Estimação em Modelos de Sobrevivência

Antonio Carlos Pedroso de Lima

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE MESTRE
EM
ESTATÍSTICA

Área de Concentração: **Estatística**
Orientador: **Prof. Dr. Heleno Bolfarine**

–São Paulo, Dezembro de 1988–

AGRADECIMENTOS

- Ao Heleno, pela orientação e pelas discussões que levaram ao tema deste trabalho.
- A Beti, pelo apoio e grande colaboração em momentos difíceis.
- Ao Koba, Lisa, Angela, Coelho, Célia e Fernando, pela amizade.
- Ao Beto, pelo desenho dos gráficos .
- Ao Clóvis, pelo incentivo e amizade a toda hora demonstrados.
- Ao Gubi, pelas infindáveis discussões a respeito do TEX.
- Ao Nélson, pelas discussões sobre alguns aspectos deste trabalho.
- A todos os membros do departamento de estatística, incluindo os colegas da pós-graduação, cujo relacionamento diário tem sido enriquecedor.

ÍNDICE

Capítulo 0. Introdução 1

Capítulo 1. Estimação da Função de Sobrevivência 5
1. Introdução 6
2. Estimação não paramétrica 7
2.1. Estimador da amostra reduzida (R.S.) 8
2.2. Estimador da tábua de vida padrão (S.D.) 9
2.3. Estimador produto–limite (P.L.) 17
3. Estimação paramétrica 29
3.1. Método iterativo de Newton–Raphson 30
3.2. Distribuição Assintótica 31
4. Estimação não paramétrica x estimação paramétrica 33
4.1. Tempos de vida com distribuição exponencial 33
4.2. Tempos de vida com distribuição Weibull 35
4.3. Comentários 38

Capítulo 2. Estimação em Modelos de Regressão 40
1. Introdução 41
2. Estimação não paramétrica 42
2.1. Estimadores de Miller 42
2.2. Estimadores de Buckley–James 46
2.3. Método de Cox 50
3. Estimação paramétrica 55
3.1. Estimação de máxima verossimilhança 56
4. Estimação não paramétrica x estimação paramétrica 58
4.1. Amostras sem censuras 58
4.2. Amostras censuradas 61
4.2.1. Tempos de vida com distribuição exponencial 61
4.2.2. Tempos de vida com distribuição gama 63
4.3. Comentários 64

Capítulo 3. Estimação Bayesiana em Modelos de Sobrevivência 67
1. Introdução 68
2. Método de aproximação de Laplace 69
3. Modelo log-linear com erro gama generalizado 70
3.1. Distribuição a posteriori para α, β, σ e k 71
3.2. Comparação de dois tratamentos 74
3.3. Exemplo 76
4. Transformação para a distribuição de valor extremo 78
4.1. Densidade a posteriori para μ, σ e λ 78
4.2. Exemplo 81
5. Densidade preditiva com um modelo gama generalizado 83
5.1. Densidade preditiva para σ e k conhecidos 83
5.2. Comparação de dois tratamentos 85
5.3. Exemplo 85

Apêndice 1. O Uso do GLIM em Modelos de Sobrevivência 87

Referências Bibliográficas 96

Capítulo 0

Introdução

A análise de sobrevivência está estreitamente relacionada com a observação do tempo necessário para a ocorrência de algum resultado experimental no qual se tem interesse. Para exemplificar, suponha um experimento que consiste em observar o tempo de sobrevida de pacientes submetidos a um certo tratamento, com o intuito de verificar a proporção de indivíduos vivos no decorrer do tempo. Esta situação, apesar de bastante comum entre pesquisadores da área médica, ocorre em diversas outras áreas como por exemplo em engenharia, onde n componentes podem ser colocados em teste e são observados seus instantes de falha. Em ambos os exemplos, o interesse principal está relacionado ao tempo decorrido até que um determinado evento se verifique. Mais formalmente, pode-se considerar que as observações consistem de realizações da variável aleatória não-negativa T, que representa o *tempo de sobrevivência* ou *instante de falha*. Suponha que associada a esta variável existe uma função densidade de probabilidade $f(t)$. Segue que a proporção de sobreviventes inicialmente mencionada será representada pela *função de sobrevivência*, dada por

$$S(t) = P\{T > t\} = \int_t^{+\infty} f(x)dx.$$

que representa a probabilidade de T exceder a um dado t. Desta forma, uma primeira questão que se coloca diz respeito à estimação de $S(t)$.

A função de sobrevivência e a função densidade de probabilidade podem ser escritas através da *função de risco* $\lambda(t)$, dada por

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

de onde são tiradas as relações

$$S(t) = \epsilon^{-\int_0^t \lambda(u)du} = \epsilon^{-\Lambda(t)}$$

e

$$f(t) = \lambda(t)\epsilon^{-\int_0^t \lambda(u)du} = \lambda(t)\epsilon^{-\Lambda(t)},$$

onde $\Lambda(t) = \int_0^t \lambda(u)du$ representa a função de risco *acumulada*.

Algumas vezes, o tempo gasto até a ocorrência do evento de interesse não pode ser observado por razões externas ao processo experimental. Para unidades experimentais nesta situação são observados apenas os instantes de *censura*, que correspondem ao instante em que se perde contato com a unidade experimental. Os processos de censura são comumente classificados em três tipos: censura de *tipo I*, *tipo II* e censura *aleatória*.

A censura de tipo I ocorre frequentemente em experimentos que têm período de duração fixado. Por exemplo, suponha que os n componentes colocados em teste só são acompanhados durante um certo período de tempo, após o qual o experimento é encerrado. Aquelas unidades que permaneciam em funcionamento quando do término do experimento não terão mais o seu tempo de falha observado; no entanto, é sabido que não falharam até o final do ensaio. Portanto, os tempos de vida das unidades experimentais só serão observados se forem menores ou iguais ao tempo de execução do experimento; caso contrário, consistem de observações censuradas.

A censura de tipo II é típica de experimentos que terminam a partir do momento em que um certo número de falhas é atingido. No exemplo anterior, suponha que o custo dos componentes em teste é considerável, de modo que o pesquisador só se dispõe a perder r componentes, após o que o ensaio é encerrado. Nesta situação, se o experimento tem início com n unidades experimentais, após a r-ésima falha ($r \leq n$) ocorrerão $(n - r)$ observações censuradas.

A censura aleatória ocorre quando a cada unidade experimental é associado um limite de obervação com tempo também aleatório, a partir do qual a unidade não poderá ter seu tempo de falha observado. Novamente no caso dos n componentes colocados em teste, suponha que variações de voltagem, que não são controladas pelo pesquisador danificam alguns dos componentes, impedindo a observação dos instantes de falha. Se estas variações ocorrem em tempos aleatórios, os limites de observação de cada componente serão dados pelo tempo que demora para ocorrer a variação de voltagem para aquele componente. Configura-se então um esquema de censuras aleatórias.

As censuras de tipo I e aleatória são mais comuns na área médica, ao passo que a censura de tipo II é frequentemente mencionada em experimentos na área de engenharia.

Alguns autores afirmam que o aparecimento da análise de sobrevivência ocorreu conjuntamente com as tabelas de mortalidade a mais de um século; posteriormente, sofreu grandes avanços durante a segunda guerra mundial, quando a confiabilidade de equipamentos militares se tornou estratégica. Tal fato originou o desenvolvimento de procedimentos paramétricos de estimação. Mais recentemente, a utilização em ensaios médicos motivou a elaboração de alguns procedimentos não paramétricos. Ambos os enfoques têm sido comparados por Miller(1983), considerando a função de sobrevivência e amostras censuradas. Cox e Hinkley(1968) também comparam ambas abordagens, porém em modelos de regressão e sem levar em conta a possibilidade de censuras.

O objetivo deste trabalho é enfocar alguns métodos de estimação desenvolvidos

em análise de sobrevivência, considerando-se na maior parte das vezes a ocorrência de observações censuradas sob o esquema de censuras aleatórias. No capítulo 1 são abordados aspectos da estimação da função de sobrevivência, considerando-se abordagens não-paramétrica e paramétrica. Os dois métodos são comparados assintóticamente em alguns casos particulares e para pequenas amostras são obtidas indicações a respeito do comportamento dos estimadores através de amostras simuladas. No capítulo 2 são considerados procedimentos de estimação em modelos de regressão, também sob o enfoque não-paramétrico e paramétrico, que são comparados através de simulações. Finalmente, no capítulo 3 é estudada a estimação de modelos paramétricos sob o ponto de vista Bayesiano.

Capítulo 1

Estimação da Função de Sobrevivência

1. Introdução

Suponha que $T_1, \ldots, T_n$ sejam variáveis aleatórias não negativas e independentes com distribuição comum $F^0(t) = P\{T_i \leq t\}$, representando tempos de vida ; suponha também que cada T_i "compete" com L_i, $i = 1, \ldots, n$, no sentido de que o tempo efetivamente observado é:

$$t_i = \min\{T_i, L_i\}. \tag{1.1}$$

A cada t_i é associada a quantidade

$$\delta_i = I_{\{T_i \leq L_i\}}. \tag{1.2}$$

Os termos $L_1, \ldots, L_n$ podem ser variáveis aleatórias não negativas ou constantes e são denominados limites de observação, supostos independentes de $T_1, \ldots, T_n$. Portanto utilizando os termos usuais em análise de sobrevivência,

$$\delta_i = 1 \Rightarrow T_i \leq L_i \Rightarrow t_i = T_i \Rightarrow \text{ocorre uma } \textit{morte} \text{ ;}$$
$$\delta_i = 0 \Rightarrow T_i > L_i \Rightarrow t_i = L_i \Rightarrow \text{ocorre uma } \textit{censura} \text{ .}$$

Assim, morte e censura formam duas classes de eventos mutuamente exclusivos.

Este capítulo tem como objetivo estudar a estimação da proporção de sobreviventes além de um instante especificado t; isto é, a estimação da função de sobrevivência, definida por

$$S(t) = 1 - F^0(t) = P\{T > t\}. \tag{1.3}$$

Caso a forma da distribuição dos tempos de vida seja desconhecida, a classe de distribuições admissíveis torna-se geral, ou seja, a classe de todas as distribuições, configurando-se então um procedimento de estimação *não paramétrico* .

Por outro lado, se a forma da distribuição é conhecida, a estimação de $S(t)$ consiste em selecionar da classe de distribuições admissíveis, aquela que melhor se ajusta aos dados, através de algum procedimento *paramétrico* de estimação.

2. Estimação Não Paramétrica

Além de ser uma alternativa natural para a estimação no caso da falta de informação a respeito da forma da distribuição, os métodos não paramétricos fornecem indicações de possíveis modelos paramétricos.

Na ausência de censuras um estimador não paramétrico de (1.3) é obtido através da função de distribuição empírica, sendo dado por

$$\hat{S}(t) = \frac{\text{num. de tempos } > t}{n}. \tag{2.1}$$

Na presença de censuras (2.1) não poderá ser utilizado sendo necessária a obtenção de estimadores mais gerais. Kaplan e Meier (1958) estudam a estimação não paramétrica com observações incompletas discutindo propriedades do estimador *amostra reduzida* ou *reduced-sample* (R.S.), estimadores *atuariais*, entre os quais o *estimador da tábua de vida padrão* ou *standart life estimate* (S.D.), que são aproximações para o estimador *produto limite* (P.L.), sendo este último o mais utilizado atualmente.

A fim de introduzir as notações que serão utilizadas nas próximas seções, suponha que o período de observação de n indivíduos seja $[0, T]$. e que este intervalo é particionado segundo constantes $0 = \xi_0 < \xi_1 < \ldots < \xi_k < \xi_{k+1} = T$. Em cada intervalo $I_j = (\xi_{j-1}, \xi_j]$ serão calculadas estimativas das *probabilidades condicionais de morte*

$$q_j = P\{T_i \leq \xi_j \mid T_i > \xi_{j-1}\} = \frac{P\{\xi_{j-1} < T_i \leq \xi_j\}}{P\{T_i > \xi_{j-1}\}} = \frac{F^0(\xi_j) - F^0(\xi_{j-1})}{1 - F^0(\xi_{j-1})}. \tag{2.2}$$

de tal forma que $1 - F^0(\xi_j) = p_1 p_2 \cdots p_j$, onde $p_i = 1 - q_i$. Para cada intervalo I_j podem ser calculadas as estatísticas: $N_{1j} = \sum_{i=1}^{n} I_{\{t_i > \xi_{j-1}, L_i > \xi_j\}}$, $N_{2j} = \sum_{i=1}^{n} I_{\{t_i > \xi_{j-1}, L_i \in I_j\}}$, $D_{1j} = \sum_{i=1}^{n} I_{\{t_i \in I_j, L_i > \xi_j, \delta_i = 1\}}$, $D_{2j} = \sum_{i=1}^{n} I_{\{t_i \in I_j, L_i \in I_j, \delta_i = 1\}}$ e $W_j = \sum_{i=1}^{n} I_{\{t_i \in I_j, \delta_i = 0\}}$. Definem-se então $N_j = N_{1j} + N_{2j}$, o número de indivíduos *em risco* em I_j e $D_j = D_{1j} + D_{2j}$, o número de indivíduos não censurados que não sobrevivem a I_j.

Quando uma censura ocorre no instante ξ_j, desloca-se seu instante uma quantidade infinitesimal para a direita, de modo que as quantidades acima ficam perfeitamente definidas inclusive no caso em que $T_i = L_i$, isto é, quando morte e censura ocorrem simultaneamente.

