Um novo modelo de cura paramétrico

Ana M. Abreu
DME e CCM, Universidade da Madeira
Cristina S. Rocha
DEIO e CEAUL, Faculdade de Ciências da Universidade de Lisboa

Resumo: Os modelos de cura, nomeadamente os de mistura, foram desenvolvidos para dar resposta a situações em que há suspeita da existência de indivíduos imunes (ou não susceptíveis) numa população. Estes modelos podem incluir, ou não, covariáveis, que representam factores de risco associados a cada indivíduo. Neste trabalho, iremos apresentar um novo modelo de cura paramétrico, no sentido em que é baseado numa distribuição ainda pouco conhecida, proposta por Chen (2000). Este modelo revelou-se um forte concorrente a modelos já existentes, nomeadamente no que diz respeito à flexibilidade da função hazard, facilidade de tratamento matemático e qualidade de ajustamento a dados reais já analisados anteriormente.

Palavras–chave: Algoritmo EM, distribuição de Chen, modelo de cura.

Abstract: Cure models, namely cure mixture models, were developed to contemplate situations where one suspects the presence of cured individuals in the population. These models may include covariables which represent risk factors associated to each individual. In this work, we present a new parametric cure model, in the sense that it is based on a still fairly unknown distribution, proposed by Chen (2000). This model has shown to be a strong contender to existent models, namely regarding hazard function flexibility, easiness for mathematical modelling and quality of fit to real data previously analised.

Keywords: EM algorithm, Chen distribution, cure model.

1 Introdução

A maior área de aplicação da Análise de Sobrevivência é sem dúvida a da Medicina, a qual tem sofrido uma grande evolução, nomeadamente no que diz respeito à possibilidade de cura de certas doenças. Não é pois de estranhar que tenham sido desenvolvidos novos modelos, que permitem acompanhar esta evolução. Neste contexto, surgem os modelos de cura, os quais contemplam situações onde se suspeita da existência de indivíduos curados na população. Uma propriedade importante dos modelos de cura é o facto de a correspondente função de sobrevivência ser imprópria, que é o mesmo que dizer que a função hazard cumulativa é limitada. Em termos gerais, a primeira característica (função de sobrevivência imprópria) está associada aos modelos de cura de mistura, e a segunda (função

hazard cumulativa limitada) aos modelos de cura de não-mistura.

Frequentemente a cura é um acontecimento que não é observável, pelo que a suspeita é sustentada por algumas características associadas, em particular, pela existência de muitas observações censuradas para além do último tempo de vida observado. Assim, o nivelamento da estimativa de Kaplan-Meier da função de sobrevivência é um forte indicador da possível presença de indivíduos curados na população em estudo.

Note-se que na Análise de Sobrevivência clássica também existem observações censuradas. No entanto, existe uma diferença fundamental: admite-se que o acontecimento de interesse irá ocorrer para todos os indivíduos, eventualmente fora do período de observação.

Neste trabalho, vamos considerar a distribuição de Chen (2000) e, a partir dela, construir dois novos modelos de cura, um de mistura e um de não-mistura. A secção 2 trata destes assuntos. Na secção 3, é abordada a questão da estimação dos parâmetros do modelo, a que se seguem, na secção 4, algumas aplicações. A secção 5 é destinada aos comentários finais.

2 A distribuição de Chen e os modelos de cura

Nesta secção iremos mostrar como a distribuição de Chen pode ser usada num modelo de cura. Os modelos de cura podem ser caracterizados como modelos de cura de mistura e como modelos de cura de não-mistura embora, com a parametrização adequada, dêem origem ao mesmo modelo. Comecemos por definir sucintamente os dois tipos de modelos.

Seja T uma variável aleatória que representa o tempo de vida de um indivíduo numa dada população, na qual se admite existirem indivíduos doentes (ou susceptíveis) e indivíduos curados (ou não susceptíveis ou imunes). Seja ainda Y uma variável binária que toma o valor 1 se o indivíduo é susceptível e zero se está curado. O modelo de cura de mistura, escrito com base na função de sobrevivência, é da forma

$$S(t) = p + (1-p)S(t|Y=1), \tag{1}$$

onde p representa a taxa de cura e $S(t|Y=1)$ é a função de sobrevivência (própria) dos indivíduos susceptíveis. Para um indivíduo cujo tempo de vida é censurado, Y não é observável, logo não sabemos se se trata de um indivíduo susceptível ou curado. Uma vez que $S(t) \to p$ quando $t \to \infty$, $S(t)$ é uma função de sobrevivência imprópria.

