MODELO DE INTENSIDADE HÍBRIDO COM APLICAÇÃO EM EVENTOS RECORRENTES

Sabrina Luzia CAETANO*

RESUMO

Uma particularidade da análise de sobrevivência e confiabilidade se refere ao fato de que existem situações onde um evento de interesse do pesquisador pode ocorrer várias vezes em uma unidade amostral. Neste trabalho estudamos o modelo de intensidade híbrido Poisson - Número de eventos para dados de eventos recorrentes.

PALAVRAS-CHAVE: Eventos recorrentes. Função de intensidade híbrida. Processo de Poisson. Teoria Assintótica. Estimação.

INTRODUÇÃO

Uma particularidade da análise de sobrevivência e confiabilidade se refere ao fato de que existem situações nas quais um evento de interesse pode ocorrer várias vezes para uma mesma unidade amostral. Nesta estrutura de dados, para cada unidade amostral é observado o número de eventos ocorridos em um determinado período de tempo, os tempos de ocorrência de cada um destes eventos e, além disso, pode-se ter disponível um vetor de covariáveis e um vetor de indicadores de censura. Diante disto, o interesse reside em entender e caracterizar o processo de recorrência de eventos nas unidades amostrais e possivelmente a comparação de tratamentos, com base no tempo para cada evento distinto, o número de eventos, os tipos de eventos e as interdependências entre eventos, visando explicar a natureza da variação entre unidades amostrais em termos de tratamento, covariáveis e de outros fatores que podem ser não observáveis. Alguns exemplos específicos nesta área podem ser encontrados em Cook & Lawless (2002).

MATERIAL E MÉTODOS

Processos envolvendo eventos recorrentes são comuns em confiabilidade e várias outras áreas, desta forma considere um sistema reparável, observando um tempo $t \geq 0$, e suponha que os eventos ocorram nos tempos $(t_1, t_2, ...)$, tal que $t_1 < t_2$. Dado que $x_i = t_i - t_{i-1}$(com $t_0=0$ e i=1,2,....), assim x_i denota o tempo entre os eventos, e $N(s,t)$ denota o número de eventos no intervalo [s,t), a notação $N(t)$ também pode ser usada para representar $N(0,t)$, ver Lawless e Thiagarajah (1996).

Geralmente, um modelo de probabilidade para um determinado ponto no processo pode ser especificado em termos de uma condicional ou uma função intensidade completa (FDC). Definimos $H_t = \{N(s): 0 \leq s < t\}$ como o histórico do processo acima para o tempo t. Então a FDC é dada por,

$$\lambda(t; H_t) = \lim_{\Delta t \downarrow 0} \frac{P\{N(t, t+\Delta t) = 1 | H_t\}}{\Delta t} \qquad (1)$$

* Professora da Faculdade de Tecnologia de Taquaritinga e doutoranda da Unesp - Jaboticabal em Estatística Aplicada a Genética e Melhoramento Animal

Particularmente podemos reescrever a equação em (1) na forma,

$$\lambda(t,H_t) = e^{\theta z(t)}, \qquad (2)$$

em que $z(t)=(z_1(t),.....,z_p(t))'$ é um vetor de funções que pode depender de t e H_t, e $\theta=(\theta_1,....,\theta_p)'$ é um vetor de parâmetros desconhecidos. Todos os vetores são considerados vetor coluna.
Inferência paramétrica para os parâmetros de modelos da forma (2) são diretos. Este trabalho desenvolve e investiga procedimentos de inferência, sendo que o modelo tratado tem como $z(t)=(1,t,N(t))'$ e $\theta=(\alpha,\beta,\gamma)'$, assim a FDC pode ser escrita na forma,

$$\lambda(t,H_t) = e^{\alpha+\beta t+\gamma N(t)}, \qquad (3)$$

No modelo (3) o parâmetro α denota o risco de base, o parâmetro β é caracterizado como o coeficiente do tempo, ou seja, ele será significante se o tempo interferir na função intensidade. O último parâmetro que é dado por γ, também só terá significância se a função de risco, no caso a intensidade por se tratar de eventos recorrentes, depender da quantidade de eventos.
A verossimilhança é proporcional para densidade de probabilidade para dados observados, na qual é da forma {n eventos, no tempo $t_1<......<t_n\leq T$}; em que $n\geq 0$. Desta forma a função de verossimilhança considerada, é dada por,

$$L(\theta) = \prod_{i=1}^{n} e^{\alpha+\beta t+\gamma N(t)} \cdot \exp\{-\int_0^T e^{\alpha+\beta t+\gamma N(t)}dt\}. \qquad (4)$$

Como o principal interesse é encontrar os estimadores de máxima verossimilhança, neste modelo não encontraremos problemas para estimá-los, uma vez que a integral envolvida é de fácil resolução, diferentemente do modelo proposto por Lawless e Thiagarajah (1996).

RESULTADOS E DISCUSSÕES

Para ilustrar o uso do modelo (4) e associá-lo com processo de inferência, nós consideramos o subconjunto de dados gerados pela inversa da função densidade condicional, sendo esta dada por:

$$f(t_i / t_{i-1}) = e^{\alpha+\beta t+\gamma (i-1) - \frac{e^{\alpha+(i-1)\gamma}(e^{\beta t_i}-e^{\beta t_{i-1}})}{\beta}}, \qquad (5)$$

Desta forma conseguimos estimar os parâmetros para o modelo e também verificar a consistência do mesmo através de testes de hipóteses.
Os dados artificiais gerados referem-se à quantidade de eventos recorrentes em apenas um componente ou indivíduo, sendo aqui considerado igual a 10, 25, 50, 100, 150 e 300. A Tabela 1 apresenta o número de eventos recorrentes, e os respectivos intervalos de confiança em relação aos três parâmetros estimados α, β e γ.