Utilizando a notação adotada serão definidos a seguir o estimador amostra reduzida (R.S.), o estimador tábua de vida padrão (S.D.) e o estimador produto-limite (P.L.).

2.1 Estimador da Amostra Reduzida (R.S.)

O estimador R.S. é o mais simples e intuitivo, uma vez que utiliza apenas as observações não censuradas para estimar $S(t)$. Isto implica que todos os limites de observação devem ser conhecidos. É definido por:

$$q_j^* = \frac{D_{1j}}{N_{1j}}. \tag{2.3}$$

Supondo um esquema de censuras aleatórias independentes dos tempos de vida T_i com função de distribuição acumulada dada por

$$H(t) = P\{L_i \leq t\}, \tag{2.4}$$

tem-se para o primeiro intervalo I_1,

$$\begin{aligned} E(D_{11}) &= E(\sum_{i=1}^{n} I_{\{t_i \in I_1, L_i > \xi_1, \delta_i = 1\}}) = \sum_{i=1}^{n} P\{T_i \in I_1, L_i > \xi_1\} \\ &= nF^0(\xi_1)[1 - H(\xi_1)] \end{aligned}$$

e

$$\begin{aligned} E(N_{11}) &= E(\sum_{i=1}^{n} I_{\{t_i > \xi_0, L_i > \xi_1\}}) = \sum_{i=1}^{n} P\{T_i > \xi_0, L_i > \xi_1\} \\ &= n[1 - H(\xi_1)]. \end{aligned}$$

Portanto, como D_{11} e N_{11} são somas de n variáveis de Bernoulli, segue pela Lei Forte dos Grandes Números que:

$$\frac{D_{11}}{n} \xrightarrow{\text{q.c.}} F^0(\xi_1)[1 - H(\xi_1)]$$

e

$$\frac{N_{11}}{n} \xrightarrow{\text{q.c.}} 1 - H(\xi_1),$$

de modo que

$$\frac{D_{11}}{N_{11}} \xrightarrow{\text{q.c.}} F^0(\xi_1).$$

Ou seja, $p_1^* = 1 - q_1^*$ é um estimador consistente para $S(\xi_1)$. Estes resultados podem ser generalizados para os outros intervalos.

O inconveniente no estimador R.S. é o fato de que a informação contida em observações censuradas é desprezada, do que decorre a sua pouca utilização e o desenvolvimento de estimadores mais refinados.

2.2 Estimador Tábua de Vida Padrão (S.D.)

O método da tábua de vida padrão assume que, em média, indivíduos censurados em I_j permanecem em risco durante metade do intervalo. Define-se então o *número efetivo de indivíduos em risco* no intervalo I_j como:

$$N'_j = N_j - \frac{1}{2}W_j. \tag{2.5}$$

O estimador S.D. pode ser definido a partir de (2.5), sendo dado por

$$\tilde{q}_j = \frac{D_j}{N'_j}. \tag{2.6}$$

Portanto uma estimativa de $S(\xi_k)$ será dada por

$$\tilde{S}_k = \tilde{S}(\xi_k) = \prod_{j=1}^{k}(1 - \tilde{q}_j). \tag{2.7}$$

Para estudar as propriedades assintóticas de (2.6) será utilizado o enfoque adotado por Breslow e Crowley(1974). Inicialmente note que

$$\begin{aligned}\tilde{q}_j = \frac{D_j}{N_j - \frac{1}{2}W_j} &= \frac{D_{1j}}{N_j - \frac{1}{2}W_j} + \frac{D_{2j}}{N_j - \frac{1}{2}W_j} \\ &= \frac{N_{1j}}{N_j - \frac{1}{2}W_j}\left(\frac{D_{1j}}{N_{1j}}\right) + \frac{N_{2j} - \frac{1}{2}W_j}{N_j - \frac{1}{2}W_j}\left(\frac{D_{2j}}{N_{2j} - \frac{1}{2}W_j}\right).\end{aligned} \tag{2.8}$$

Isto é, a expressão (2.6) pode ser escrita como uma combinação convexa envolvendo o estimador R.S. e o estimador S.D. restrito aos indivíduos com limites de observação em I_j. Por (2,8) nota-se que nos intervalos onde não há censura o estimador S.D. é igual ao estimador R.S..

Com o intuito de estudar a consistência de (2.8), considere o intervalo I_1. Pode-se então, supondo censura aleatória com distribuição dada por (2.4), calcular a esperança de D_{21}, N_{21} e W_1:

$$\begin{aligned}E(D_{21}) &= E(\sum_{i=1}^{n} I_{\{t_i \in I_1, L_i \in I_1, \delta_i = 1\}}) = \sum_{i=1}^{n} P\{0 < T_i \leq L_i \leq \xi_1\} \\ &= n\int_0^{\xi_1} F^0(u)dH(u), \\ E(N_{21}) &= E(\sum_{i=1}^{n} I_{\{t_i > \xi_0, L_i \in I_1\}}) = \sum_{i=1}^{n} P\{L_i \leq \xi_1\} \\ &= nH(\xi_1)\end{aligned}$$

e

$$E(W_1) = E(\sum_{i=1}^{n} I_{\{t_i \in I_1, \delta_i = 0\}}) = \sum_{i=1}^{n} P\{0 < L_i < T_i \leq \xi_1\}$$
$$= n \int_0^{\xi_1} [1 - F^0(u)] dH(u).$$

Logo,

$$\frac{D_{21}}{n} \xrightarrow{\text{q.c.}} \int_0^{\xi_1} F^0(u) dH(u),$$
$$\frac{N_{21}}{n} \xrightarrow{\text{q.c.}} H(\xi_1)$$

e

$$\frac{W_1}{n} \xrightarrow{\text{q.c.}} \int_0^{\xi_1} [1 - F^0(u)] dH(u). \tag{2.9}$$

Assim sendo, partindo de (2.8) restrito ao primeiro intervalo e lembrando a consistência do estimador R.S., segue que

$$\tilde{q}_1 \xrightarrow{\text{q.c.}} \frac{1 - H(\xi_1)}{1 - \frac{1}{2}\int_0^{\xi_1} [1 - F^0(u)] dH(u)} F^0(\xi_1) +$$
$$+ \frac{H(\xi_1) - \frac{1}{2}\int_0^{\xi_1} [1 - F^0(u)] dH(u)}{1 - \frac{1}{2}\int_0^{\xi_1} [1 - F^0(u)] dH(u)} \frac{\int_0^{\xi_1} F^0(u) dH(u)}{H(u) - \frac{1}{2}\int_0^{\xi_1} [1 - F^0(u)] dH(u)}.$$

Portanto, $\tilde{q}_1$ será consistente para $F^0(\xi_1)$ se e somente se

$$F^0(\xi_1) = \frac{\int_0^{\xi_1} F^0(u) dH(u)}{H(u) - \frac{1}{2}\int_0^{\xi_1} [1 - F^0(u)] dH(u)}, \qquad 0 < \xi_1 < T. \tag{2.10}$$

Para o caso em que $H(x)$ é absolutamente contínua, Breslow e Crowley(1974) resolvem (2.10) utilizando equações diferenciais, de modo que o estimador S.D. será consistente em cada ponto ξ_i se e somente se

$$F^0(x) = 1 - \left(\frac{1}{1 + cH(x)}\right)^{\frac{1}{2}}, \tag{2.11}$$

para algum $c > 0$.

Dada a forte restrição imposta por (2.11), o estimador S.D. geralmente será inconsistente. Crowley(1970) estudou o vício assintótico através de simulações, considerando censuras distribuídas uniformemente em $[0, T]$ e distribuições de sobrevivência uniforme

e exponencial. Em ambos os casos ele observa que à medida que é aumentado o número de partições em $[0,T]$ ocorre uma diminuição no vício assintótico, isto é, à medida que diminue a amplitude dos intervalos I_j o estimador S.D. tende a ser não viciado.

Para obter a distribuição assintótica do estimador S.D., considere o vetor $\tilde{\mathbf{q}}_j = (\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_j)'$. Segue o teorema:

Teorema 1. *A distribuição assintótica do vetor normalizado* $n^{\frac{1}{2}}(\tilde{\mathbf{q}}_j - \mathbf{q}_j^*)$ *é normal multivariada com vetor de médias* $\mathbf{0}$ *e matriz de covariância diagonal, cujos elementos são dados por*

$$(\pi_k^N - \frac{1}{2}\pi_k^W)^{-1}\left[q_k^* - (q_k^*)^2\frac{(\pi_k^N - \frac{1}{4}\pi_k^W)}{(\pi_k^N - \frac{1}{2}\pi_k^W)}\right] \quad k = 1, \ldots, j, \tag{2.12}$$

onde $q_j^* = \frac{\pi_j^D}{(\pi_j^N - \frac{1}{2}\pi_j^W)}$ *com* π_j^D, π_j^N *e* π_j^W *representando, respectivamente, probabilidades de morte, presença e censura no intervalo* I_j.

Prova:

Considerando os j primeiros intervalos em $[0,T]$, tem-se que dada uma observação, necessariamente ela estará contida em uma e apenas uma das categorias definidas pelos elementos do vetor

$$\mathbf{M}_j = (D_1, W_1, \ldots, D_j, W_j, N_{j+1})'. \tag{2.13}$$

que tem distribuição multinomial cujos parâmetros são representados por

$$\mathbf{\Pi}_j = (\pi_1^D, \pi_1^W, \ldots, \pi_j^D, \pi_j^W, \pi_{j+1}^N)'. \tag{2.14}$$

Os elementos de (2.14) podem ser calculados a partir de expressões derivadas de (2.9).

Das relações de convergência tiradas anteriormente, decorre que $\tilde{q}_j$ converge quase certamente para

$$q_j^* = \frac{\pi_j^D}{\pi_j^N - \frac{1}{2}\pi_j^W}.$$

Dada a distribuição de $\mathbf{M}_j$, temos pelo Teorema Limite Central multivariado (Rao, 1973) que

$$n^{\frac{1}{2}}(\mathbf{M}_j - n\mathbf{\Pi}_j) \xrightarrow{D} N_j(\mathbf{0}, \mathbf{\Sigma}), \tag{2.15}$$

com

$$\mathbf{\Sigma} = \begin{pmatrix} \pi_1^D(1-\pi_1^D) & -\pi_1^W\pi_1^D & -\pi_1^D\pi_2^D & \ldots & -\pi_1^D\pi_{j+1}^N \\ -\pi_1^W\pi_1^D & \pi_1^W(1-\pi_1^W) & -\pi_1^W\pi_2^D & \ldots & -\pi_1^W\pi_{j+1}^N \\ -\pi_2^D\pi_1^D & -\pi_2^D\pi_1^W & \pi_2^D(1-\pi_2^D) & \ldots & -\pi_D^2\pi_{j+1}^N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\pi_{j+1}^N\pi_1^D & -\pi_{j+1}^N\pi_1^W & -\pi_{j+1}^N\pi_2^D & \ldots & \pi_{j+1}^N(1-\pi_{j+1}^N) \end{pmatrix}.$$

Uma vez que $\tilde{\mathbf{q}}_j$ é uma função contínua de elementos de (2.13), de (2.15) vem que (Rao, 1973, 6a.2, (*ii*))

$$n^{\frac{1}{2}}(\tilde{\mathbf{q}}_j - \mathbf{q}_j^*) \xrightarrow{D} N_j(\mathbf{0}, \boldsymbol{\Sigma}^*). \tag{2.16}$$

Para obter a matriz $\boldsymbol{\Sigma}^*$ considere inicialmente a expansão em série de Taylor de $\tilde{q}_j - q_j^*$ ao redor do ponto $n(\pi_j^N, \pi_j^D, \pi_j^W)$:

$$\begin{aligned}
&\tilde{q}_j - q_j^* = \\
&= \left(\frac{-\pi_j^D}{n(\pi_j^N - \frac{1}{2}\pi_j^W)^2}, \frac{1}{n(\pi_j^N - \frac{1}{2}\pi_j^W)}, \frac{\frac{1}{2}\pi_j^D}{n(\pi_j^N - \frac{1}{2}\pi_j^W)^2} \right) \begin{pmatrix} N_j - n\pi_j^N \\ D_j - n\pi_j^D \\ W_j - n\pi_j^W \end{pmatrix} + O_p(n^{-1}) \\
&= \frac{(\pi_j^N - \frac{1}{2}\pi_j^W)^{-2}}{n} \left(-\pi_j^D, \pi_j^N - \frac{1}{2}\pi_j^W, \frac{1}{2}\pi_j^D \right) \left\{ \begin{pmatrix} N_j \\ D_j \\ W_j \end{pmatrix} - n \begin{pmatrix} \pi_j^N \\ \pi_j^D \\ \pi_j^W \end{pmatrix} \right\} + O_p(n^{-1}) \\
&= \frac{(\pi_j^N - \frac{1}{2}\pi_j^W)^{-2}}{n} \mathbf{c}_j'(\mathbf{v}_j - n\mathbf{p}_j) + O_p(n^{-1}),
\end{aligned} \tag{2.17}$$

onde

$$\begin{aligned}
\mathbf{c}_j' &= (-\pi_j^D, \pi_j^N - \tfrac{1}{2}\pi_j^W, \tfrac{1}{2}\pi_j^D), \\
\mathbf{v}_j' &= (N_j, D_j, W_j)
\end{aligned}$$

e

$$\mathbf{p}_j' = (\pi_j^N, \pi_j^D, \pi_j^W).$$

A partir da expressão (2.17) pode-se calcular o termo (k, l) de $\boldsymbol{\Sigma}^*$, para $k < l$:

$$\begin{aligned}
&Cov\{n^{\frac{1}{2}}(\tilde{q}_k - q_k^*), n^{\frac{1}{2}}(\tilde{q}_l - q_l^*)\} \cong \\
&\quad \cong n \frac{(\pi_k^N - \frac{1}{2}\pi_k^W)^{-2}}{n} \frac{(\pi_l^N - \frac{1}{2}\pi_l^W)^{-2}}{n} Cov\{\mathbf{c}_k'(\mathbf{v}_k - n\mathbf{p}_k), \mathbf{c}_l'(\mathbf{v}_l - n\mathbf{p}_l)\} \\
&\quad = \frac{(\pi_k^N - \frac{1}{2}\pi_k^W)^{-2}(\pi_l^N - \frac{1}{2}\pi_l^W)^{-2}}{n} \mathbf{c}_k' Cov(\mathbf{v}_k, \mathbf{v}_l)\mathbf{c}_l.
\end{aligned}$$

Para calcular a covariância entre $\mathbf{v}_k$ e $\mathbf{v}_l$, utilizam-se as relações

$$N_j = N_{j-1} - D_{j-1} - W_{j-1} \Rightarrow N_l = N_k - \sum_{j=k}^{l-1}(D_j + W_j) \tag{2.18}$$

e

$$\sum_{j=k}^{l-1}(\pi_j^D + \pi_j^W) = \pi_k^N - \pi_l^N, \tag{2.19}$$

sendo a última devida à distribuição de $\mathbf{M}_k$ e $\mathbf{M}_l$.