Defina-se agora $H(t)$ tal que $S(t) = \exp[-H(t)]$. Se $S(\infty) > 0$, existe $\theta < \infty$, tal que $H(\infty) = \theta$. De acordo com Yakovlev *et al.* (1993), uma forma possível de caracterizar esta propriedade consiste em considerar $H(t) = \theta F(t)$, onde $F(t)$ designa a função de distribuição (própria) de uma variável aleatória não negativa. Então, o modelo de cura de não-mistura pode ser escrito na forma

$$S(t) = \exp[-\theta F(t)]. \tag{2}$$

A correspondente função hazard é

$$h(t) = \theta f(t),$$

onde $f(t)$ é a função densidade correspondente a $F(t)$. Neste caso, a probabilidade de cura é $e^{-\theta}$.

2.1 A distribuição de Chen

A função de distribuição proposta por Chen (2000) é

$$F(t) = 1 - \exp[\lambda(1 - \exp(t^{\beta}))], \quad t > 0, \quad \lambda,\ \beta > 0, \tag{3}$$

onde λ é o parâmetro de escala e β o parâmetro de forma. As correspondentes função de sobrevivência e função hazard são, respectivamente,

$$\overline{F}(t) = \exp[\lambda(1 - \exp(t^{\beta}))], \quad t > 0, \tag{4}$$

$$h^*(t) = \lambda\beta t^{\beta-1} \exp(t^{\beta}), \quad t > 0.$$

O autor refere que $h^*(t)$ pode ser *bathtub-shaped* quando $\beta < 1$ e que é crescente quando $\beta \geq 1$, o que não é habitual nas distribuições mais usadas em análise de sobrevivência. Na verdade, como

$$h^{*'}(t) = \lambda\beta t^{\beta-2} \exp(t^{\beta})((\beta - 1) + \beta t^{\beta}),$$

para $\beta < 1$ temos que $h^*(t)$ é decrescente para $t \in]0, (\frac{1}{\beta} - 1)^{\frac{1}{\beta}}]$ e, para $t \geq (\frac{1}{\beta} - 1)^{\frac{1}{\beta}}$, $h^*(t)$ é crescente. Assim, o intervalo para t onde $h^*(t)$ é decrescente será tanto maior quanto menor o valor de β. Deste modo, para valores de β próximos de zero, por exemplo, $\beta = 0.1$, o intervalo tem uma amplitude tão grande que, do ponto de vista prático, é como se tivessemos uma função hazard decrescente. Já para valores de β próximos de 1, o intervalo em que a função hazard é decrescente é tão pequeno que é quase como se a função hazard fosse sempre crescente.

2.2 Os novos modelos de cura: de mistura e de não-mistura

Modelo de mistura

Admitamos que o tempo de vida dos indivíduos susceptíveis segue a distribuição de Chen, dada por (3). Então, o novo modelo de cura, que se obtém ao substituir em (1) $S(t|Y=1)$ pela expressão (4), escreve-se na forma

$$S(t) = p + (1-p)\exp[\lambda(1-\exp(t^{\beta}))], \quad t > 0,\ \lambda, \beta > 0.$$

Se o modelo for definido à custa da função hazard, temos

$$h(t) = \frac{(1-p)\lambda\beta t^{\beta-1}\exp(t^{\beta})\exp[\lambda(1-\exp(t^{\beta}))]}{p + (1-p)\exp[\lambda(1-\exp(t^{\beta}))]}.$$

Importa notar que a função hazard populacional é decrescente para $\beta \leq 1$ e é unimodal para $\beta > 1$, o que a torna perfeitamente adequada à situação de existência simultânea de indivíduos doentes e de indivíduos curados numa população. No caso em que $\beta > 1$, o valor da moda será tanto maior quanto menor for o valor de p, como seria de esperar uma vez que p representa a proporção de indivíduos curados. Neste caso p é o parâmetro de escala. Relativamente ao parâmetro λ, de certa forma vai definir o valor de t a partir do qual a função hazard irá estabilizar num valor próximo de zero, ou seja, neste contexto, o valor de t a partir do qual poderemos esperar que os indivíduos que sobrevivam até esse instante estejam curados. Mais especificamente, quanto menor for o valor de λ, maior será o valor de t a partir do qual a função hazard é aproximadamente zero.