Tabela 1. Intervalo de Confiança para cada parâmetro

n	α	β	γ
10	(-0.902933, 2.885241)	(12.97509, 45.09368)	(-5.429661, -1.574914)
25	(0.6152125, 2.6195736)	(0.1207661, 5.8338198)	(-0.576146131, 0.006695533)
50	(1.938203, 3.046013)	(-0.1100208, 2.0716776)	(-0.289147514, 0.005421464)
100	(1.724362, 2.529397)	(0.04917632, 1.33738202)	(-0.172174444,-0.006328347)
150	(1.596579, 2.295810)	(0.1951839, 1.1897598)	(-0.14628934, -0.02354922)
300	(1.761850, 2.238151)	(0.3765116, 1.2234865)	(-0.15247742, -0.04757104)

O que podemos verificar na Tabela 1, é que quando o número de eventos recorrentes é muito baixo, ou seja, 10, 25 ou até mesmo 50, os intervalos de confiança contém o zero, assim podemos dizer que, quando somente pequenas quantidades de ocorrências são observadas, existe dificuldade em verificarmos o efeito do tempo e do número de eventos na função de intensidade.

Os testes de hipóteses também serão realizados para verificação da significância dos parâmetros, inicialmente consideramos o teste de Wald, que se baseia nas estatísticas $W^2=\alpha/s^2(\alpha), W^2=\beta/s^2(\beta)^2$ e $W^2=\gamma/s^2(\gamma)$, em que $s^2(\alpha)=(I(\theta)^{-1})_{3x3}$, $s^2(\beta)=(I(\theta)^{-1})_{3x3}$ e $s^2(\gamma)=(I(\theta)^{-1})_{3x3}$ para os parâmetros α, β e γ respectivamente, tal que a hipótese nula $H_0=\alpha=\beta=\gamma=0$, em que se H é verdadeira é aproximadamente uma distribuição $\chi_1^2=3.841$.

Tabela 2. Estatística de Wald

n	W_α	W_β	W_γ
10	1.051949	12.55696	12.68481
25	10.00579	4.173291	3.667103
50	77.76348	3.105766	3.5639991
100	107.2578	4.450594	4.450324
150	119.0430	7.449044	7.355491
300	270.9373	13.70913	13.96943

A Tabela 2 apresenta a Estatística de Wald para diferentes quantidades de eventos recorrentes. Podemos observar que todos os parâmetros do modelo se tornam significativos após 50 recorrências, confirmando o já havia sido obtido anteriormente através dos intervalos de confiança.

Foi realizado também a verificação da probabilidade de cobertura para os intervalos de confiança assintóticos com probabilidade de cobertura nominal de 95%. Os resultados são apresentados na Tabela 3.

Tabela 3. Probabilidade de cobertura para cada parâmetro

n	α	β	γ
10	94,6%	77,1%	74%
25	93,6%	76,5%	76,2%
50	94,0%	73,3%	74,4%
100	93,6%	83,8%	84,3%
150	96,7%	87,3%	87,3%

Através da Tabela 3 constatamos que a probabilidade de cobertura aumenta com a quantidade de eventos recorrentes.

CONSIDERAÇÕES FINAIS

Neste trabalho propomos uma nova função de intensidade para modelar eventos recorrentes que depende tanto do efeito do tempo quanto da quantidade de eventos recorrentes observados. Mais estudos de simulação são necessários para verificar a probabilidade de cobertura dos intervalos de confiança, bem como para determinar as distribuições empíricas das estatísticas de testes.

ABSTRACT

It is common we find situation where an individual can be committed by several events. In this work we study a new hybrid intensity model for fitting recurrent event data. The idea is to consider a model that account for the total time to event and for the number of recurrent event.

KEYWORDS: *Recurrent event. Hybrid intensity function. Poisson process. Maximum likelihood estimation.*

REFERÊNCIAS

COX, D. R. *Renewal Theory*, (1962).

LAWLESS, J.F. & THIAGARAJAH,K.(1996). *A point-process model incorporating renewals and time trends, with application to repairable systems*. Technometrics 38,131-38.

LOUZADA-NETO, F.; MAZUCHELLI, J.; ACHCAR, J.A. (2002). *Introdução à Análise de Sobrevivência e Confiabilidade*. III Jornada Regional de Estatística.

NELSON, W. (1982). *Applied Life Data Analysis*. John Wiley and Sons, New York, NY.

NELSON, W. (1988). *Graphical Analysis of System Repair Data.* Journal of Quality Technology, 20,24-35.

NELSON, W. (1995). Confidence *Limits for Recurrence Data - Applied to Cost or Number of Product Repair.* Technometrics, 37, p.147-157.

PRENTICE, R.L., WILLIAMS, B. J. AND PETERSON, A. V. (1981). *On the regression analysis of multivariate failure time data.* Biometrika, 68, 373-79.

ROSS, S. M. (1983), *Stochastic Processes,* New York: John Wiley.