Portanto, usando (2.18) e propriedades da covariância,

$$\begin{aligned} Cov(N_k, N_l) &= Cov\{N_l + \sum_{j=k}^{l-1}(D_j + W_j), N_l\} \\ &= Var(N_l) + \sum_{j=k}^{l-1}\{Cov(D_j, N_l) + Cov(W_j, N_l)\} \\ &= n\pi_l^N(1-\pi_l^N) - n\sum_{j=k}^{l-1}(\pi_j^D\pi_l^N + \pi_j^W\pi_l^N) \\ &= n\pi_l^N\{1-\pi_l^N - \sum_{j=k}^{l-1}(\pi_j^D + \pi_j^W)\}. \end{aligned}$$

Utilizando a relação (2.19) vem que

$$Cov(N_k, N_l) = n\pi_l^N(1-\pi_l^N-\pi_k^N+\pi_l^N) = n\pi_l^N(1-\pi_k^N).$$

Da mesma forma,

$$\begin{aligned} Cov(N_k, D_l) &= Cov\{N_{l+1} + \sum_{j=k}^{l}(D_j + W_j), D_l\} \\ &= Cov(N_{l+1}, D_l) + \sum_{j=k}^{l}\{Cov(D_j, D_l) + Cov(W_j, D_l)\} \\ &= -n\pi_{l+1}^N\pi_l^D - n\sum_{j=k}^{l-1}(\pi_j^D\pi_l^D + \pi_j^W\pi_l^D) + n\pi_l^D(1-\pi_l^D) - n\pi_l^D\pi_l^W \\ &= n\pi_l^D\{1-\pi_{l+1}^N - \sum_{j=k}^{l}(\pi_j^D + \pi_j^W)\}, \end{aligned}$$

que devido a (2.19) torna-se

$$Cov(N_k, D_l) = n\pi_l^D(1-\pi_k^N).$$

Analogamente,

$$Cov(N_k, W_l) = n\pi_l^W(1-\pi_k^N).$$

Logo,

$$Cov(\mathbf{v_k}, \mathbf{v_l}) = n \begin{pmatrix} \pi_l^N(1-\pi_k^N) & \pi_l^D(1-\pi_k^N) & \pi_l^W(1-\pi_k^N) \\ -\pi_k^D\pi_l^N & -\pi_k^D\pi_l^D & -\pi_k^D\pi_l^W \\ -\pi_k^W\pi_l^N & -\pi_k^W\pi_l^D & -\pi_k^W\pi_l^W \end{pmatrix}$$

$$= n \begin{pmatrix} 1-\pi_k^N \\ -\pi_k^D \\ -\pi_k^W \end{pmatrix} (\pi_l^N, \pi_l^D, \pi_k^W)$$

$$= n \begin{pmatrix} 1-\pi_k^N \\ -\pi_k^D \\ -\pi_k^W \end{pmatrix} \mathbf{p'}_l.$$

Assim sendo,

$$Cov\{n^{\frac{1}{2}}(\tilde{q}_k - q_k^*), n^{\frac{1}{2}}(\tilde{q}_l - q_l^*)\} \cong (\pi_k^N - \frac{1}{2}\pi_k^W)^{-2}(\pi_l^N - \frac{1}{2}\pi_l^W)^{-2}\mathbf{c}_k' \begin{pmatrix} 1-\pi_k^N \\ -\pi_k^D \\ -\pi_k^W \end{pmatrix} \mathbf{p}_l'\mathbf{c}_l.$$

Como $\mathbf{p}_l'\mathbf{c}_l = 0$, os estimadores S.D. das probabilidades condicionais de morte são assintoticamente não correlacionados.

Para completar a prova falta encontrar a variância assintótica de $\mathbf{q}_j$. Utilizando o mesmo esquema anterior,

$$Var\{n^{\frac{1}{2}}(\tilde{\mathbf{q}}_j - \mathbf{q}_j^*)\} \cong nVar\left\{\frac{(\pi_j^N - \frac{1}{2}\pi_j^W)^{-2}}{n}\mathbf{c}_j'(\mathbf{v}_j - n\mathbf{p}_j)\right\}$$

$$= \frac{(\pi_j^N - \frac{1}{2}\pi_j^W)^{-4}}{n}\mathbf{c}_j' Var(\mathbf{v}_j)\mathbf{c}_j.$$

A covariância de N_j e D_j é calculada através da relação (2.18):

$$\begin{aligned} Cov(N_j, D_j) &= Cov(N_{j+1} + D_j + W_j, D_j) \\ &= Cov(N_{j+1}, D_j) + Var(D_j) + Cov(W_j, D_j) \\ &= -n\pi_{j+1}^N\pi_j^D + n\pi_j^D(1-\pi_j^D) - n\pi_j^W\pi_j^D \\ &= n\pi_j^D(1-\pi_{j+1}^N - \pi_j^D - \pi_j^W). \end{aligned}$$

Mas $\sum_{j=1}^{k}(\pi_j^D + \pi_j^W) = 1 - \pi_{k+1}^N \Rightarrow 1 - \pi_k^N = \sum_{j=1}^{k-1}(\pi_j^D + \pi_j^W) = 1 - \pi_{k+1}^N - \pi_k^D - \pi_k^W$. Portanto,

$$Cov(N_j, D_j) = n\pi_j^D(1-\pi_j^N).$$

Analogamente,

$$Cov(N_j, W_j) = n\pi_j^W(1-\pi_j^N).$$

Assim, a variância de $\mathbf{v}_j$ será dada pela matriz

$$n\begin{pmatrix} \pi_j^N(1-\pi_j^N) & \pi_j^D(1-\pi_j^N) & \pi_j^W(1-\pi_j^N) \\ \pi_j^D(1-\pi_j^N) & \pi_j^D(1-\pi_j^D) & -\pi_j^W\pi_j^D \\ \pi_j^W(1-\pi_j^N) & -\pi_j^D\pi_j^W & \pi_j^W(1-\pi_j^W) \end{pmatrix},$$

que pode ser escrita como

$$n\left[\begin{pmatrix} 1-\pi_j^N \\ -\pi_j^D \\ -\pi_j^W \end{pmatrix}\mathbf{p}_j' + \begin{pmatrix} 0 & 0 & 0 \\ \pi_j^D & \pi_j^D & 0 \\ 0 & \pi_j^W & \pi_j^W \end{pmatrix}\right].$$

No cálculo da variância de $n^{\frac{1}{2}}(\tilde{q}_j - q_j^*)$, o primeiro termo da soma acima envolve $\mathbf{p}_j'\mathbf{c}_j$ que é igual a zero. Consequentemente, a variância será dada por

$$\begin{aligned} Var\{n^{\frac{1}{2}}(\tilde{q}_j - q_j^*)\} &\cong (\pi_j^N - \frac{1}{2}\pi_j^W)^{-4}\mathbf{c}_j'\begin{pmatrix} 0 & 0 & 0 \\ \pi_j^D & \pi_j^D & 0 \\ 0 & \pi_j^W & \pi_j^W \end{pmatrix}\mathbf{c}_j \\ &= (\pi_j^N - \frac{1}{2}\pi_j^W)^{-4}\left\{(-\pi_j^D)^2\pi_j^N + (\pi_j^N - \frac{1}{2}\pi_j^W)^2 + (\frac{1}{2}\pi_j^D)^2\pi_j^W\right\} \\ &= -(q_j^*)^2\frac{\pi_j^N}{(\pi_j^N - \frac{1}{2}\pi_j^W)^2} + \frac{q_j^*}{\pi_j^N - \frac{1}{2}\pi_j^W} + (\frac{1}{2}q_j^*)^2\frac{\pi_j^W}{(\pi_j^N - \frac{1}{2}\pi_j^W)^2}. \end{aligned}$$

Escrevendo convenientemente,

$$Var\{n^{\frac{1}{2}}(\tilde{q}_j - q_j^*)\} \cong (\pi_j^N - \frac{1}{2}\pi_j^W)^{-1}\left\{q_j^* - (q_j^*)^2\frac{\pi_j^N - \frac{1}{4}\pi_j^W}{\pi_j^N - \frac{1}{2}\pi_j^W}\right\},$$

O que completa a prova do teorema. □

Como consequência do teorema 1 segue a distribuição assintótica de (2.7):

Corolário 1. *A distribuição assintótica de* $n^{\frac{1}{2}}(\tilde{\mathbf{S}} - \mathbf{S}^*)$ *é normal multivariada com média* $\mathbf{0}$ *e matriz de covariância cujos elementos são dados por*

$$S_k^* S_l^* \sum_{j=1}^{k} Var\{n^{\frac{1}{2}}(\tilde{q}_j - q_j^*)\}/(1-q_j^*)^2, \qquad k \leq l, \tag{2.20}$$

com $S_l^* = \prod_{j=1}^{l}(1-q_j^*)$.

Prova:

Dado que os elementos de $\tilde{\mathbf{S}}$ são funções contínuas dos elementos de $\tilde{\mathbf{q}}$, utiliza-se o método delta (Rao, 1973). Assim sendo, desenvolve-se $\tilde{S}_k - S_k^* = \prod_{j=1}^{k}(1-\tilde{q}_j) - \prod_{j=1}^{k}(1-q_j^*)$ em torno do ponto $(q_1^*, q_2^*, \ldots, q_k^*)$. O vetor de derivadas terá como elementos $\partial(\tilde{S}_k - S_k^*)/\partial(1-\tilde{q}_i) = \prod_{j=1}^{k}(1-\tilde{q}_j)/(1-\tilde{q}_i), i = 1, \ldots, k$.
Então,

$$\begin{aligned}
\tilde{S}_k - S_k^* &= \prod_{j=1}^{k}(1-q_j^*)\left(\frac{1}{1-q_1^*}, \ldots, \frac{1}{1-q_k^*}\right)\begin{pmatrix}\tilde{q}_1 - q_1^* \\ \vdots \\ \tilde{q}_k - q_k^*\end{pmatrix} + O_p(n^{-1}) \\
&= -\prod_{j=1}^{k}(1-q_j^*)\sum_{i=1}^{k}\frac{\tilde{q}_i - q_i^*}{1-q_i^*} + O_p(n^{-1}) \\
&= -S_k^*\sum_{i=1}^{k}\frac{\tilde{q}_i - q_i^*}{1-q_i^*} + O_p(n^{-1}).
\end{aligned}$$

Portanto, para $k \leq l$,

$$\begin{aligned}
Cov\{n^{\frac{1}{2}}(\tilde{S}_k - S_k^*), n^{\frac{1}{2}}(\tilde{S}_l - S_l^*)\} &\cong Cov\left\{-S_k^* n^{\frac{1}{2}}\sum_{i=1}^{k}\frac{\tilde{q}_i - q_i^*}{1-q_i^*}, -S_l^* n^{\frac{1}{2}}\sum_{j=1}^{l}\frac{\tilde{q}_j - q_j^*}{1-q_j^*}\right\} \\
&= S_k^* S_l^* \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{Cov\{n^{\frac{1}{2}}(\tilde{q}_i - q_i^*), n^{\frac{1}{2}}(\tilde{q}_j - q_j^*)\}}{(1-q_i^*)(1-q_j^*)}.
\end{aligned}$$

A expressão acima é igual a zero para $i \neq j$. Portanto, assintoticamente,

$$Cov\{n^{\frac{1}{2}}(\tilde{S}_k - S_k^*), n^{\frac{1}{2}}(\tilde{S}_l - S_l^*)\} \cong S_k^* S_l^* \sum_{j=1}^{k}\frac{Var\{n^{\frac{1}{2}}(\tilde{q}_j - q_j^*)\}}{(1-q_j^*)^2},$$

conforme a expressão (2.20). □

Classicamente, a estimação da variância de $\tilde{q}_j$ é feita considerando-se que o número de mortes no intervalo I_j segue uma distribuição binomial onde o número de indivíduos no início do intervalo é dado por N_j'. Portanto uma estimativa para a variância será dada por

$$\widehat{\mathrm{Var}}(\tilde{q}_j) = \frac{\tilde{q}_j - (\tilde{q}_j)^2}{N_j'}. \tag{2.21}$$

Utilizando (2.20) e (2.21) decorre que

$$\widehat{\mathrm{Var}}(\tilde{S}_j) = \frac{1}{n}(\tilde{S}_j)^2 \sum_{i=1}^{j} \frac{\tilde{q}_j(1-\tilde{q}_j)}{N'_j(1-\tilde{q}_j)^2}$$

$$= \frac{1}{n}(\tilde{S}_j)^2 \sum_{i=1}^{j} \frac{D_j}{N'_j(N'_j - D_j)}. \tag{2.22}$$

A expressão (2.22) é a aproximação clássica estabelecida por Greenwood(1926).

