
O USO DO MODELO POLI-LOG-LOGÍSTICO EM ANÁLISE DE SOBREVIVÊNCIA E CONFIABILIDADE

CARRASCO, Cleber Giuglioli

RESUMO: Em análise de sobrevivência e confiabilidade, além de estudos que englobam apenas uma causa de falha, são comuns pesquisas em áreas médicas e industriais centradas em causas que competem entre si para causarem a ocorrência do evento de interesse, como o óbito do paciente ou a falha de um determinado componente. Entretanto, em muitos experimentos a causa de morte do paciente ou falha do componente pode ser desconhecida, dando origem aos chamados riscos competitivos latentes que podem ser modelados via modelo poli-log-logístico, em particular, se a função de risco for multimodal. Uma técnica de verificação gráfica para ajuste do modelo é apresentada.

Palavras-chave: Análise de sobrevivência e confiabilidade. Função de risco. Função de sobrevivência. Modelo poli-log-logístico. Riscos competitivos.

SUMMARY: In survival analysis and reliability studies it is common we find data sets where several failure causes competing between themselves to cause the occurrence of the event of interest, such as the patient death or the failure of a certain component. However, in many experiments the cause of the patient death or the failure of the component can be unknown, leading to the latent competing risk problems. This kind of data can be modeled through the poly-log-logistic model, in particular, if the risk function is multimodal.

Keywords: Competing risk. Hazard function. Poly-log-logistic model. Reliability and survival analysis.

INTRODUÇÃO

Em análise de sobrevivência, um paciente cujo sistema imunológico é deficiente está sujeito a contrair inúmeros tipos de doenças, tendo assim várias possíveis causas de falha competindo entre si, para ocasionar a morte do paciente. Em confiabilidade, cada componente de um sistema elétrico ligado em série pode ocasionar a falha do sistema, tendo assim várias possíveis causas de falhas competindo entre si. Neste tipo de problema, o tempo de vida associado a um particular fator de risco não é observado, mas somente o tempo mínimo dos tempos de vida relacionados a todos os fatores de riscos e, em geral, as causas de falhas são latentes, no sentido de não termos informação sobre qual fator de risco foi responsável pela morte do indivíduo ou falha do componente. Em muitas situações, esta informação não está disponível ou é impossível para um especialista especificar a real causa da morte ou falha. Quando a causa

morte ou falha não é conhecida, temos um problema de riscos competitivos latentes, que estão presentes em diversas áreas, como por exemplo, ciências atuais, confiabilidade industrial, demografia, estudos biomédicos e saúde pública.

Há uma extensa literatura com procedimentos estatísticos de análise baseados em modelos de riscos para tratarmos esses problemas, quando as causas de falhas são conhecidas (COX; OAKES, 1984). Porém, tais procedimentos não podem ser utilizados se as causas de mortes ou falhas são latentes, ou seja, se não sabemos qual o fator de risco foi responsável pela morte do indivíduo ou falha do componente.

Uma possível formulação para esse problema é assumir que os riscos afetam os indivíduos independentemente. Esta suposição está de acordo com a discussão apresentada por Tsistis (1975) que enfatiza que a hipótese de independência não é diretamente testável, e Cox (1962) e Tsistis (1975) que argumentam existir uma distribuição conjunta com tempos de falhas independentes para qualquer distribuição conjunta de todos os tempos de vida relacionados a cada fator de risco. Dessa forma, se assumirmos que o tempo de vida relacionado a um particular risco tem uma distribuição log-logística, então o mínimo tempo de vida entre todos os tempos de vida terá uma distribuição poli-log-logística. Essa distribuição é muito útil para modelarmos dados com funções de riscos multimodais, o que será discutido neste trabalho.

