempo contínuo de dois estados)

Y_1^j e T_1^j denotam a sequência de 1 até j de estados observados da doença e tempos de avaliação
(O índice i denotando o indivíduo é omitido)

Suposição markoviana para o processo oculto da doença

$$\begin{aligned}P[X(t_j)|X(t_1), \dots, X(t_{j-1}), Y_1^{j-1}, T_1^j = t_1^j] &= P[X(t_j)|X(t_{j-1}), T_{j-1}^j = t_{j-1}^j] \\ &= P_{x_{j-1}, x_j}(t_j - t_{j-1})\end{aligned}$$

Suposição markoviana para o processo oculto da doença

$$\begin{aligned}P[X(t_j)|X(t_1), \dots, X(t_{j-1}), Y_1^{j-1}, T_1^j = t_1^j] &= P[X(t_j)|X(t_{j-1}), T_{j-1}^j = t_{j-1}^j] \\ &= P_{x_{j-1}, x_j}(t_j - t_{j-1})\end{aligned}$$

Probabilidade de classificar corretamente ou incorretamente o estado da doença dado o verdadeiro estado

$$P[Y_j|X(t_1), \dots, X(t_j), Y_1^{j-1}, T_1^j = t_1^j] = P[Y_j|X(t_j), T_j = t_j] = f(y_j|x_j)$$

Processo oculto em tempo contínuo de dois estados

Processo oculto em tempo contínuo de dois estados

- ▶ Intensidade de aquisição da manifestação da doença, ν

Processo oculto em tempo contínuo de dois estados

- ▶ Intensidade de aquisição da manifestação da doença, u
- ▶ Intensidade de remoção, v

Processo oculto em tempo contínuo de dois estados

- ▶ Intensidade de aquisição da manifestação da doença, u
- ▶ Intensidade de remoção, v

Probabilidades de transição

$$P_{01}(t) = \frac{u}{u+v} (1 - \exp\{-(u+v)t\}) \quad P_{10}(t) = \frac{v}{u+v} (1 - \exp\{-(u+v)t\})$$

Intensidade de aquisição sobre o intervalo $[t_{j-1}, t_j]$ como função de p covariáveis

$$u_{i,j}(z_{ij}) = \exp \left\{ \theta_{u0} + \sum_{k=1}^p \theta_{uk} z_{ijk} \right\},$$

Intensidade de aquisição sobre o intervalo $[t_{j-1}, t_j]$ como função de p covariáveis

$$u_{i,j}(z_{ij}) = \exp \left\{ \theta_{u0} + \sum_{k=1}^p \theta_{uk} z_{ijk} \right\},$$

a expressão para a intensidade de remoção é similar

Intensidade de aquisição sobre o intervalo $[t_{j-1}, t_j]$ como função de p covariáveis

$$u_{i,j}(z_{ij}) = \exp \left\{ \theta_{u0} + \sum_{k=1}^p \theta_{uk} z_{ijk} \right\},$$

a expressão para a intensidade de remoção é similar

Verossimilhança para as m observações de um indivíduo

Intensidade de aquisição sobre o intervalo $[t_{j-1}, t_j]$ como função de p covariáveis

$$u_{i,j}(z_{ij}) = \exp \left\{ \theta_{u0} + \sum_{k=1}^p \theta_{uk} z_{ijk} \right\},$$

a expressão para a intensidade de remoção é similar

Verossimilhança para as m observações de um indivíduo

$$\begin{aligned} L(\theta) &= P[Y_1^m | T_1^m = t_1^m, Z_1^m = z_1^m, \theta] \\ &= \sum_{x_1 \dots x_m} P[Y_1^m, X_1^m | T_1^m = t_1^m, Z_1^m = z_1^m, \theta] \\ &= \sum_{x_1 \dots x_m} \pi_{x_1 | z_1} f(y_1 | x_1, z_1) \prod_{j=2}^m P_{x_{j-1}, x_j}(t_j - t_{j-1}) f(y_j | x_j, z_j) \end{aligned}$$

Aplicações



Leucoplasia pilosa

**Leucoplasia pilosa**

Leucoplasia pilosa



Leucoplasia pilosa

Lesão na língua causada pelo Vírus Epstein-Barr (VEB), acredita-se
Muito frequente em pacientes infectados pelo vírus HIV



Leucoplasia pilosa

Lesão na língua causada pelo Vírus Epstein-Barr (VEB), acredita-se
Muito frequente em pacientes infectados pelo vírus HIV

Estudo

334 homens do *San Francisco Men's Health Study*



Leucoplasia pilosa

Lesão na língua causada pelo Vírus Epstein-Barr (VEB), acredita-se Muito frequente em pacientes infectados pelo vírus HIV

Estudo

334 homens do *San Francisco Men's Health Study*

- ▶ Duração: fevereiro de 1987 até maio 1993
- ▶ Avaliação: Semestral

Leucoplasia pilosa

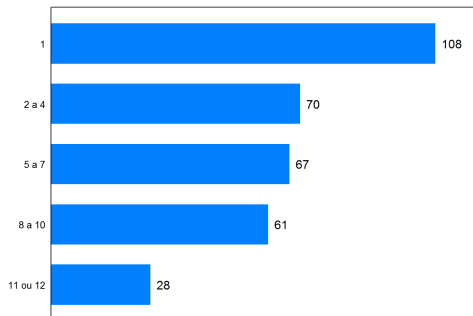
Lesão na língua causada pelo Vírus Epstein-Barr (VEB), acredita-se Muito frequente em pacientes infectados pelo vírus HIV