Modelo de não-mistura

Consideremos agora um modelo de cura de não-mistura (Yakovlev *et al.* (1993), Tsodikov (1998)), tal como em (2). O modelo que propomos, definido por

$$S(t) = \exp[\lambda(1-\exp(t^{\beta}))], \quad t > 0,\ \lambda, \beta < 0, \tag{5}$$

é um modelo de cura de não-mistura, onde a probabilidade de cura é dada por e^{λ} e $F(t) = 1-\exp(t^{\beta})$. Este modelo resulta de alterar o espaço dos parâmetros, de forma a que a função de sobrevivência da população seja imprópria. Assim, em vez de os parâmetros serem estritamente positivos, são estritamente negativos. Designámos a distribuição (5) por distribuição de Chen modificada.

3 Estimação dos parâmetros

O processo de estimação dos parâmetros do novo modelo de cura de mistura será abordado nesta secção. Para tal, vamos aplicar o método da máxima

verosimilhança. No entanto, como estamos a lidar com dados omissos, usamos o algoritmo EM (Dempster *et al.*, 1977), seguindo no essencial a abordagem de Steele (2003), embora a autora se referisse a tempos de vida discretos.

3.1 Construção da função de verosimilhança

Admitamos que a censura é não informativa. Consideremos uma amostra de dimensão n e designemos por $t_1, ..., t_n$ os tempos de vida dos n indivíduos. Suponhamos que os dados estão na forma (t_i, δ_i), $i = 1, ..., n$, onde $\delta_i = 1$ se t_i é observado e $\delta_i = 0$ caso contrário. Sem perda de generalidade, suponhamos que os primeiro m $(m < n)$ tempos de vida são censurados. Então, temos também

$$\delta_i = \begin{cases} 0 & \text{se } 1 \leq i \leq m \\ 1 & \text{se } m+1 \leq i \leq n \end{cases}.$$

A contribuição de um indivíduo para a função de verosimilhança, quando o acontecimento de interesse foi observado em t_i, é

$$(1-p)f(t_i|Y=1),$$

onde $f(t_i|Y=1)$ representa a função densidade de T condicional ao indivíduo ser susceptível, ou seja, a função densidade dos indivíduos susceptíveis. Se o acontecimento de interesse não é observado até t_i, então a contribuição do indivíduo para a função de verosimilhança é

$$p + (1-p)S(t_i|Y=1).$$

Assim, a função de verosimilhança observada é

$$L_O = \prod_{i=1}^{n} [(1-p)f(t_i|Y=1)]^{\delta_i} [p + (1-p)S(t_i|Y=1)]^{1-\delta_i},$$

a qual, ao usar a distribuição de Chen para a distribuição do tempo de vida dos indivíduos susceptíveis, se escreve na forma

$$L_O = \prod_{i=1}^{n} \left[(1-p)\lambda\beta t^{\beta-1} \exp\left[t^{\beta} + \lambda(1-e^{t^{\beta}})\right]\right]^{\delta_i} \left[p + (1-p)\exp\left[\lambda(1-e^{t^{\beta}})\right]\right]^{1-\delta_i}.$$

Sejam $y_1, ..., y_n$ tais que

$$y_i = \begin{cases} 0 & \text{se o indivíduo é imune} \\ 1 & \text{se o indivíduo é susceptível} \end{cases}.$$

Se todos os y_i's fossem observados, a verosimilhança completa seria

$$L_C = \prod_{i=1}^{n} [[(1-p)f(t_i|Y=1)]^{y_i}]^{\delta_i} [p^{1-y_i}[(1-p)S(t_i|Y=1)]^{y_i}]^{1-\delta_i}.$$

Considerando $q = 1-p$, após alguns cálculos, a expressão anterior pode ser escrita no modo que se segue

$$L_C = \prod_{i=1}^{n} q^{y_i}(1-q)^{1-y_i} \prod_{i=1}^{n} h(t_i|Y=1)^{y_i\delta_i} S(t_i|Y=1)^{y_i} = L_{C_1} L_{C_2}. \quad (6)$$