Em análise de sobrevivência e confiabilidade é comum termos a presença de censuras nos dados. Vários fatos podem colaborar para a ocorrência de censuras. Entre outros podemos citar o evento de interesse pode não ocorrer até o final do tempo de estudo, devido a diversas peculiaridades como, por exemplo, o abandono de pacientes antes do término da pesquisa ou a falha de um determinado componente devido a outras causas não consideradas no estudo. Neste estudo não discutiremos este assunto, mas em um trabalho recente Carrasco e Louzada-Neto (2001) trabalharam com censura do tipo aleatória, o que é muito comum na área médica. Isto ocorre quando um paciente é retirado no decorrer do experimento sem ter ocorrido o evento de interesse, ou, se o paciente falecer por uma razão diferente da estudada.

Neste trabalho faz-se uma pequena introdução à análise de sobrevivência e confiabilidade, apresentando a função de sobrevivência e a função de risco. Descrevemos o modelo poli-log-logístico e suas propriedades, apontando as motivações para o seu uso no contexto de dados de riscos competitivos latentes e, em particular, para dados que apresentam função de risco multimodal. Uma técnica de verificação gráfica para ajuste do modelo também é apresentada.

1 ESPECIFICANDO O TEMPO DE FALHA

Seja T uma variável aleatória não-negativa a qual representa o tempo de falha, tal variável é usualmente especificada pela sua função de sobrevivência ou função de falha (ou risco). A seguir apresentamos estas duas funções.

1.1 A FUNÇÃO DE SOBREVIVÊNCIA

A função de sobrevivência $S(t)$ é definida como a probabilidade de uma observação não falhar até um certo tempo t . Desta forma, podemos escrever $S(t)$ como: $S(t) = P(T > t)$.

A tabela 1 apresenta algumas funções de sobrevivência que podem ser expressas analiticamente e que são normalmente utilizadas em análise de sobrevivência e confiabilidade.

Tabela 1: Algumas funções de sobrevivência paramétricas.

Distribuição	Função de Sobrevivência
Exponencial	$e^{-\frac{t}{\mu}}$
Weibull	$e^{-\left(\frac{t}{\mu}\right)^\beta}$
Log-Logística	$\frac{\mu^\beta}{\mu^\beta + t^\beta}$

1.2 A FUNÇÃO DE RISCO

A função de risco, representada por $h(t)$, especifica a taxa de falha instantânea no tempo t condicionada à sobrevivência até o tempo t e é dada por (COX ; OAKES, 1984).

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq t + \Delta t / T \geq t)}{\Delta t}.$$

Observe que os valores da função de risco (2.2) são números positivos, não limitados superiormente.

Existe uma relação matemática entre a função de risco $h(t)$ e a função de sobrevivência $S(t)$, dada através da seguinte expressão.

$$h(t) = \frac{f(t)}{S(t)},$$

Onde $f(t)$ é conhecida como função de densidade de probabilidade de T .

Outra relação muito utilizada em análise de sobrevivência, relaciona a função de sobrevivência com a função de risco acumulada $H(t)$, dada por

$$H(t) = \int_0^t h(u) du = -\log(S(t)).$$

Devido a sua interpretação, a função de risco tem sido preferida por muitos autores, uma vez que ela descreve como a probabilidade instantânea de falha se modifica com o passar do tempo. Além disso, através da função de risco podemos obter classes especiais de distribuições de sobrevivência. Por exemplo, a distribuição exponencial acomoda funções de riscos constantes, isto é, $h(t) = 1/\mu$, enquanto que se a função de risco for monotonicamente decrescente ou crescente em t , temos uma distribuição de Weibull (LAWLESS, 1982). As distribuições log-logística e log-normal acomodam funções de risco unimodais (KALBFLEISH; PRENTICE, 1980).

A tabela 2 apresenta algumas funções de risco que podem ser expressas analiticamente e que são comumente utilizadas em análise de sobrevivência paramétrica.