A medida que o número de intervalos cresce(e portanto as amplitudes diminuem) as quantidades $\tilde{q}_j$ e q_j^* tendem a ser cada vez menores, de modo que os termos quadráticos em (2.12) e (2.21) tornam-se desprezíveis. Isto implica que (2.21) será um estimador consistente de (2.12) quando $\sup_{1\leq j\leq k+1} |\xi_{j-1} - \xi_j| \longrightarrow 0$.

2.3 Estimador Produto–Limite (P.L.)

Proposto inicialmente por Bölmer(1912), o estimador produto-limite teve suas propriedades estudadas muitos anos depois por Kaplan e Meier(1958) sendo por isso conhecido também como estimador de Kaplan-Meier.

O estimador P.L. pode ser visto como o caso limite dos estimadores atuariais quando o número de partições no intervalo $[0, T]$ tende a infinito. Em particular, os estimadores R.S. e S.D. podem ser encarados como aproximações para o estimador P.L. . Neste contexto utiliza-se a notação $\tilde{S}_k^K$ para representar o estimador S.D. no intervalo I_k, calculado sob uma partição composta por K sub-intevalos $(\xi_{0,K}, \xi_{1,K}], \ldots, (\xi_{K,K}, \xi_{K+1,K}]$, de modo que o estimador produto-limite pode ser definido como

$$\hat{S}(t) = \lim_{K\to+\infty} \tilde{S}_k^K, \qquad t \in [\xi_{k,K}, \xi_{k+1,K}), \tag{2.23}$$

dado que $\sup_{0\leq k\leq K} |\xi_{k,K} - \xi_{k+1,K}| \overset{K\uparrow+\infty}{\longrightarrow} 0$.

A notação acima é proposta por Breslow e Crowley(1974) que através de argumentos heurísticos deduzem a distribuição assintótica de $\hat{S}(t)$.

Ao ser tomado o limite em (2.23) torna-se necessário reafirmar a convenção adotada anteriormente a fim de evitar imprecisões na determinação de N_k, D_k e W_k. Portanto novamente será assumido que instantes de censura coincidentes com ξ_j são deslocados um infinitésimo de tempo além de ξ_j. Consequentemente em caso de empate entre censura e morte, assume-se que a última ocorre antes e censuras ocorrendo no instante ξ_j não

contribuirão para W_j. Assim sendo, quando $K \longrightarrow +\infty$, os intervalos que permaneciam fixados na seção anterior terão agora comprimentos variáveis convergindo para valores $\tau_1 < \tau_2 < \ldots < \tau_q, \quad q \leq n$.

Redefinindo $N_k = \sum_{j=1}^{n} I_{\{t_j \geq \tau_k\}}$ e $D_k = \sum_{j=1}^{n} I_{\{t_j = \tau_k, \delta_j = 1\}}$ segue agora que a probabilidade condicional de morte em I_i será estimada por

$$\hat{q}_i = \frac{D_i}{N_i}.$$

Supondo $D_i \leq 1, \quad N_j = n - j + 1$ e

$$1 - \hat{q}_j = \begin{cases} 1 - \frac{1}{N_j}, & \text{se } \delta_j = 1; \\ 1, & \text{se } \delta_j = 0. \end{cases}$$

Portanto, a função de sobrevivência será estimada por

$$\begin{aligned} \hat{S}(t) &= \prod_{i:t_i<t} (1 - \hat{q}_i)^{\delta_i} = \prod_{i:t_i<t} \left(1 - \frac{1}{N_i}\right)^{\delta_i} \\ &= \prod_{i:t_i<t} \left(\frac{N_i - 1}{N_i}\right)^{\delta_i} = \prod_{i:t_i<t} \left(\frac{n-i}{n-i+1}\right)^{\delta_i}. \end{aligned} \tag{2.24}$$

Quando $D_j \geq 1$ para alguns instantes, (2.24) pode ser reescrita considerando-se que as D_j mortes que ocorrem no mesmo instante podem ser separadas por um refinamento maior na partição em $[0, T]$, de modo que a probabilidade condicional de sobrevivência será dada por

$$\frac{N_j - 1}{N_j} \times \frac{N_j - 2}{N_j - 1} \times \cdots \times \frac{N_j - D_j}{N_j - D_j + 1} = \frac{N_j - D_j}{N_j}.$$

Portanto, no caso de empates, (2.24) pode ser escrita como

$$\hat{S}(t) = \prod_{j:t_j<t} \left(\frac{N_j - D_j}{N_j}\right)^{\delta_j}. \tag{2.25}$$

O estimador P.L. também pode ser visto como o estimador de máxima verossimilhança para a função de sobrevivência se for considerado o enfoque dado por Kiefer e Wolfowitz(1956). Para tanto, suponha que $T_1, T_2, \ldots, T_K$ representam instantes de morte com d_i mortes no instante T_i; para cada intervalo $[T_i, T_{i+1})$ ocorrem c_i censuras em instantes $L_j^{c_i}, \quad j = 1, \ldots, c_i$ e $i = 0, \ldots, K$ com $T_0 = 0$ e $T_{K+1} = +\infty$. Suponha também

que $S(T_i - 0)$ exclue as mortes que ocorrem no instante T_i ao passo que $S(T_i)$ as inclui. Com a notação definida, a verossimilhança pode ser escrita como

$$\begin{aligned} L &= \prod_{j=1}^{c_0} S(L_j^{c_0})[S(T_1 - 0) - S(T_1)]^{d_1} \times \prod_{j=1}^{c_1} S(L_j^{c_1})[S(T_2 - 0) - S(T_2)]^{d_2} \times \ldots \\ &\quad \times \prod_{j=1}^{c_{K-1}} S(L_j^{c_{K-1}})[S(T_K - 0) - S(T_K)]^{d_K} \times \prod_{j=1}^{c_K} S(L_j^{c_K}) \\ &= \prod_{j=1}^{c_0} S(L_j^{c_0}) \prod_{i=1}^{K}[S(T_i - 0) - S(T_i)]^{d_i} \prod_{j=1}^{c_i} S(L_j^{c_i}). \end{aligned} \tag{2.26}$$

Como $S(\cdot)$ é uma função monótona não crescente e $T_i \leq L_j^{c_i} < T_{i+1}$, segue que $S(T_i) \geq S(L_j^{c_i}) \geq S(T_{i+1})$ de modo que para maximizar (2.26) ambos os termos $S(L_j^{c_i})$ e $S(T_i - 0)$, devem assumir o maior valor possível e $S(T_i)$ o menor valor possível, ou seja, $S(L_j^{c_0}) = S(T_i) = S(T_{i+1} - 0) = P_i$. Portanto, (2.26) fica

$$L = \prod_{i=1}^{K}[P_{i-1} - P_i]^{d_i} \prod_{j=1}^{c_i} P_i = \prod_{i=1}^{K}[P_{i-1} - P_i]^{d_i} P_i^{c_i}.$$

Escrevendo em termos das probabilidades condicionais $p_i = 1 - q_i$, tem-se que $P_i = p_i p_{i-1} p_{i-2} \cdots p_1$. Logo,

$$\begin{aligned} L &= \prod_{i=1}^{K}(p_{i-1}p_{i-2}\cdots p_1 - p_i \cdots p_1)^{d_i}(p_i p_{i-1} \cdots p_1)^{c_i} \\ &= \prod_{i=1}^{K}[(1 - p_i)p_{i-1} \cdots p_1)]^{d_i}(p_i p_{i-1} \cdots p_1)^{c_i} \\ &= \prod_{i=1}^{K} q_i^{d_i} p_i^{c_i} (p_{i-1} \cdots p_1)^{(c_i + d_i)} \\ &= q_1^{d_1} p_1^{c_1} \times q_2^{d_2} p_2^{c_2} p_1^{(c_2+d_2)} \times \ldots \times q_K^{d_K} p_K^{c_K} (p_{K-1} \cdots p_1)^{(c_K + d_K)} \\ &= q_1^{d_1} \ldots q_K^{d_K} p_1^{\sum_{j=1}^{K}(c_j+d_j)-d_1} \times p_2^{\sum_{j=2}^{K}(c_j+d_j)-d_2} \times \ldots \times p_K^{\sum_{j=K}^{K}(c_j+d_j)-d_K}. \end{aligned}$$

Fazendo $N_i = \sum_{j=i}^{K}(c_j + d_j)$ vem

$$L = \prod_{i=1}^{K} q_i^{d_i} p_i^{N_i - d_i} \Rightarrow \log L = \sum_{i=1}^{K} d_i \log q_i + (N_i - d_i) \log p_i$$

Derivando-se em relação a p_i e igualando a zero, segue que

$$\hat{p}_i = \frac{N_i - d_i}{N_i}.$$

O produto dos valores $\hat{p}_i$ fornece o estimador de máxima verossimilhança cuja expressão coincide com (2.25). Nota-se que $N_i = \sum_{j=i}^{K}(c_j + d_j)$ é a soma de todas as perdas e mortes excluindo-se aquelas dos intervalos anteriores a $[T_i, T_{i+1})$ ou seja, N_i representa o conjunto dos indivíduos em risco e não censurados no instante T_i.

Para estudar algumas propriedades assintóticas podem ser utilizados os resultados obtidos nas seções 2.1 e 2.2, contanto que a troca dos limites para $K \to +\infty$ e $n \to +\infty$ possa ser efetuada. Neste caso o estimador P.L. é o limite do estimador R.S. quando $K \to +\infty$ e uma vez que este é consistente, temos que $\hat{S}(t)$ será um estimador consistente para $S(t)$.

Uma prova formal da consistência de $\hat{S}(t)$ é dada por Peterson(1977) considerando censuras aleatórias com distribuição $H(\cdot)$ conforme o teorema:

Teorema 2. *O estimador P.L. é fortemente consistente para* $S(t)$.

Prova: Seja S^M definido por $S^M(t) = P\{\min(T_i, L_i) > t\} = [1 - F^0(t)][1 - H(t)]$. Considere as *funções de sub-sobrevivência* definidas por

$$S^M_{NC}(t) = P\{\min(T_i, L_i) > t, \delta_i = 1\} = P\{T_i > t; T_i < L_i\} = \int_t^{+\infty} [1 - H(u)]dF^0(u)$$

e

$$S^M_C(t) = P\{\min(T_i, L_i) > t, \delta_i = 0\} = P\{L_i > t; L_i < T_i\} = \int_t^{+\infty} [1 - F^0(u)]dH(u).$$

Decorre daí que $S^M(t) = S^M_{NC}(t) + S^M_C(t)$uma vez que δ_i define duas classes de eventos mutuamentes exclusivos. O passo seguinte é mostrar que $S(t) = 1 - F^0(t)$ pode ser escrito como função de $S^M_{NC}(t)$ e de $S^M_C(t)$. As funções F^0 e H não precisam ser contínuas, porém supõe-se que não têm pontos de descontinuidade em comum.

Supondo que $S(t) = 1 - F^0(t)$ é uma função mista (contínua e discreta), para intervalos onde F^0 é contínua a *função de risco* é definida por $\lambda_c^{F^0}(s) = -S'(s)/S(s)$ e para os pontos de descontinuidade, $\lambda_d^{F_0}(s) = (F^0(s^+) - F^0(s^-))/(1 - F^0(s-))$ onde $F^0(s^-) = \lim_{x \to s^-} F^0(x)$ e $F^0(s^+) = \lim_{x \to s^+} F^0(x)$. Conforme Cox(1972), a função de risco acumulada pode ser definida como

$$\Lambda_{F^0}(t) = c \int_0^t \lambda_c^{F^0}(s)ds + \sum_{s \in \mathcal{A}_t} -\log[1 - \lambda_d^{F^0}(s)], \qquad (2.27)$$

onde $c\int_0^t$ indica integração sobre os intervalos de pontos menores que t para os quais S_{NC}^M é contínua e $\mathcal{A}_t$ é o conjunto dos pontos de descontinuidade de F^0 tais que $s \leq t$.