O logaritmo de (6) é dado por

$$\log L_C = \sum_{i=1}^{n} [y_i \log q + (1-y_i)\log(1-q) + y_i\delta_i \log h(t_i|Y=1) + y_i \log S(t_i|Y=1)]. \quad (7)$$

3.2 Algoritmo EM

Quando existem observações omissas, ou seja, dados incompletos, o algoritmo EM é uma ferramenta muito utilizada para maximizar a função de verosimilhança. Em termos gerais, em vez de maximizar a função de verosimilhança, maximizamos o seu valor esperado condicional aos valores actuais dos parâmetros e aos dados observados. Assim, os valores omissos são identificados com o respectivo valor esperado condicional.

Mais especificamente, a etapa E do algoritmo EM consiste em determinar o valor esperado do logaritmo da verosimilhança completa em relação à distribuição dos Y's não observados, condicional aos valores actuais dos parâmetros e aos dados observados $\mathcal{O}$, onde $\mathcal{O} = \{y_i \text{ observado}, (t_i, \delta_i), i = 1, ..., n\}$. No entanto, como $\log L_C$ é linear em Y, para calcular o valor esperado de $\log L_C$ basta substituir em (7), os valores não observados de Y pelos respectivos valores esperados, denotados por τ_i. Então, temos:

$$\tau_i = E(Y|\mathcal{O}) = P(Y=1|T_i > t_i, \delta_i = 0) = \frac{(1-p)S(t_i|Y=1)}{S(t_i)}. \quad (8)$$

Deste modo, no logaritmo da verosimilhança completa, cada y_i é substituído por ω_i, onde ω_i é definido na forma que se segue

$$\omega_i = \begin{cases} 1 & \text{se } \delta_i = 1 \\ \tau_i & \text{se } \delta_i = 0 \end{cases}$$

e designa a probabilidade de o i-ésimo indivíduo ser susceptível.

Para efectuar a etapa M, é necessário maximizar as duas componentes do valor esperado do logaritmo da função de verosimilhança

$$\begin{aligned}\log L_{E_1} &= \textstyle\sum_{i=1}^{n}[\omega_i \log q + (1-\omega_i)\log(1-q)] \\ &= (n-m)\log q + m\log(1-q) + \textstyle\sum_{i=1}^{m}\tau_i[\log q - \log(1-q)],\end{aligned}$$

$$\begin{aligned}\log L_{E_2} &= \textstyle\sum_{i=1}^{n}[\delta_i\omega_i \log h(t_i|Y=1) + \omega_i \log S(t_i|Y=1)] \\ &= \textstyle\sum_{i=1}^{m}\tau_i \log S(t_i|Y=1) + \sum_{i=m+1}^{n}[\log h(t_i|Y=1) + \log S(t_i|Y=1)],\end{aligned}$$

Relativamente a $\log L_{E_1}$, após alguns cálculos, obtemos a expressão do estimador de q na $(k+1)$-ésima iteração

$$q^{(k+1)} = \frac{1}{n}\Big[(n-m) + \sum_{i=1}^{m}\tau_i^{(k)}\Big].$$

Uma vez que estamos a usar a distribuição de Chen para os indivíduos susceptíveis, por (8), temos

$$\tau_i = \frac{q\exp[\lambda(1-exp(t_i^{\beta}))]}{1-q+q\exp[\lambda(1-exp(t_i^{\beta}))]}. \tag{9}$$

Em relação a $\log L_{E_2}$, uma vez que se escreve na forma

$$\begin{aligned}\log L_E = \;& \lambda\textstyle\sum_{i=1}^{m}\tau_i[1-\exp(t_i^{\beta})] + (n-m)(\log\lambda + \log\beta) + \\ & (\beta-1)\textstyle\sum_{i=m+1}^{n}\log t_i + \sum_{i=m+1}^{n} t_i^{\beta} + \lambda\sum_{i=m+1}^{n}[1-\exp(t_i^{\beta})],\end{aligned}$$

após alguns cálculos, é possível obter uma expressão explícita para o estimador de λ,

$$\hat{\lambda} = \frac{n-m}{\sum_{i=1}^{m}\tau_i\Big[\exp(t_i^{\beta})-1\Big] + \sum_{i=m+1}^{n}\Big[\exp(t_i^{\beta})-1\Big]},$$

onde τ_i é dado por (9), mas não para o estimador de β. Por conseguinte, é necessário usar simultaneamente um outro processo de maximização como, por exemplo, o método de Newton-Raphson.