Tabela 2: Algumas funções de risco paramétricas

Distribuição	Função de Risco
Exponencial	$\frac{1}{\mu}$
Weibull	$\frac{\beta t^{\beta-1}}{\mu^\beta}$

Log-Logística	$\frac{\beta t^{\beta-1}}{\mu^\beta + t^\beta}$
---------------	---

Apesar de raras, funções de risco multimodais e em forma de “U” também podem ser observadas. Louzada-Neto (1999) propôs uma função de riscos múltiplos, também conhecida como modelo de riscos múltiplos, que acomoda esses tipos de comportamento, definida por

$$h(t) = \sum_{j=1}^m \frac{\beta_j t^{\beta_j-1}}{\mu_j^{\beta_j}} q_j(t; \mu_j, \beta_j, k),$$

onde μ_j , β_j e k são parâmetros positivos desconhecidos e $q_j(\cdot)$ é uma função monótona de forma paramétrica conhecida que é igual a um, quando todos os seus argumentos são iguais a zero.

Uma vantagem da função de múltiplos riscos (2.5) é sua capacidade de acomodar várias famílias de modelos usuais, como por exemplo, os modelos de riscos gamma generalizado, log-normal, log-logístico e o Weibull. Tais funções permitem não somente o ajuste de funções de riscos constantes, monótonas crescente e decrescente, mas também o ajuste de funções de riscos em forma de “U” e multimodais.

2 O MODELO LOG-LOGÍSTICO SIMPLES

Uma variável aleatória T tem distribuição log-logística simples se a função de risco e a função de sobrevivência para T , são dadas respectivamente por (LAWLESS, 1982).

$$h(t) = \frac{\beta t^{\beta-1}}{\mu^\beta + t^\beta} \quad (3.1)$$

$$S(t) = \frac{\mu^\beta}{\mu^\beta + t^\beta}$$

onde μ e β são respectivamente os parâmetros de escala e forma respectivamente.

A função de densidade de probabilidade para T é dada a partir de (2.3) e (3.1) por

$$f(t) = \left(\frac{\beta t^{\beta-1}}{\mu^\beta + t^\beta} \right) \left(\frac{\mu^\beta}{\mu^\beta + t^\beta} \right) \quad (3.2)$$

Assim a média e a variância de T são respectivamente dadas por,

$$E(t) = \pi \csc(\pi / \beta) / (\beta(1 / \mu^\beta)^{1/\beta}) \text{ se } \beta > 1 \text{ e } \text{Var}(t) = 2 \pi \csc(2 \pi / \beta) / (\beta(1 / \mu^\beta)^{2/\beta}) - E(t)^2, \\ \text{se } \beta > 2. \text{ O } j\text{-ésimo percentil } t_j \text{ é definido por } t_j = (j / ((1 / \mu^\beta)(1 - j)))^{1/2}.$$

Pode-se mostrar que a distribuição log-logística simples tem função de risco cuja forma depende do parâmetro de forma, isto é, para $\beta \leq 1$ a função de risco é monotonicamente decrescente e para $\beta > 1$, a função de risco tem forma côncava, ou seja, inicialmente cresce para o máximo no tempo $((\beta - 1) / (1 / \mu^\beta))^{1/\beta}$ e então decresce para zero quando o tempo se aproxima do infinito. O modelo log normal também pode acomodar funções de risco na forma côncava, sendo similar à distribuição log-logística (Klein; Moeschberger, 1997). A vantagem do modelo log-logístico em relação à log-normal é apresentar uma expressão simples para a função de risco e para a função de sobrevivência.

3 O MODELO POLI-LOG-LOGÍSTICO

Considere que indivíduos ou componentes são submetidos a m causas de mortes ou falhas diferentes $m \leq 2$. Seja X_j o tempo de vida relacionado com a j-ésima causa de morte ou falha, com uma função de densidade de probabilidade dada por $f_j(x)$. Somente o tempo mínimo entre os vários riscos, $T = \min(X_1, \dots, X_m)$, é observado para cada indivíduo ou componente e, os X_j 's são supostos independentes. Assim, a função de risco e a função de sobrevivência para T são dadas respectivamente, por