Estudo

334 homens do *San Francisco Men's Health Study*

- ▶ Duração: fevereiro de 1987 até maio 1993
- ▶ Avaliação: Semestral

Frequência do número de avaliações por indivíduo





Intensidades de transição (transição/mês) estimadas para um modelo markoviano e para o modelo markoviano oculto

	Aquisição			Remoção		
	Estimativa	Erro-padrão	Tempo esperado	Estimativa	Erro-padrão	Tempo esperado
Markov	0.0193	0.0016	51.8 meses	0.0837	0.0069	11.9 meses
HMM	0.0043	0.0011	233 meses	0.0130	0.0044	76.9 meses



Intensidades de transição (transição/mês) estimadas para um modelo markoviano e para o modelo markoviano oculto

	Aquisição			Remoção		
	Estimativa	Erro-padrão	Tempo esperado	Estimativa	Erro-padrão	Tempo esperado
Markov	0.0193	0.0016	51.8 meses	0.0837	0.0069	11.9 meses
HMM	0.0043	0.0011	233 meses	0.0130	0.0044	76.9 meses

Estimativas dos efeitos das covariáveis em modelos com apenas uma covariável

	Aquisição		Remoção	
	Razão de risco	Intervalo de 95% de confiança	Razão de risco	Intervalo de 95% de confiança
Candidíase oral	19.80	[8.16, 48.07]	1.70	[0.55, 5.25]
Diagnóstico de AIDS	2.11	[1.28, 3.50]	2.78	[1.61, 4.81]
100 unidades de CD4	0.66	[0.51, 0.84]	0.74	[0.48, 1.14]
Fumo	3.94	[1.40, 11.05]	1.35	[0.39, 3.08]
Aciclovir	3.17	[0.37, 27.2]	7.84	[0.56, 109.14]

	Falso positivo		Falso negativo	
	Razão de chance	Intervalo de 95% de confiança	Razão de chance	Intervalo de 95% de confiança
Candidíase oral	4.10	[1.81, 12.70]	0.83	[0.42, 1.63]



Infecção por HPV

**Infeção por HPV**



Infecção por HPV



Infecção por HPV

O HPV (papilomavírus humano) compreende uma diversidade grande de subtipos (mais de 70), que provocam desde o aparecimento de verrugas na pele e nas mucosas até doenças graves como o câncer do colo do útero

Os subtipos 16 e 18 são de alto risco para o câncer do colo do útero



Estudo



Estudo

663 mulheres em San
Francisco



Estudo

663 mulheres em San Francisco

- ▶ Início: 1990
- ▶ Avaliação:
Quadrimestral

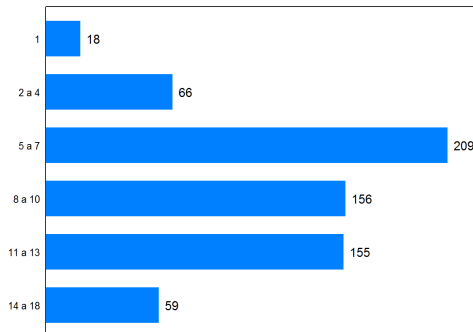


Estudo

663 mulheres em San Francisco

- ▶ Início: 1990
- ▶ Avaliação: Quadrimestral

Frequência do número de avaliações por indivíduo





Itensidades de transição (transição/mês) estimadas para o modelo markoviano oculto

Tipo de HPV	Aquisição			Remoção		
	Estimativa	Erro-padrão	Tempo esperado	Estimativa	Erro-padrão	Tempo esperado
16	0.0037	0.0010	270 meses	0.0569	0.0077	17.6 meses
18	0.0020	0.0006	500 meses	0.0875	0.0250	11.4 meses



Itensidades de transição (transição/mês) estimadas para o modelo markoviano oculto

Tipo de HPV	Aquisição			Remoção		
	Estimativa	Erro-padrão	Tempo esperado	Estimativa	Erro-padrão	Tempo esperado
16	0.0037	0.0010	270 meses	0.0569	0.0077	17.6 meses
18	0.0020	0.0006	500 meses	0.0875	0.0250	11.4 meses

Probabilidades de observação dado os estados ocultos

Tipo de HPV	P[negativo estado = 1]		P[positivo estado = 0]	
	Estimativa	Erro-padrão	Estimativa	Erro-padrão
16	0.1879	0.0427	0.0032	0.0030
18	0.3325	0.1167	0.0000	---

Considerações finais

Considerações finais

Considerações finais

Extensões interessantes a serem desenvolvidas:

- ▶ Possibilitar a interação entre múltiplos processos ocultos

Considerações finais

Extensões interessantes a serem desenvolvidas:

- ▶ Possibilitar a interação entre múltiplos processos ocultos

Modificações para possibilitar que intensidades de transições variem ao longo do tempo

Considerações finais

Extensões interessantes a serem desenvolvidas:

- ▶ Possibilitar a interação entre múltiplos processos ocultos

Modificações para possibilitar que intensidades de transições variem ao longo do tempo

Elaboração de técnicas de diagnóstico mais sofisticadas

Obrigado por seu tempo!