4 Aplicações

Uma consequência habitual do desenvolvimento de um novo modelo é a sua aplicação a dados, quer estes sejam reais ou simulados. A nossa opção foi a aplicação a dados reais, com duplo objectivo. Por um lado, naturalmente, a verificação da sua exequibilidade e adequabilidade. Por outro, averiguar se apresenta vantagens em relação a outros modelos previamente estudados. Com este objectivo em mente, ajustámos este modelo a dois conjuntos de dados distintos.

Os ajustamentos foram efectuados recorrendo a algoritmos programados na linguagem R (Ihaka e Gentleman, 1996).

4.1 Dados de leucemia

Klein e Moeschberger (1998, p.10) referem um estudo envolvendo 101 pacientes com leucemia mielógena aguda avançada. A todos estes pacientes foi feito um transplante da medula óssea, 51 com a sua própria medula (os do grupo designado por *auto*) e 50 com medula de algum parente compatível (os do grupo designado por *allo*). O tempo de vida é definido como o tempo, em meses, desde o transplante até à recaída. Como o grupo *allo* tem uma apreciável quantidade de observações censuradas correspondentes aos valores mais elevados das observações, apresenta boas perspectivas de conter indivíduos curados, daí a razão de termos optado por trabalhar com os dados deste grupo. Além disso, o nivelamento da estimativa de Kaplan-Meier da função de sobrevivência após aproximadamente 20 meses (conforme a Fig. 1), num valor próximo de 0.5, é outra indicação da provável existência de indivíduos curados.

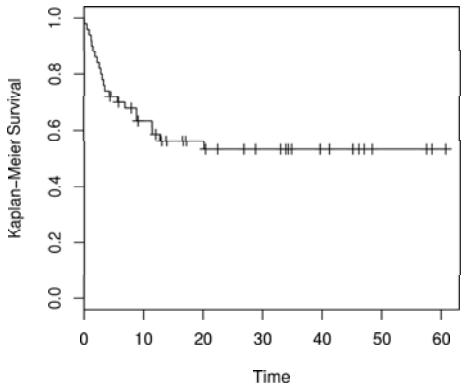

Figura 1: Estimativa de Kaplan-Meier da função de sobrevivência.

Neste caso, os valores iniciais que usámos no algoritmo foram $q^{(0)} = 0.46786$, $\lambda^{(0)} = 0.183395$ e $\beta^{(0)} = 0.5$. As estimativas que obtivemos foram $\hat{q} = 0.47060$

(0.08809), $\hat{\lambda} = 0.12508(0.02965)$ e $\hat{\beta} = 0.42100(0.03058)$, onde os valores entre parêntesis indicam o desvio padrão correspondente. Ao representarmos a correspondente função de sobrevivência, constatámos que o ajustamento aos dados é quase perfeito, como está patente na Fig. 2.

Apesar de parecer evidente a superioridade do nosso modelo, comparámos o valor da estatística $-2\log\hat{L}$ para o modelo de cura baseado na distribuição log-logística (Abreu, 2004, p.104) e para o novo modelo que propomos. De acordo com o esperado, no primeiro caso foi 386.4602 e no segundo apenas 106.2315. De referir ainda o valor ligeiramente inferior do desvio padrão de $\hat{q}$ quando se utiliza a distribuição de Chen: 0.08809 em vez de 0.09087.

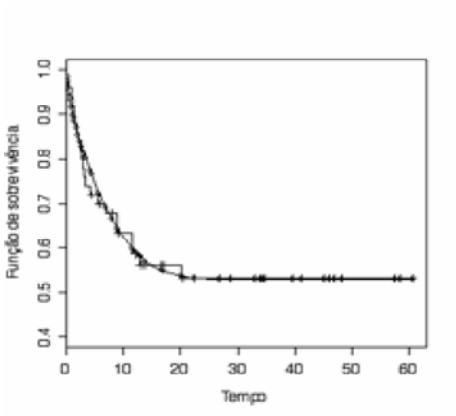

Figura 2: Curvas de sobrevivência correspondentes à estimativa de Kaplan-Meier e ao novo modelo de cura (dados de Klein e Moeschberger, 1998).