$$h(t) = \sum_{j=1}^m h_j(t) \quad (4.1)$$

$$S(t) = \prod_{j=1}^m S_j(t)$$

De (2.3) e (4.1), temos que a função de densidade de probabilidade é dada por

$$f(t) = \sum_{j=1}^m h_j(t) \prod_{j=1}^m S_j(t). \quad (4.2)$$

Quando assumimos que o tempo de vida relacionado a um particular risco tem uma distribuição log-logística, então de (3.1) e (3.2) o mínimo tempo de vida entre vários riscos terá uma distribuição poli-log-logística (Louzada-Neto, 1999) dada por

$$h(t) = \sum_{j=1}^m \frac{\beta_j t^{\beta_j - 1}}{\mu_j^{\beta_j}} \quad (4.3)$$

onde μ_j , β_j e k são parâmetros positivos a serem estimados, com μ_j denotando o parâmetro de escala, representando o 63.2 percentil da distribuição e β_j representa o parâmetro de forma (Nelson, 1990).

De (3.1) e (3.2) a função de sobrevivência do mínimo tempo de vida T é definida por

$$S(t) = \prod_{j=1}^m \frac{\mu_j^{\beta_j}}{\mu_j^{\beta_j} + t^{\beta_j}} \quad (4.4)$$

Substituindo (4.3) e (4.4) em (4.2), obtemos a função de densidade de probabilidade definida por

$$f(t) = \sum_{j=1}^m \left(\frac{\beta_j t^{\beta_j - 1}}{\mu_j^{\beta_j} + t^{\beta_j}} \right) \prod_{j=1}^m \left(\frac{\mu_j^{\beta_j}}{\mu_j^{\beta_j} + t^{\beta_j}} \right). \quad (4.5)$$

Quando apenas dois riscos são considerados isto é, $m = 2$, o modelo poli-log-logístico (4.3) é dado por

$$h(t) = \frac{\beta_1 t^{\beta_1 - 1}}{\mu_1^{\beta_1} + t^{\beta_1}} + \frac{\beta_2 t^{\beta_2 - 1}}{\mu_2^{\beta_2} + t^{\beta_2}} \quad (4.6)$$

Assim, com $m = 2$, a esperança da variável T pode ser calculada como aleatória

$$E(T) = \int_0^{\infty} \left(\frac{\beta_1 t^{\beta_1} \mu_1^{\beta_1} \mu_2^{\beta_2}}{(\mu_1^{\beta_1} + t^{\beta_1})^2 (\mu_2^{\beta_2} + t^{\beta_2})} + \frac{\beta_2 t^{\beta_2} \mu_2^{\beta_2} \mu_1^{\beta_1}}{(\mu_2^{\beta_2} + t^{\beta_2})^2 (\mu_1^{\beta_1} + t^{\beta_1})} \right) dt. \quad (4.7)$$

Enquanto que a variância de T pode ser obtida através da forma usual

$$\text{Var}(T) = E(T^2) - (E(T))^2. \quad (4.8)$$

As integrais em 4.7 e 4.8 não têm soluções explícitas conhecidas e, portanto, métodos numéricos devem ser aplicados para a obtenção de suas soluções.

4 MOTIVAÇÃO PARA O USO DO MODELO

Além de modelar dados de riscos competitivos latentes, o modelo poli-log-logístico também pode ser utilizado para modelar dados com função de risco multimodais, sendo esta outra motivação para sua formulação.

Considere o caso onde os dados apresentam uma função de risco bi-modal. O modelo (4.6), é suficientemente flexível para ajustar este tipo de conjuntos de dados pois, além de acomodar funções de risco unimodais, também pode acomodar funções de risco bi-modais, conforme é mostrado na Figura 1 a qual apresenta formas típicas da função de risco do modelo log-logístico duplo. Em sua generalização o modelo (4.3) pode acomodar funções de riscos multimodais.