Considerando o primeiro termo da soma em (2.27),

$$c\int_0^t \lambda_c^{F^0}(s)ds = c\int_0^t -\frac{dS(s)}{S(s)} = c\int_0^t \frac{dF^0(s)}{1-F^0(s)} = c\int_0^t \frac{[1-H(s)]dF^0(s)}{[1-H(s)][1-F^0(s)]}$$
$$= c\int_0^t -\frac{dS_{NC}^M(s)}{S_{NC}^M(s)+S_C^M(s)}.$$

Para o segundo termo da soma (2.27), é utilizada a suposição feita a respeito da continuidade de F^0 e H. Portanto, como a soma é efetuada sob pontos de descontinuidade de F^0, $H(s^-) = H(s^+)$, segue que

$$\sum_{s\in\mathcal{A}_t} -\log[1-\lambda_d^{F^0}(s)] = \sum_{s\in\mathcal{A}_t} -\log\left[\frac{1-F^0(s^+)}{1-F^0(s^-)}\right]$$
$$= \sum_{s\in\mathcal{A}_t} -\log\left\{\frac{[1-F^0(s^+)][1-H(s^+)]}{[1-F^0(s^-)][1-H(s^-)]}\right\}$$
$$= \sum_{s\in\mathcal{A}_t} -\log\left[\frac{S_{NC}^M(s^+)+S_C^M(s^+)}{S_{NC}^M(s^-)+S_C^M(s^-)}\right].$$

Como a relação entre $S(t)$ e $\Lambda_{F^0}(t)$ é dada por $S(t) = e^{-\Lambda_{F^0}(t)}$ tem-se que

$$S(t) = \exp\left\{c\int_0^t -\frac{dS_{NC}^M(s)}{S_{NC}^M(s)+S_C^M(s)} + \sum_{s\in\mathcal{A}_t} -\log\left[\frac{S_{NC}^M(s^+)+S_C^M(s^+)}{S_{NC}^M(s^-)+S_C^M(s^-)}\right]\right\}. \qquad (2.28)$$

A expressão (2.28) é conhecida como *representação de Peterson* e mostra que $S(t)$ pode ser representada como função de S_{NC}^M, S_C^M e t, isto é,

$$S(t) = \Psi(S_{NC}^M, S_C^M; t).$$

Definindo as *funções de sub-sobrevivência empíricas*

$$\hat{S}_{NC}(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{t_i>t;\delta_i=1\}}$$

e

$$\hat{S}_C(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{t_i>t;\delta_i=0\}},$$

Peterson(1977) prova que são estimadores de máxima verossimilhança de modo que o estimador P.L., sendo também um estimador de máxima verossimilhança, pode ser escrito como

$$\hat{S}(t) = \Psi(\hat{S}_{NC}, \hat{S}_C; t),$$

sendo que agora Ψ envolve apenas a soma sob os saltos de descontinuidade de $\hat{S}$.

Utilizando o teorema de Glivenko-Cantelli (Billingsley, 1968), segue que $\hat{S}_{NC}(t) \xrightarrow{q.c.} S^M_{NC}(t)$ e $\hat{S}_C(t) \xrightarrow{q.c.} S^M_C(t)$ uniformemente em t.

A função $\Psi(G, J; t)$ é uma função contínua dos argumentos $G(\cdot)$ e $J(\cdot)$ quando a métrica no espaço de descontinuidades é tal que se

$$\|\hat{S}_{NC} - G\| = \sup |\hat{S}_{NC}(t) - G(t)| \longrightarrow 0$$

e

$$\|\hat{S}_C - J\| \longrightarrow 0,$$

então,

$$\Psi(\hat{S}_{NC}, \hat{S}_C; t) \xrightarrow{q.c.} \Psi(G, J; t).$$

Fazendo $G = S^M_{NC}$ e $J = S^M_C$ segue que

$$\hat{S}(t) = \Psi(\hat{S}_{NC}, \hat{S}_C; t) \xrightarrow{q.c.} \Psi(S^M_{NC}, S^M_C; t) = S(t).$$

Portanto, o estimador P.L. é fortemente consistente. □

Uma vez que o estimador Produto-Limite pode ser visto como caso limite do estimador S.D. cujas propriedades assintóticas foram vistas na seção anterior, podem ser determinados alguns resultados assintóticos supondo válida a troca de limites entre k e n. Inicialmente, considere o processo estocástico $Z_n(t) = n^{\frac{1}{2}}(\hat{S}(t) - S(t))$. Breslow e Crowley(1974) através de argumentos heurísticos concluem que o processo $Z_n(t)$ converge fracamente para um processo Gaussiano $Z(t)$ com média zero e covariância dada por

$$Cov(Z(s), Z(t)) = S(t)S(s) \int_0^{t \wedge s} \frac{dF^0(x)}{[1 - F(x)][1 - F^0(x)]}, \tag{2.29}$$

onde $1 - F(\cdot) = [1 - F(\cdot)][1 - H^0(\cdot)]$, e $t \wedge s$ indica o mínimo entre t e s.

O resultado acima pode ser formalizado em alguns casos particulares. Supondo que a função de distribuição dos tempos de vida é discreta, tem-se o seguinte teorema:

Teorema 3. *Se $S(t)$ é uma função discreta assumindo valores nos instantes $\tau_1 < \tau_2 < \ldots < \tau_q$, então o vetor normalizado $n^{\frac{1}{2}}(\hat{\mathbf{S}} - \mathbf{S})$ terá distribuição assintótica normal multivariada com vetor de médias* $\mathbf{0}$ *e matriz de covariância cujos elementos serão dados por*

$$S_k S_l \sum_{j=1}^{k \wedge l} \frac{q_j}{\pi_j^N p_j},$$

onde $\pi_j^N = 1 - F(\tau_j - 0)$; $q_j = 1 - p_j = \frac{F^0(\tau_j) - F^0(\tau_j - 0)}{1 - F^0(\tau_j - 0)}$.

Prova: Foi visto que $\hat{q}_j = D_j/N_j$. A quantidade D_j é o número de indivíduos que morrem no instante τ_j e não são censurados, isto é, D_j tem uma distribuição binomial com probabilidade dada por $\pi_j^D = P\{t_i = \tau_j; \delta_i = 1\} = P\{T_i = \tau_j; T_i \leq L_i\} = [F^0(\tau_j) - F^0(\tau_j - 0)][1 - H(\tau_j - 0)]$. Analogamente, N_k, o número de indivíduos que estão em risco em τ_k, tem distribuição binomial com probabilidade dada por $\pi^{N_j} = P\{t_i \geq \tau_j\} = [1 - F^0(\tau_j - 0)][1 - H(\tau_j - 0)]$. Portanto, pela Lei Forte dos Grandes Números, $D_j/n \xrightarrow{q.c.} [F^0(\tau_j) - F^0(\tau_j - 0)][1 - H(\tau_j - 0)]$ e $N_j/n \xrightarrow{q.c.} [1 - F^0(\tau_j - 0)][1 - H(\tau_j - 0)]$. Assim, $\hat{q}_j \xrightarrow{q.c.} [F^0(\tau_j) - F^0(\tau_j - 0)]/[1 - F^0(\tau_j - 0)] = q_j$. A distribuição assintótica do vetor $n^{\frac{1}{2}}(\hat{\mathbf{q}} - \mathbf{q})$ pode então ser obtida, a exemplo do que foi feito na seção anterior. Expandindo $\hat{\mathbf{q}} - \mathbf{q}$ em série de Taylor ao redor de $(n\pi_j^N, n\pi_j^D)$.

$$\begin{aligned}
\hat{q}_j - q_j &= \frac{1}{n}\left(-\frac{\pi_j^D}{(\pi_j^N)^2}, \frac{1}{\pi_j^N}\right)\begin{pmatrix} N_j - n\pi_j^N \\ D_j - n\pi_j^D \end{pmatrix} + O_p(n^{-1}) \\
&= \frac{1}{(\pi_j^N)^2}\frac{1}{n}\left(-\pi_j^D, \pi_j^N\right)\left[\begin{pmatrix} N_j \\ D_j \end{pmatrix} - n\begin{pmatrix} \pi_j^N \\ \pi_j^D \end{pmatrix}\right] + O_p(n^{-1}).
\end{aligned}$$

Procedendo com os cálculos de forma análoga à seção anterior, temos que

$$\begin{aligned}
Cov\{n^{\frac{1}{2}}(\hat{q}_j - q_j); n^{\frac{1}{2}}(\hat{q}_l - q_l)\} &\cong \\
&\cong (\pi_j^N \pi_l^N)^{-2}\left(\pi_j^N, -\pi_j^D\right)\begin{pmatrix} 1 - \pi_j^N \\ -\pi_j^D \end{pmatrix}\left(\pi_l^N, \pi_l^D\right)\begin{pmatrix} \pi_l^N \\ \pi_k^N \end{pmatrix} \\
&= 0.
\end{aligned}$$

Portanto os estimadores $\hat{q}_j$, $j = 1, \ldots, K$ são assintoticamente não correlacionados.

Para obter a variância tem-se que

$$
\begin{aligned}
Var\{n^{\frac{1}{2}}(\hat{q}_j - q_j)\} &\cong \frac{(\pi_j^N)^{-4}}{n}\,(\pi_j^D, \pi_j^N)\, Var\left\{\begin{pmatrix} N_j \\ D_j \end{pmatrix}\right\}\begin{pmatrix} -\pi_j^D \\ \pi_j^N \end{pmatrix} \\
&\cong \frac{\pi_j^N}{n}\,(-\pi_j^D, \pi_j^N)\, n \begin{pmatrix} \pi_j^N(1-\pi_j^N) & \pi_j^D(1-\pi_j^N) \\ \pi_j^D(1-\pi_j^N) & \pi_j^D(1-\pi_j^D) \end{pmatrix}\begin{pmatrix} -\pi_j^D \\ \pi_j^N \end{pmatrix} \\
&= (\pi_j^N)^{-4}\,(-\pi_j^D, \pi_j^N)\left[\begin{pmatrix} 1-\pi_j^N \\ -\pi_j^D \end{pmatrix}(\pi_j^N, \pi_j^D) + \begin{pmatrix} 0 & 0 \\ \pi_j^D & \pi_j^D \end{pmatrix}\right]\begin{pmatrix} -\pi_j^D \\ \pi_j^N \end{pmatrix} \\
&= (\pi_j^N)^{-4}\,(-\pi_j^D, \pi_j^N)\begin{pmatrix} 0 & 0 \\ \pi_j^D & \pi_j^D \end{pmatrix}\begin{pmatrix} -\pi_j^D \\ \pi_j^N \end{pmatrix} \\
&= (\pi_j^N)^{-4}\left\{-\pi_j^N(\pi_j^D)^2 + (\pi_j^N)^2\pi_j^D\right\} = \frac{\pi_j^D}{\pi_j^N}\left(\frac{\pi_j^N}{(\pi_j^N)^2} - \frac{\pi_j^D}{(\pi_j^N)^2}\right) \\
&= p_j\frac{1-p_j}{\pi_j^N} = \frac{p_j q_j}{\pi_j^N}.
\end{aligned}
$$

Finalmente, conforme visto na prova do corolário 1 da seção anterior,

$$
\begin{aligned}
Cov\{n^{\frac{1}{2}}(\hat{S}_k - S_k); n^{\frac{1}{2}}(\hat{S}_l - S_l)\} &\cong S_k S_l \sum_{j=1}^{k\wedge l} \frac{Var\{n^{\frac{1}{2}}(\hat{q}_j - q_j)\}}{(1-q_j)^2} \\
&= S_k S_l \sum_{j=1}^{k\wedge l} \frac{p_j q_j}{p_j^2 \pi_j^N} = S_k S_l \sum_{j=1}^{k\wedge l} \frac{q_j}{p_j \pi_j^N}.
\end{aligned}
$$

□

Outro caso de interesse é quando as distribuições $F^0(\cdot)$ e $H(\cdot)$ são contínuas. Para determinar a distribuição assintótica do estimador P.L. utiliza-se a *função de risco acumulada empírica*, definida por

$$
\Lambda^{\epsilon}(t) = \sum_{i:t_i \le t} \frac{\delta_i}{n-i+1}. \tag{2.30}
$$

A função de risco acumulada empírica é uma estimativa da função de risco acumulada, que na notação utilizada na prova do teorema 2, pode ser escrita como

$$
\Lambda(t) = \int_0^t \lambda_c^{F^0}(u)du = -\int_0^t \frac{S'(u)}{S(u)} = \int_0^t \frac{dF^0(u)}{1-F^0(u)}. \tag{2.31}
$$

Definindo a função de subdistribuição para dados não censurados,

$$
F_{NC}(t) = P\{\min(T_i, L_i) \le t, \delta_i = 1\} = \int_0^t [1-H(u)]dF^0(u),
$$

tem-se que $dF^0(t) = [1 - H(t)]^{-1} dF_{NC}(t)$, de modo que a expressão (2.31) fica

$$\Lambda(t) = \int_0^t \frac{dF_{NC}(u)}{[1 - H(u)][1 - F^0(u)]} = \int_0^t \frac{dF_{NC}(u)}{1 - F(u)}. \tag{2.32}$$

Seja a função de distribuição empírica $F^\epsilon(t) = \frac{1}{n} \sum I_{\{t_i \leq t\}}$ e a função de subdistribuição empírica restrita aos dados não censurados, $F^\epsilon_{NC} = \frac{1}{n} \sum I_{\{t_i \leq t; \delta_i = 1\}}$. Então com base em (2.32), a função de risco acumulado empírica pode ser escrita como

$$\Lambda^\epsilon(t) = \int_0^t \frac{dF^\epsilon_{NC}(u)}{1 - F^\epsilon(u)}.$$

Pela Lei Forte dos Grandes Números segue que $F^\epsilon(t)$ e $F^\epsilon_{NC}(t)$ convergem quase certamente para $F(t)$ e $F_{NC}(t)$ respectivamente, de modo que $\Lambda^\epsilon(t) \xrightarrow{q.c.} \Lambda(t)$.