Yu *et al.* (2004) salientam que a duração do *follow-up* tem um papel importante na obtenção de melhores estimativas dos parâmetros nos modelos de cura de mistura. De facto, estes autores enfatizam a necessidade de o *follow-up* ser maior do que o valor da mediana do tempo de vida dos indivíduos susceptíveis, por forma a que a maioria dos acontecimentos possa ser observada antes do fim do *follow-up*. Os dados que escolhemos estão precisamente nessa situação, já que a mediana é 4.25 meses e a maior observação é 60.625 meses.

4.2 Outros dados de leucemia

Apesar do quase perfeito ajustamento do modelo de cura com a distribuição de Chen aos dados anteriores, decidimos considerar outros dados para verificar se o modelo era suficientemente flexível para se ajustar bem aos novos dados. Kersey *et al.* (1987) compararam o efeito do tratamento da leucemia através de transplante de medula óssea, em 46 doentes com a sua própria medula (grupo 2)

e em 44 doentes com medula de algum parente compatível (grupo 1). O tempo de vida é definido como o tempo, em anos, desde o transplante até à recaída. Devido à grande quantidade de observações censuradas correspondentes aos valores mais elevados das observações em ambos os grupos, optámos por efectuar o ajustamento para cada um dos grupos em separado.

Assim, para o grupo 1, os valores iniciais foram $q^{(0)} = 0.73662$, $\lambda^{(0)} = 0.9930455$ e $\beta^{(0)} = 1$, e as estimativas foram $\hat{q} = 0.7285978$, $\hat{\lambda} = 0.7611151$ e $\hat{\beta} = 0.6139747$. Com estes valores das estimativas, obtivemos o gráfico da função de sobrevivência ajustada (Fig. 3), a qual, uma vez mais, se ajustou bastante bem aos dados.

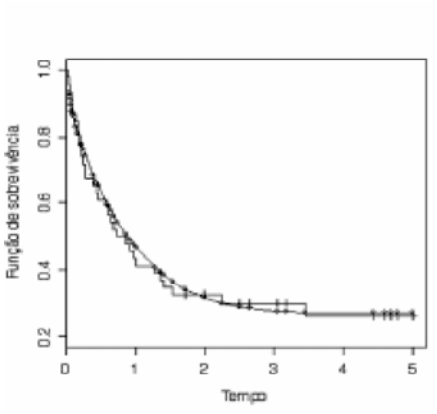

Figura 3: Curvas de sobrevivência correspondentes à estimativa de Kaplan-Meier e ao modelo de cura com a distribuição de Chen (dados de Kersey *et al.*, 1987, grupo1).

Para o grupo 2, os valores iniciais foram $q^{(0)} = 0.8134885$, $\lambda^{(0)} = 0.8246753$ e $\beta^{(0)} = 0.5$, e as estimativas foram $\hat{q} = 0.8181942$, $\hat{\lambda} = 1.7420948$ e $\hat{\beta} = 0.9199423$. De modo análogo ao grupo 1, representámos a função de sobrevivência ajustada, patente na Fig. 4. Podemos verificar que, embora o ajustamento não seja tão bom como no grupo 1, ainda é bastante razoável.

Este mesmo tipo de diferença na qualidade do ajustamento em relação aos dois grupos ocorreu ao ser considerado um modelo de cura baseado na distribuição de Weibull (Maller e Zhou, 1996, p.108 e p.117) e na distribuição de Burr de tipo XII (Shao e Zhou, 2004), o que reforça o facto de o nosso modelo ser uma alternativa a ter em conta.