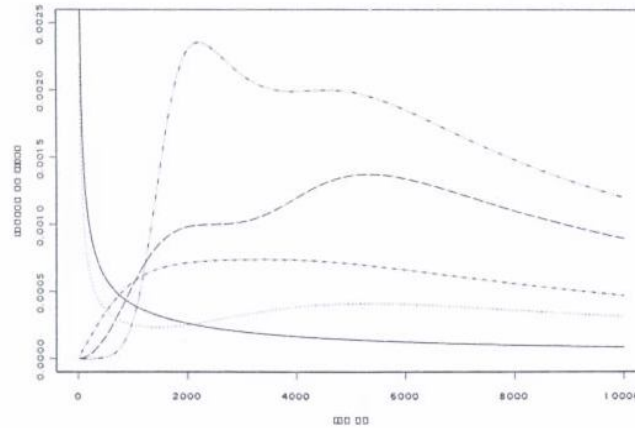


Figura 1: funções de risco para o modelo (2.8). Para todas as curvas $\mu_1 = 1650$. (---): $\beta_1 = 0,75$, (...): $\beta_1 = 0,5$ e $\beta_2 = 3$, (-.-.-.-): $B_1 = 2$ e $\beta_2 = 3$, (- - - -): $\beta_1 = 3$ e $\beta_2 = 6$ e (-...-...-): $\beta_1 = 6$ e $\beta_2 = 6$.

A forma dessas funções de risco depende dos valores conjuntos dos parâmetros de forma β_1 e β_2 . Se o $\max(\beta_1, \beta_2) < 1$, o risco é decrescente; se $\beta_1 < 1$ e $\beta_2 > 1$ a função de risco decresce-decresce e se o $\min(\beta_1, \beta_2) > 1$ o risco é uni ou bi-modal.

Apesar de haver modelos de riscos que acomodam funções de risco unimodais, como é o caso da distribuição log-logística simples (KALBFLEISCH; PRENTICE, 1980) e da distribuição Weibull-exponenciada (MUDHOLKAR et al., 1995), a literatura nessa área é carente de modelos que acomodam funções de riscos multimodais. Alguns exemplos numéricos com função de risco multimodais relacionadas a pacientes transplantados podem ser encontrados em Klein; Moeschberger (1977). Neste contexto, o risco poli-log-logístico (2.5) é suficientemente flexível para acomodar esse tipo de conjunto de dados, sendo esta, uma outra motivação para sua formulação.

5 REVISÃO

A escolha do modelo a ser utilizado é um tópico extremamente importante na análise paramétrica de dados de tempo de vida. O procedimento de estimação só deverá ser aplicado após ter sido estabelecido o modelo probalístico adequado aos dados. Assim, uma implementação

de técnicas gráficas de adequação para ajuste do modelo se faz necessária. Essas técnicas gráficas são a forma mais simples de escolher o modelo a ser utilizado para um determinado conjunto de dados, além de serem utilizadas com bastante frequência para verificar a adequação do ajuste de modelos estatísticos. Tais técnicas geralmente são baseadas na construção de gráficos da forma.

$$g_1(H(t)) \text{ versus } g_1(t)$$

onde $H(t) = \int_0^t h(a)da$ é a função de risco acumulada (LAWLESS, 1982) e g_1 é uma função monotonicamente crescente tal que $\lim_{g_1(H(\cdot))} = -\infty$, $\lim_{g_1(H(\cdot))} = \infty$.

A idéia básica deste procedimento é obter gráficos que devem ser aproximadamente lineares caso o modelo selecionado seja adequado. Violações da linearidade podem ser verificadas visualmente.

6 ILUSTRAÇÕES NUMÉRICAS

Considerando $g_1(\cdot) = \log(\cdot)$, quando um gráfico de $g_1(H(t)) = \log(e^{H(t)-1})$ versus $g_1(t) = \log(t)$ for aproximadamente linear, o modelo log-logístico simples é adequado aos dados (LOUZADA-NETO, 1997).