A fim de determinar a distribuição assintótica de $\Lambda^\epsilon(t)$, considere o desenvolvimento em série de Taylor

$$\frac{1}{1-x} = \frac{1}{1-x_0} + \frac{x - x_0}{(1-x_0)^2} + \frac{(x-x_0)^2}{(1-x_0)^3} + \ldots,$$

de modo que

$$\frac{1}{1-x} = \frac{1}{1-x_0} + \frac{x - x_0}{(1-x_0)} \left(\frac{1}{1-x} \right),$$

e

$$\frac{1}{1-x} = \frac{1}{1-x_0} + \frac{x - x_0}{(1-x_0)^2} + \frac{(x-x_0)^2}{(1-x_0)^2} \left(\frac{1}{1-x} \right).$$

Substituindo $F^\epsilon(\cdot)$ e $F(\cdot)$ no lugar de x e x_0 vem

$$\begin{aligned}
\Lambda^\epsilon(t) &= \int_0^t \frac{dF^\epsilon_{NC}(u)}{1 - F^\epsilon(u)} = \int_0^t \frac{dF_{NC}(u)}{1 - F^\epsilon(u)} + \int_0^t \frac{d(F^\epsilon_{NC}(u) - F_{NC}(u))}{1 - F^\epsilon(u)} \\
&= \int_0^t \left[\frac{1}{1 - F(u)} + \frac{F^\epsilon(u) - F(u)}{(1 - F(u))^2} + \frac{(F^\epsilon(u) - F(u))^2}{(1 - F(u))^2} \left(\frac{1}{1 - F^\epsilon(u)} \right) \right] dF_{NC}(u) \\
&+ \int_0^t \left[\frac{1}{1 - F(u)} + \frac{F^\epsilon(u) - F(u)}{1 - F(u)} \left(\frac{1}{1 - F^\epsilon(u)} \right) \right] d(F^\epsilon_{NC} - F_{NC}) = \\
&= \Lambda(t) + \int_0^t \frac{F^\epsilon(u) - F(u)}{(1 - F(u))^2} dF_{NC}(u) + \frac{F^\epsilon_{NC}(t) - F_{NC}(t)}{1 - F(t)} \\
&- \int_0^t \frac{F^\epsilon_{NC}(u) - F_{NC}}{(1 - F(u))^2} dF(u) + \int_0^t \frac{(F^\epsilon(u) - F(u))^2}{(1 - F^\epsilon(u))^2} \left(\frac{1}{1 - F^\epsilon(u)} \right) dF_{NC}(u) \\
&+ \int_0^t \frac{F^\epsilon(u) - F(u)}{1 - F(u)} \left(\frac{1}{1 - F^\epsilon} \right) d(F^\epsilon_{NC}(u) - F_{NC}(u)).
\end{aligned} \tag{2.33}$$

Utilizando o espaço de funções descontínuas $D[0,T]$ e uma métrica conveniente, dada pela topologia de Skorohod (Billingsley, 1968), Breslow & Crowley(1974) provam que os dois últimos termos de (2.33) convergem para zero em probabilidade e que $X^e = n^{\frac{1}{2}}(F^e - F)$ e $Y^e = n^{\frac{1}{2}}(F^e_{NC} - F_{NC})$ convergem fracamente para um processo gaussiano bivariado (X,Y) com média zero e covariâncias dadas por

$$\begin{aligned} Cov\{X(s), X(t)\} &= F(s)[1 - F(t)], \\ Cov\{Y(s), Y(t)\} &= F_{NC}(s)[1 - F_{NC}(t)], \\ Cov\{Y(s), X(t)\} &= F_{NC}(s)[1 - F(t)], \\ Cov\{X(s), Y(t)\} &= F_{NC}(s) - F(s)F_{NC}(t), \end{aligned}$$

para $s \leq t$. Consequentemente, segue que

$$n^{\frac{1}{2}}(\Lambda^e(t) - \Lambda(t)) = \int_0^t n^{\frac{1}{2}} \frac{X^e(u)}{(1 - F(u))^2} dF_{NC}(u) + n^{\frac{1}{2}} \frac{Y^e(t)}{1 - F(t)} - \int_0^t \frac{Y^e(u)}{(1 - F(u))^2} dF(u)$$

converge fracamente para

$$Z(t) = \int_0^t \frac{X(u)}{(1 - F(u))^2} dF_{NC}(u) + \frac{Y(t)}{1 - F(t)} + \int_0^t \frac{Y(u)}{(1 - F(u))^2} dF(u).$$

O processo $Z(t)$ é um processo gaussiano, pois é uma combinação linear de processos gaussianos. Sua média é zero e a covariância assintótica é dada por

$$\begin{aligned} Cov\{Z(s), Z(t)\} &= \int_0^t (1 - F(u))^{-1}(1 - F^0(u))^{-1} dF^0(u) \\ &= \int_0^s (1 - F(u))^{-2} dF_{NC}(u). \end{aligned} \tag{2.34}$$

A partir dos resultados acima, segue a distribuição assintótica do estimador P.L.:

Teorema 4. *Supondo que $T < \infty$ satisfaz $F(T) < 1$ e que $F^0(\cdot)$ e $H(\cdot)$ são contínuas, a função aleatória $n^{\frac{1}{2}}(\hat{S}(t) - S(t))$, para $0 < t < T$ converge fracamente para um processo gaussiano $Z^0(t)$ com média 0 e covariância dada por*

$$Cov\{Z^0(s), Z^0(t)\} = S(s)S(t) \int_0^t \frac{dF^0(u)}{[1 - H(u)][S(u)]^2}, \qquad s \leq t.$$

Prova:

Utilizando a relação $S(t) = e^{-\Lambda(t)}$,

$$\begin{aligned} n^{\frac{1}{2}}(\hat{S}(t) - S(t)) &= n^{\frac{1}{2}}(\hat{S}(t) - e^{-\Lambda(t)} + e^{-\Lambda^e(t)} - e^{-\Lambda^e(t)}) \\ &= n^{\frac{1}{2}}(e^{-\Lambda^e(t)} - e^{-\Lambda(t)}) + n^{\frac{1}{2}}(\exp\{\log \hat{S}(t)\} - e^{-\Lambda^e(t)}). \end{aligned}$$

Aplicando-se o método delta, a distribuição assintótica para o estimador P.L. pode ser encontrada. Considere os seguintes desenvolvimentos em série de Taylor:

$$e^{-x} = e^{-x_0} - e^{-x_0^*}(x - x_0), \qquad \text{com } x_0^* \text{ entre } x \text{ e } x_0$$

e

$$e^{-y} = e^{-y_0} - e^{-y_0}(y - y_0) + e^{-y_0^{**}}\frac{(y - y_0)^2}{2}, \qquad \text{com } y_0^{**} \text{ entre } y \text{ e } y_0.$$

Fazendo $x = -\log \hat{S}(t), x_0 = \Lambda^e(t)$ e $y = \Lambda^e(t), y_0 = \Lambda(t)$ vem que

$$\begin{aligned} n^{\frac{1}{2}}(\hat{S}(t) - S(t)) = -n^{\frac{1}{2}}e^{-\Lambda(t)}(\Lambda^e(t) - \Lambda(t)) + \frac{e^{-\Lambda^{**}(t)}}{2}n^{\frac{1}{2}}(\Lambda^e(t) - \Lambda(t))^2 \\ - e^{-\Lambda^*(t)}n^{\frac{1}{2}}(-\log\{\hat{S}(t)\} - e^{-\Lambda^e(t)}). \end{aligned} \tag{2.35}$$

onde $\Lambda^*(t)$ situa-se entre $-\log\{\hat{S}(t)\}$ e $\Lambda^e(t)$; e $\Lambda^{**}(t)$ situa-se entre $\Lambda^e(t)$ e $\Lambda(t)$, segundo a métrica de Skorohod.

Pode ser mostrado (Breslow & Crowley, 1974) que os dois últimos termos de (2.35) convergem para zero em probabilidade, enquanto que o primeiro termo converge fracamente para $Z^0(t) = -S(t)Z(t)$. Logo, como $F(T) < 1$, $Z^0(t)$ é um processo gaussiano com média zero cuja covariância, para $s \leq t$, será dada por

$$\begin{aligned} &Cov\{Z^0(s), Z^0(t)\} \cong \\ &\cong S(s)S(t)Cov\{Z(s), Z(t)\} = S(s)S(t)\int_0^s [1 - F(u)]^{-1}[1 - F^0(u)]^{-1}dF^0(u) \\ &= S(s)S(t)\int_0^s \frac{dF^0(u)}{[1 - H(u)][1 - F^0(u)]^2} = S(s)S(t)\int_0^s \frac{dF^0(u)}{[1 - H(u)][S(u)]^2}. \end{aligned}$$

o que completa a prova do teorema. □

Segue do teorema 4 que a variância do estimador P.L. pode ser aproximada por

$$\begin{aligned} Var\{\hat{S}(t)\} &\cong \frac{1}{n}[S(t)]^2\int_0^t \frac{dF^0(u)}{[1 - H(u)][S(u)]^2} \\ &= \frac{1}{n}[S(t)]^2\int_0^t \frac{dF_{NC}(u)}{[1 - F(u)]^2}. \end{aligned} \tag{2.36}$$

Para calcular uma estimativa da variância, substitui-se no denominador da integral acima a expressão $[1 - F(u)][1 - F(u - 0)]$. Assim supondo que não hajam empates,

$dF_{NC}(t_i), 1 - F(t_i)$ e $1 - F(t_i - 0)$ podem ser estimados por δ_i/n, $(n-i)/n$ e$(n-i+1)/n$ respectivamente. Portanto,

$$\widehat{Var}\{\hat{S}(t)\} = [\hat{S}(t)]^2 \sum_{j:t_j \leq t} \frac{\delta_i}{(n-i)(n-i+1)}.$$

No caso de empates, (2.36) torna-se

$$\widehat{Var}\{\hat{S}(t)\} = \frac{1}{n}[\hat{S}(t)]^2 \sum_{j:t_j \leq t} \frac{D_j}{N_j(N_j - D_j)}, \tag{2.37}$$

onde D_J e N_J foram definidos no inicio desta seção. Note que a expressão (2.37) coincide com a fórmula de Greenwood.

Utilizando o estimador P.L. e a relação $S(t) = \exp\{-\Lambda(t)\}$, Peterson (1977) sugere estimar a função de risco acumulada por

$$\hat{\Lambda}(t) \cong \sum_{j:t_j \leq t} -\log\left(1 - \frac{\delta_i}{n-i+1}\right). \tag{2.38}$$

Peterson(1977) ainda comenta que a expressão (2.38) converge para a função de risco acumulada $\Lambda(\cdot)$.

3. Estimação Paramétrica

Considere agora a situação em que os tempos de vida correspondem a valores de uma variável aleatória associada a uma distribuição de probabilidade conjunta P, indexada pelo vetor de parâmetros $\boldsymbol{\theta}$ pertencente ao espaço paramétrico $\Theta \subset R^k$. Isto é, P pertence a uma classe $\mathcal{P}$ tal que

$$\mathcal{P} = \{P_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta\}. \tag{3.1}$$

O problema de estimação da função de sobrevivência pode então ser colocado da seguinte maneira:

i. É dada uma função a valores reais g com domínio no espaço paramétrico Θ e cujo valor em $\boldsymbol{\theta}$ precisa ser estimado.

ii. Tem-se uma quantidade aleatória $\mathbf{T} = (T_1, \ldots, T_n)$ tomando valores de acordo com uma distribuição $P_{\boldsymbol{\theta}}$ que pertence a $\mathcal{P}$. Os valores que são efetivamente observados são ordenados e constituem a variável $t_i = \min(T_i, L_i)$ associados com indicadores $\delta_i = I_{\{T_i \leq L_i\}}$ onde L_i representa um limite de observação para o indivíduo i e δ_i indica a ocorrência ou não de censura, $i = 1, \ldots, n$.

iii. Através do método de máxima verossimilhança será obtido um valor para $\boldsymbol{\theta}$. Substituindo este valor na função de sobrevivência, obtemos o estimador de máxima verossimilhança $S_{\widehat{\boldsymbol{\theta}}}(t)$ para $S_{\boldsymbol{\theta}}(t)$.

Suponha que as variáveis aleatórias T_i e L_i são contínuas com função de distribuição $F_{\boldsymbol{\theta}}(\cdot)$ e $H(\cdot)$ respectivamente. As derivadas destas funções fornecem as densidades $f_{\boldsymbol{\theta}}(\cdot)$ e $h(\cdot)$, que irão compor a verossimilhança.

A densidade conjunta de (t_i, δ_i) denotada por $g_{\delta_i}(t_i)$, será dada por

$$g_{\delta_i}(t_i) = \begin{cases} f_{\boldsymbol{\theta}}(t_i)[1 - H(t_i)], & \text{se } \delta_i = 1; \\ h(t_i)[1 - F_{\boldsymbol{\theta}}(t_i)], & \text{se } \delta_i = 0. \end{cases}$$

Deste modo, segue que a verossimilhança é

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} [f_{\boldsymbol{\theta}}(t_i)]^{\delta_i} [1 - F_{\boldsymbol{\theta}}(t_i)]^{1-\delta_i} [h(t_i)]^{1-\delta_i} [1 - H(t_i)]^{\delta_i}. \tag{3.2}$$

Supondo que T_i é independente de L_i, a distribuição de L_i não deverá não envolver parâmetros de interesse. Consequentemente, costuma-se reescrever a verossimilhança con-

siderando apenas o seu *núcleo*, do que resulta

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} [f_{\boldsymbol{\theta}}(t_i)]^{\delta_i} [1 - F_{\boldsymbol{\theta}}(t_i)]^{1-\delta_i} \\ &= \prod_{i=1}^{n} [f_{\boldsymbol{\theta}}(t_i)]^{\delta_i} [S_{\boldsymbol{\theta}}(t_i)]^{1-\delta_i}. \end{aligned} \tag{3.3}$$

Tomando-se o logaritmo da verossimilhança, seguem

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\delta_i \log f_{\boldsymbol{\theta}}(t_i) + (1 - \delta_i) \log S_{\boldsymbol{\theta}}(t_i)\right], \tag{3.4}$$

e as equações normais

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_i} = 0, \qquad i = 1, \ldots, k.$$

Na maior parte das vezes a maximização de $L(\boldsymbol{\theta})$ é feita através de métodos iterativos que fornecem resultados aproximados para as equações normais. Um dos mais utilizados é o método de Newton-Raphson.