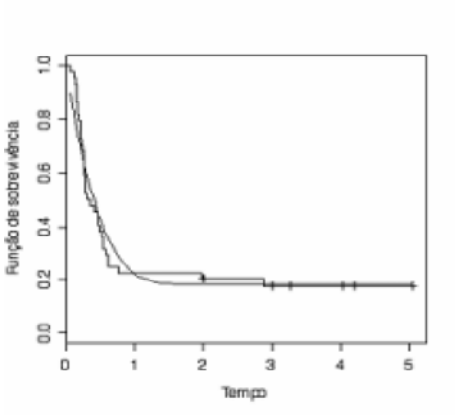

Figura 4: Curvas de sobrevivência correspondentes à estimativa de Kaplan-Meier e ao modelo de cura com a distribuição de Chen (dados de Kersey *et al.*, 1987, grupo2).

5 Comentários finais

À medida que a cura de certas doenças se torna, cada vez mais, uma realidade na Medicina, o interesse por modelos de sobrevivência que permitem a inclusão de indivíduos curados aumenta.

O nosso objectivo neste artigo foi o de reforçar o uso de modelos de cura, quando tal se torna adequado, apresentando um novo modelo. Sendo baseado numa distribuição recente e ainda pouco divulgada, a distribuição de Chen, esta foi também uma forma de alargar o leque de escolhas para a distribuição do tempo de vida. Além disso, pelos resultados apresentados nas aplicações, esta distribuição parece ter flexibilidade suficiente para se tornar adequada em várias situações.

A distribuição de Chen modificada está ainda numa fase embrionária, com vários campos por explorar. O mais imediato tem a ver com a questão da estimação dos parâmetros do modelo e eventuais aplicações.

Também por explorar encontra-se o desenvolvimento de modelos de regressão, a partir dos dois modelos propostos.

Agradecimentos

Este trabalho foi elaborado com o apoio parcial do CITMA (para Ana M. Abreu) e da FCT / POCTI / FEDER (para Cristina S. Rocha).

Referências

[1] Abreu, A.M. (2004). *Modelos de Sobrevivência para Populações com Indivíduos Imunes.* Tese de Doutoramento. Universidade da Madeira.

[2] Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics and Probability Letters*, Vol.49, p. 155-161.

[3] Dempster, A.P., Laird, N.M. e Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of Royal Statistical Society*, B, Vol.39, p. 1-38.

[4] Ihaka, R. e Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, Vol.5, p. 299-314.

[5] Kersey, J.H., Weisdorf, D., Nesbit, M.E., LeBien, T.W., Woods, W.G., McGlave, P.B., Kim, T., Vallera, D.A., Goldman, A.I., Bostrom, B., Hurd, D. e Ramsay, N.K.C. (1987). Comparison of autologous and allogeneic bone marrow transplantation for treatment of high-risk refractary acute lymphoblastic leukaemia. *New England Journal of Medicine*, Vol.317, p. 461-467.

[6] Klein, J.P. e Moeschberger, M.L. (1998). *Survival Analysis. Techniques for Censored and Truncated Data*, 2ª impressão corrigida. Springer-Verlag, New York.

[7] Maller, R.A. e Zhou, S. (1996). *Survival Analysis with Long-Term Survivors.* John Wiley, New York.

[8] Shao, Q. e Zhou, X. (2004). A new parametric model for survival data with long-term survivors. *Statistics in Medicine*, Vol.23, p. 3525-3543.

[9] Steele, F., (2003). A discrete-time multilevel mixture model for event history data with long-term survivors, with an application to an analysis of contraceptive sterilisation in Bangladesh. *Lifetime Data Analysis*, Vol.9, p. 155–174.

[10] Tsodikov, A., (1998). A proportional hazards model taking account of long-term survivors. *Biometrics*, Vol.54, p. 1508-1516.

[11] Yakovlev, A.Y., Asselain, B., Bardou, V.J. Fourquet, A., Hoang, R., Rochefediere, A., e Tsodikov, A. D. (1993). *A simple stochastic model of tumor recurrence and its applications to data on premenopausal breast cancer.* Em *Biometrie et Analyse de Donnees Spatio-Temporelles.* Editores: Asselain, B., Boniface, M., Duby, C., Lopez, C., Masson, J.P. e Tranchefort, J., Vol.12, p. 66-82. Societé Française de Biométrie, ENSA Rennes, France.

[12] Yu, B., Tiwari, R.C., Cronin, K.A. e Feuer, E.J., (2004). Cure fraction from the mixture cure models for grouped survival data. *Statistics in Medicine*, Vol.23, p. 1733–1747.