A figura 2 mostra gráficos para dois exemplos gerados com tamanhos amostrais iguais a 80. Para o primeiro exemplo os parâmetros foram fixados em $\mu_1 = 1000$, $\beta_1 = 3$, $\mu_2 = 2000$, $\beta_2 = 3$. O primeiro exemplo representa uma situação onde o modelo log-logístico simples não é adequado, mas um modelo log-logístico duplo possivelmente seja, enquanto o segundo exemplo representa uma situação onde uma distribuição log-logística simples será suficiente para o ajuste.

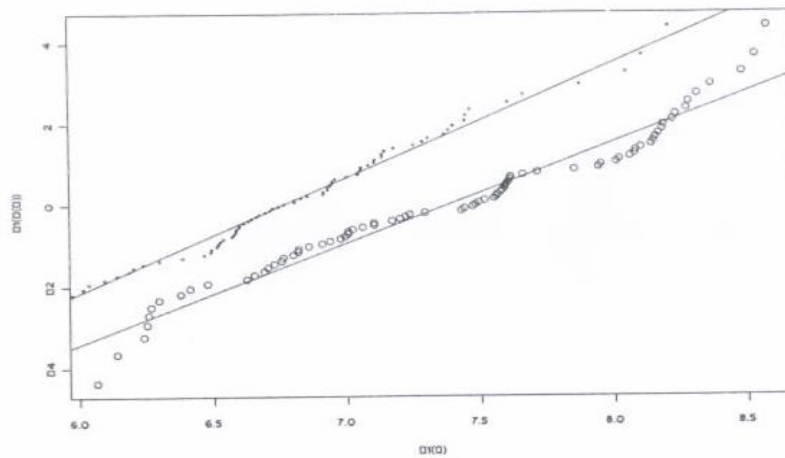


Figura 2: Verificação gráfica para ajuste do modelo: (o o o) exemplo 1 e (...) exemplo 2.

CONCLUSÃO

O modelo poli-log-logístico (4.3) pode ser utilizado para modelagem de riscos competitivos latentes, ou seja, quando não temos informação de qual fator de risco foi responsável pela morte do indivíduo ou falha do componente. Também pode ser utilizado com sucesso par ajuste de dados com funções de riscos multimodais, sendo essa sua grande vantagem em relação aos outros modelos que ajustam dados de riscos competitivos. Entretanto, podemos também utilizar a técnica de verificação gráfica para auxiliar na escolha do modelo a ser ajustado.

REFERÊNCIAS

CARRASCO, C. G; LOUZADA-NETO, F. **Intervalos de confiança para os parâmetros do Modelo log-logístico duplo na presença de censuras aleatórias**. São Carlos: UFScar, 2001. (Relatório Técnico do Departamento de Estatística, v.66).

COX, D. R. **Oakes, D. analysis survival data**. London: Chapman and Hall, 1984. 201p.

COX, D. R. **Renewal Theory**. London: Methuen. 1962.

KALBFLEISCH, J. D; PRENTICE, R. L. **The statistical analysis of failure time data.** New York: Wiley, 1980.

KLEIN, J. P; MOESCHBERGER, L. M. **Survival analysis:** techniques for censored and truncated data. New York: Springer, 1997, 501p.

LAWLESS, J. F. **Statistical models and methods for lifetime data.** New York: Wiley, 1982. 580p.

LOUZADA-NETO, F. Extended hazard regression model for reliability and survival analysis. **Lifetime Data Analysis**, v.3, p.367-81, 1997.

_____ Poly-hazard regression models for lifetime data. **Biometrics**, v.55, p.1121-5, 1999

MUDHOLKAR, G. S; SRIVASTAVA, D. K; FREIMER, M. The exponentiated Weibull family: a reanalysis of the bus-motor-failure-data. **Technometrics**, v.37, p.436-45, 1995.

NELSON, W. **Accelerated Testing:** Statistical Models, Test Plans and Data Analysis. New York: Wiley, 1990. 601p.

TSIATIS, A. A. A nonidentifiability aspect of the problem of competing risks. **Proc. Nat. Acad. Sci. U.S.A.**, v.72, p.20-22, 1975.