3.1. Método Iterativo de Newton-Raphson

Seja $\mathbf{U}(\boldsymbol{\theta})$ um vetor coluna de dimensão k cujos elementos são dados por $\frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_i}$ e $\mathbf{V}(\boldsymbol{\theta})$ uma matriz quadrada de dimensão k cujos elementos são dados por $\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$, $i, j = 1, \ldots n$. Seja $\boldsymbol{\theta}^0 = (\theta_1^0, \ldots, \theta_k^0)'$ um vetor coluna de dimensão k cujos componentes são valores próximos dos componentes de $\boldsymbol{\theta}$. Expandindo $\mathbf{U}(\boldsymbol{\theta})$ em série de Taylor em torno de $\boldsymbol{\theta}^0$ tem-se que

$$\mathbf{U}(\boldsymbol{\theta}) \cong \mathbf{U}(\boldsymbol{\theta}^0) + \mathbf{V}(\boldsymbol{\theta}^0)(\boldsymbol{\theta} - \boldsymbol{\theta}^0). \tag{3.5}$$

Supondo $\mathbf{V}$ uma matriz não singular, pode-se escrever

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}^0 + \mathbf{V}^{-1}(\boldsymbol{\theta}^0)[\mathbf{U}(\widehat{\boldsymbol{\theta}}) - \mathbf{U}(\boldsymbol{\theta}^0)]. \tag{3.6}$$

Se $\widehat{\boldsymbol{\theta}}$ estiver suficientemente próximo do estimador de máxima verossimilhança de $\boldsymbol{\theta}$, a expressão (3.6) pode ser escrita como

$$\widehat{\boldsymbol{\theta}} \cong \boldsymbol{\theta}^0 - \mathbf{V}^{-1}(\boldsymbol{\theta}^0)\mathbf{U}(\boldsymbol{\theta}^0), \tag{3.7}$$

que fornece a base do processo iterativo, isto é, dado um valor inicial $\boldsymbol{\theta}^0$ obtem-se o valor $\widehat{\boldsymbol{\theta}}$ através de (3.7). A quantidade obtida é considerada o valor inicial da iteração seguinte.

O processo segue até satisfazer alguma regra de parada, como por exemplo $\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 \| < \epsilon$, $\epsilon > 0$.

Se $L(\boldsymbol{\theta})$ for uma função "bem comportada" no sentido que $\log L(\boldsymbol{\theta})$ pode ser aproximada por uma expressão quadrática na vizinhança de $\boldsymbol{\theta}$, a convergência ocorrerá na maior parte das vezes; contudo, em algumas situações isto pode não ocorrer. Em tais casos procedimentos alternativos podem ser utilizados conforme discute Chambers(1977). Uma das modificacões mais comuns é a substituição da matriz **V** pelo inverso da esperança da matriz de informação de Fisher que será vista adiante. Esta alteração leva ao chamado *método do scoring* e simplifica o algoritmo em algumas situações.

3.2. Distribuição Assintótica

Para encontrar a distribuição assintótica do estimador de máxima verossimilhança suponha que o espaço paramétrico é de dimensão 1 e que $f_\theta(t)$ satisfaz às condições de regularidade (Rao, 1973, §5f), de modo que $\widehat{\theta}$ será fortemente consistente para θ e

$$n^{\frac{1}{2}}(\widehat{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, I^{-1}(\theta)),$$

onde $I(\theta) = E\left\{\left(\frac{\partial \log f_\theta(T)}{\partial \theta}\right)^2\right\}$ é a informação de Fisher. Uma prova formal destes resultados é dada por Serfling(1980). A distribuição assintótica para o estimador $S_{\widehat{\theta}}(t)$ pode então ser encontrada.

Teorema 5. *Se $f_\theta(t)$ satisfaz às condições de regularidade, então a variável aleatória $n^{\frac{1}{2}}(S_{\widehat{\theta}}(t) - S_\theta(t))$ terá distribuição assintótica normal com média 0 e variância dada por*

$$I^{-1}(\theta)\left\{\frac{\partial S_\theta(t)}{\partial \theta}\right\}^2, \tag{3.8}$$

onde $I(\theta) = \int_0^\infty \left\{\frac{\partial \log S_\theta(u)}{\partial \theta}\right\}^2 S_\theta(u)h(u)du + \int_0^\infty \left\{\frac{\partial \log f_\theta(u)}{\partial \theta}\right\}^2 f_\theta(u)[1 - H(u)]du.$

Prova: Da expressão (3.4),

$$\frac{\partial \log L(\theta)}{\partial \theta} = \sum_{j=1}^{n} \delta_j \frac{\partial \log f_\theta(t_j)}{\partial \theta} + (1 - \delta_j)\frac{\partial \log S_\theta(t_j)}{\partial \theta} = \sum_{j=1}^{n} \frac{\partial \log L_j^{\delta_j}(\theta)}{\partial \theta}.$$

O valor $\widehat{\theta}$ que maximiza $\log L(\theta)$ tem distribuição assintótica normal com média θ e

variância $(nI(\theta))^{-1}$ onde

$$I(\theta) = E\left\{\left(\frac{\partial \log L_j^{\delta_j}(\theta)}{\partial\theta}\right)^2\right\}$$
$$= \int_0^\infty \left(\frac{\partial \log L_j^0(\theta)}{\partial\theta}\right)^2 g_0(u)du + \int_0^\infty \left(\frac{\partial \log L_j^1(\theta)}{\partial\theta}\right)^2 g_1(u)du$$
$$= \int_0^\infty \left\{\frac{\partial \log S_\theta(u)}{\partial\theta}\right\}^2 S_\theta(u)h_(u)du + \int_0^\infty \left\{\frac{\partial \log f_\theta(u)}{\partial\theta}\right\}^2 f_\theta(u)[1-H(u)]du.$$

Utilizando o desenvolvimento em série de Taylor, $S_{\widehat{\theta}}(t) = S_\theta(t) + \frac{\partial S_{\widehat{\theta}}(t)}{\partial\widehat{\theta}}|_{\widehat{\theta}=\theta}(\widehat{\theta}-\theta) + O_p(n^{-1}) \Rightarrow S_{\widehat{\theta}}(t) - S_\theta(t) = \frac{\partial S_\theta(t)}{\partial\theta}(\widehat{\theta}-\theta) + O_p(n^{-1})$. Portanto, pelo método delta, segue que $n^{\frac{1}{2}}(S_{\widehat{\theta}}(t) - S_\theta(t))$ terá distribuição assintótica normal com média 0 e variância dada por

$$I^{-1}(\theta)\left(\frac{\partial S_\theta(t)}{\partial\theta}\right)^2.$$

□

O resultado do teorema pode ser estendido para o caso em que o espaço paramétrico é de dimensão k, de modo que $n^{\frac{1}{2}}(S_{\widehat{\boldsymbol{\theta}}}(t) - S_{\boldsymbol{\theta}}(t))$ terá distribuição assintótica normal com vetor de médias $\mathbf{0}$ e variância dada por

$$\partial\mathbf{S}'\mathbf{I}^{-1}(\boldsymbol{\theta})\partial\mathbf{S}, \tag{3.9}$$

onde $\partial\mathbf{S}$ é um vetor coluna de dimensão k cujos elementos correspondem às derivadas de $S(t)$ em relação a θ_i, $i = 1,\ldots,k$ e $\mathbf{I}(\boldsymbol{\theta})$ é a matriz de informação de Fisher definida apropriadamente para observações sujeitas a censura, cujo elemento (i,j) é dado por

$$E\left\{\frac{\partial \log L_i^{\delta_i}(\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial \log L_j^{\delta_j}(\boldsymbol{\theta})}{\partial\theta_j}\right\}.$$

Para $k = 2$, analogamente ao desenvolvimento utilizado na demostração do teorema 5, segue que

$$\mathbf{I}(\boldsymbol{\theta}) = \int_0^{+\infty}\left[\mathbf{A}\mathbf{A}'S_{\boldsymbol{\theta}}(u)h(u) + \mathbf{B}\mathbf{B}'f_{\boldsymbol{\theta}}(u)(1-H(u))\right]du$$

onde $\mathbf{A}' = \left(\frac{\partial \log S_{\boldsymbol{\theta}}(u)}{\partial\theta_1}, \frac{\partial \log S_{\boldsymbol{\theta}}(u)}{\partial\theta_2}\right)$ e $\mathbf{B}' = \left(\frac{\partial \log f_{\boldsymbol{\theta}}(u)}{\partial\theta_1}, \frac{\partial \log f_{\boldsymbol{\theta}}(u)}{\partial\theta_2}\right)$.

4. Estimação Não Paramétrica x Estimação Paramétrica

Utilizando os resultados das seções anteriores pode-se comparar a eficiência do estimador produto-limite com relação ao estimador de máxima verossimilhança.

Supõe-se que os tempos de vida e os tempos de censura correspondem a variáveis aleatórias contínuas com função de distribuição dadas por $F_{\theta}(t)$ e $H(t)$, respectivamente. Com os resultados assintóticos obtidos, pode ser calculada a eficiência assintótica do estimador P.L. com relação ao estimador M.V..

Uma vez que a distribuição destes estimadores não é conhecida no caso de pequenas amostras, o comportamento da eficiência nesta situação é investigado através da razão de estimativas dos erros quadrático médio obtidas a partir da simulação de 2000 amostras de tamanhos variáveis ($n = 10, 20, 30, 50$ e 500), dadas por

$$\sum_{k=1}^{2000} \frac{\{\widehat{S}_k(t) - S(t)\}^2}{2000},$$

em situações onde a proporção de sobreviventes (PS) é 0.50, 0.25 e 0.10 e a proporção de censurados (PC) é alta (0.50), média (0.25) e nula. A escolha destes valores segue o mesmo padrão utilizado por Miller(1983) para o caso assintótico. Duas situações são consideradas: tempos de vida distribuídos exponencialmente e segundo uma Weibull. Em ambas as situações os tempos de censura seguem uma distribuição exponencial.

4.1. Tempos de Vida com Distribuição Exponencial

Considerando $S_{\theta}(t) = 1 - F_{\theta}(t) = \epsilon^{-\theta t}$ e $1 - H(t) = \epsilon^{-\mu t}$, com $\theta > 0$ e $\mu > 0$, obtem-se por (2.36) a variância assintótica do estimador P.L., dada por

$$\begin{aligned} Var\{\widehat{S}(t)\} \cong & \frac{1}{n}[S_{\theta}(t)]^2 \int_0^t \frac{f_{\theta}(u)}{[1-H(u)][S_{\theta}(u)]^2} du = \frac{1}{n}\epsilon^{-2\theta t} \int_0^t \frac{\theta \epsilon^{-\theta u}}{\epsilon^{-\mu u}\epsilon^{-2\theta u}} du \\ = & \frac{1}{n}\epsilon^{-2\theta t}\left(\epsilon^{(\theta+\mu)t} - 1\right)\frac{\theta}{\theta+\mu}. \end{aligned} \tag{4.1}$$

Para o estimador de máxima verossimilhança, segue que $I^{-1}(\theta) = \theta(\theta + \mu)$, de modo que da expressão (3.8),

$$Var\{S_{\widehat{\theta}}(t)\} \cong \frac{1}{n} I^{-1}(\theta)\left\{\frac{\partial S_{\theta}(t)}{\partial \theta}\right\} = \frac{1}{n}\theta t^2(\theta+\mu)\epsilon^{-2\theta t}. \tag{4.2}$$

Portanto a eficiência assintótica do estimador P.L. com relação ao estimador M.V. será dada por

$$Ef_A(t) = \frac{Var\{S_{\widehat{\theta}}(t)\}}{Var\{\widehat{S}(t)\}} = (\theta+\mu)^2 t^2 \left(e^{(\theta+\mu)t} - 1\right)^{-1}$$
$$= (1+\rho)^2(\theta t)^2 \left[e^{(1+\rho)\theta t} - 1\right]^{-1}, \qquad (4.3)$$

onde $\rho = \mu/\theta$.

Da expressão (4.3) pode-se concluir que a eficiência assintótica aproxima-se de zero para $t \to 0$ e para $t \to +\infty$. Além disso, quando $(1+\rho)\theta t = 1,59$ a função (4.3) atinge o máximo de 0,6476, o maior valor possível. Nota-se também que a medida que a censura torna-se mais pesada ($\rho \to +\infty$) diminui a eficiência assintótica.

A proporção de sobreviventes (PS) é dada por $P\{T > t\} = e^{-\theta t}$, e a proporção de censura (PC) por $P\{T_i > L_i\} = \frac{\rho}{1+\rho}$. Considerando estas expressões e os valores de PS e PC mencionados anteriormente, é possível calcular estimativas para os estimadores P.L. e M.V. a partir das amostras simuladas, e os correspondentes E.Q.M. estimados, de modo que no caso não assintótico,

$$\widehat{Ef}(t) = \frac{\widehat{\text{EQM}}(S_{\widehat{\theta}}(t))}{\widehat{\text{EQM}}(\widehat{S}(t))},$$

onde $\widehat{\text{EQM}}(S_{\widehat{\theta}}(t))$ e $\widehat{\text{EQM}}(\widehat{S}(t))$ são as estimativas dos erros quadrático médio para os estimadores M.V. e P.L. respectivamente. Os resultados obtidos são apresentados na tabela I. Nesta situação as estimativas M.V. de $S(t)$ podem ser obtidas sem a necessidade de processos iterativos.

PC	n	PS 0,50	0,25	0,10
0,50	10	0,5720	0,4838	0,3120
$\rho = 1$	20	0,6296	0,4662	0,2385
	30	0,6170	0,4911	0,2414
	50	0,6403	0,4591	0,1971
	500	0,6435	0,5015	0,2022
	∞	0,6402	0,5112	0,2149
0,25	10	0,5521	0,5555	0,3945
$\rho = 0.33$	20	0,5803	0,5792	0,4173
	30	0,5643	0,5987	0,4213
	50	0,5618	0,5971	0,4270
	500	0,5666	0,6368	0,4734
	∞	0,5601	0,6386	0,4608
0	10	0,4992	0,5890	0,5152
$\rho = 0$	20	0,5181	0,6000	0,5302
	30	0,5037	0,6326	0,5522
	50	0,5135	0,6305	0,5506
	500	0,4793	0,6163	0,5869
	∞	0,4791	0,6409	0,5895

TABELA I - Eficiência do estimador produto-limite em relação ao estimador de máxima verossimilhança. Tempos de vida exponenciais.

4.2. Tempos de Vida com Distribuição Weibull

Sendo $S_{\boldsymbol{\theta}}(t) = \exp\{-(\theta t)^{\alpha}\}$ com $\boldsymbol{\theta} = (\alpha; \theta)$, $\alpha > 0$, $\theta > 0$ e $1 - H(t) = \exp\{-\mu t\}$ com $\mu > 0$, segue de (2.36) que, assintóticamente,

$$\begin{aligned} Var\{\widehat{S}(t)\} &\cong \frac{1}{n}[S_{\boldsymbol{\theta}}(t)]^2 \int_0^t \frac{f_{\boldsymbol{\theta}}(u)}{[1 - H(u)][S_{\boldsymbol{\theta}}(u)]^2} du \\ &= \frac{1}{n} \exp\{-2(\theta t)^{\alpha}\} \int_0^t \alpha \theta^{\alpha} u^{\alpha - 1} \exp\{(\theta^{\alpha} + \mu)u\} du \qquad (4.4) \\ &= \frac{1}{n} \exp\{-2(\theta t)^{\alpha}\} \alpha \int_0^{\theta t} x^{\alpha - 1} \exp\{(x^{\alpha} + \rho x)\} dx, \end{aligned}$$

onde $\rho = \mu/\theta$.

Caso o parâmetro α seja conhecido, a variância assintótica do estimador M.V. segue de (3.8). Neste caso, $I^{-1}(\theta) = [(\alpha/\theta)^2(1 - \mathrm{PC})]^{-1}$, onde $\mathrm{PC} = P\{T_i > L_i\} =$

$\int_0^{+\infty} \rho \exp\{-u^\alpha - \rho u\} du$ é a proporção de indivíduos censurados. Portanto,

$$Var\{S_{\widehat{\theta}}(t)\} \cong \frac{1}{n} \frac{(\theta t)^{2\alpha} \exp\{-2(\theta t)^\alpha\}}{(1 - \text{PC})}. \tag{4.5}$$

Logo a eficiência assintótica será dada por

$$Ef_A(t) = \frac{(\theta t)^{2\alpha}}{(1 - \text{ PC })} \frac{1}{\alpha \int_0^{\theta t} x^{\alpha-1} \exp\{x^\alpha + \rho x\} dx}. \tag{4.6}$$

Calculando-se as integrais envolvidas numericamente, podem ser obtidos valores para (4.6) nas situações propostas para PS e PC. Quando não existem observações censuradas, a expressão (4.6) fica igual à expressão da eficiência assintótica no caso exponencial.

Para o caso em que as amostras são de tamanho reduzido, são feitas simulações e comparados os erros quadrático médio, a exemplo do modelo exponencial. Como α é suposto conhecido, não é necessária a utilização de processo iterativo para se obter as estimativas M.V. de $S(t)$. Os resultados obtidos seguem na tabela II com PS= $P\{T > t\} = \exp\{-(\theta t)^\alpha\}$ e $\alpha = 2$.

		PS		
PC	n	0.50	0.25	0,10
0,50	10	0.5245	0,4991	0,3988
$\rho = 1$	20	0.5733	0.4855	0,3659
	30	0,5725	0.5102	0.3561
	50	0.5853	0.5394	0.3647
	500	0,5813	0.5633	0.3855
	∞	0,5673	0,5806	0,3978
0,25	10	0.5281	0,5407	0,4444
$\rho = 0.33$	20	0.5494	0.5689	0,4698
	30	0.5445	0.5947	0,4813
	50	0,5369	0,6133	0.4873
	500	0.5256	0,6208	0,5320
	∞	0,5219	0,6305	0,5166
0	10	0,4992	0,5890	0,5152
$\rho = 0$	20	0,5181	0,6000	0,5302
	30	0,5037	0,6326	0,5522
	50	0,5135	0,6305	0,5506
	500	0,4793	0,6163	0,5869
	∞	0,4791	0,6409	0,5895

TABELA II - Eficiência do estimador produto-limite em relação ao estimador de máxima verossimilhança. Tempos de vida Weibull—α=2, conhecido.

Se o parâmetro α é desconhecido, a variância assintótica do estimador M.V. será dada por (3.9). Neste caso,

$$\mathbf{A}' = -(\theta u)^{\alpha}\,(\log\theta; \theta^{\alpha}u^{\alpha})$$

$$\mathbf{B}' = \left(\frac{1}{n} + \log\theta u - \theta^{\alpha}u^{\alpha}\log(\theta u); \frac{\alpha}{\theta}(1-\theta^{\alpha}u^{\alpha})\right)$$

Após alguns cálculos envolvendo integração por partes obtem-se a matriz de informação de Fisher:

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\alpha^3}{\theta^2}i_1 & \frac{\alpha}{\theta}i_2 \\ \frac{\alpha}{\theta}i_2 & \frac{1}{\alpha}i_3 \end{pmatrix}$$

onde $i_1 = \int_0^{\infty} u^{\alpha-1}\exp\{-\rho u - u^{\alpha}\}du$, $i_2 = \int_0^{\infty} u^{\alpha-1}\exp\{-\rho u - u^{\alpha}\}(1+\log u^{\alpha})du$ e $i_3 = \int_0^{\infty} u^{\alpha-1}\exp\{-\rho u - u^{\alpha}\}(1+\log u^{\alpha})^2 du$, com $\rho = \mu/\theta$. Como

$$\partial \mathrm{S}' = -\theta^{\alpha}t^{\alpha}\exp\{-\theta^{\alpha}t^{\alpha}\}\left(\frac{\alpha}{\theta}; \log(\theta t)\right),$$

segue que

$$Var\{S_{\widehat{\boldsymbol{\theta}}}\}(t) \cong \frac{1}{n}\frac{(\theta t)^{2\alpha}\exp\{-(\theta t)^{\alpha}\}}{i_1 i_3 - i_2^2}\left[\frac{i_3}{\alpha} - \log(\theta t)(2i_2 - \alpha i_1\log(\theta t))\right]$$

Logo a eficiência assintótica no caso em que ambos os parâmetros α e θ são desconhecidos será dada por

$$Ef_A(t) = \frac{(\theta t)^{2\alpha}\left[\frac{i_3}{\alpha} - \log(\theta t)(2i_2 - \alpha i_1\log(\theta t))\right]}{\alpha i_4(i_1 i_3 - i_2^2)} \tag{4.7}$$

As integrais envolvidas em (4.7) podem ser calculadas numericamente, de modo que as eficiências assintóticas podem ser encontradas. Para o caso de pequenas amostras, estimativas da eficiência são obtidas através de simulações. Os resultados encontram-se na tabela III. As estimativas M.V. de $S(t)$ são agora obtidas através do processo iterativo de

Newton-Raphson e na geração das amostras foi considerado $\alpha = 2$.

		PS		
PC	n	0,50	0,25	0,10
0,50	10	0,7662	0,6145	0,3689
$\rho = 1$	20	0,7216	0,5981	0,4182
	30	0,6627	0,5929	0,4416
	50	0,6531	0,6264	0,4818
	500	0,6199	0,6073	0,5393
	∞	0,6019	0,6219	0,5590
0,25	10	0,7298	0,6796	0,4460
$\rho = 0.33$	20	0,6811	0,6719	0,5400
	30	0,6893	0,6688	0,5543
	50	0,6631	0,6699	0,5861
	500	0,6456	0,6259	0,6286
	∞	0,6344	0,6335	0,6225
0	10	0,6808	0,7314	0,5365
$\rho = 0$	20	0,6905	0,6814	0,5830
	30	0,7084	0,6965	0,6069
	50	0,6800	0,6742	0,6225
	500	0,6738	0,6258	0,6392
	∞	0,6623	0,6443	0,6494

TABELA III - Eficiência do estimador produto-limite em relação ao estimador de máxima verossimilhança. Tempos de vida Weibull—α=2, desconhecido.

4.3. Comentários

A eficiência do estimador P.L. é baixa, sobretudo quando a censura é pesada e a proporção de sobreviventes é pequena (valores altos para t).

Assintoticamente, quando o valor de PS é alto (tempos próximos da origem), os estimadores têm uma precisão maior, de modo que ambos devem fornecer boas estimativas para $S(t)$. No entanto, a medida que t aumenta de valor ocorre um acréscimo nas variâncias assintóticas fazendo com que qualquer perda na eficiência represente uma real diminuição na precisão dos estimadores.

Para pequenas amostras, quando a censura é alta (PC = 0,50) os três modelos considerados têm suas eficiências diminuídas a medida que t aumenta (e portanto diminue a porcentagem de sobreviventes). Para os outros valores de PC os modelos exponencial

e Weibull com α conhecido apresentam maiores eficiências para valores intermediários de t, que declinam quando o tempo aumenta, ao passo que no modelo Weibull com α desconhecido a eficiência sempre diminui conforme o tempo se afasta da origem. No caso sem censura a mesma tendência se verifica com maior nitidez. No modelo Weibull com α desconhecido a eficiência do estimador P.L. é sempre maior que nos outros modelos devido à estimação de *dois* parâmetros que causa um aumento no erro quadrático médio do estimador paramétrico de $S(t)$. Ainda assim o estimador M.V. continua a ser mais eficiente.

Na construção das tabelas anteriores está implicitamente suposto que o modelo assumido para os dados é correto. Para grandes amostras a identificação do modelo é razoavelmente frequente e o uso da estimação paramétrica é indicado. No caso de pequenas amostras aumentam os problemas na identificação do modelo, de modo que a eficiência do estimador P.L. em relação ao estimador M.V. deve sofrer um acréscimo, uma vez que a estimativa de máxima verossimilhança será viciada tornando a estimação não paramétrica uma opção importante.

Capítulo 2

Estimação em Modelos de Regressão

1. Introdução

Em situações práticas é muito comum que a cada unidade experimental, além de se observar o tempo de morte ou falha, sejam observadas covariadas que são medidas em determinadas circunstâncias ou assumem códigos relacionados a categorias às quais as unidades podem ser classificadas. Em ensaios médicos, por exemplo, podem ser observados os tempos de vida de pacientes submetidos a transplantes de coração, controlados pela idade bem como pela ocorrência ou não de rejeição ao orgão. Em tais situações, o desenvolvimento de modelos que comportam covariadas se faz necessário, com extensão para a estimação com dados censurados.

Considere agora a situação em que as quantidades observadas correspondem ao terno $(t_i, \delta_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ onde $t_i = \min\{T_i, L_i\}$ com T_i representando o tempo de falha e L_i o limite de observação do i-ésimo indivíduo; δ_i é o indicador de censura e também é observado $\mathbf{x}_i$, o vetor de covariadas associado ao i-ésimo indivíduo. Supõe-se que T_i e L_i são variáveis aleatórias independentes, condicionalmente a $\mathbf{x}_i$.

Este capítulo trata da estimação dos parâmetros associados às covariadas. Considera-se inicialmente uma abordagem não paramétrica, que procura generalizar métodos desenvolvidos para modelos sem censura para o caso em que existem observações incompletas. Em seguida é desenvolvida a estimação paramétrica através do método de máxima verossimilhança, com censuras aleatórias. Uma comparação dos dois métodos de estimação é apresentada no final, utilizando-se resultados teóricos no caso sem censura e numéricos para amostras censuradas.

2. Estimação Não Paramétrica

No caso em que é difícil identificar a distribuição de modelos probabilísticos para os dados, alguns autores propõem estender o método de mínimos quadrados para o caso em que existem dados censurados. Nesta linha, Miller(1976) e Buckley e James(1979) generalizam o modelo linear padrão. Cox(1972) desenvolve métodos para inferência baseados em modelos de *riscos proporcionais*. Convém mencionar que apesar de não desenvolvidos aqui, outros enfoques têm sido utilizados como por exemplo o modelo de Koul, Susarla e Van Ryzin(1982).

2.1. Estimadores de Miller

Suponha que os tempos de vida estão relacionados às covariadas segundo o modelo

$$T_i = \alpha + \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i, \tag{2.1}$$

onde α é um escalar representando o intercepto, $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p)$ é um vetor de dimensão 1 x p contendo os coeficientes de regressão para as covariadas $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{pi})$ e ϵ_i corresponde a uma variável aleatória cuja distribuição $F(\cdot)$ não é especificada mas sabe-se que $E\{\epsilon_i\} = 0$ e $Var\{\epsilon_i\} = \sigma^2 < +\infty$. Para $i = 1, \ldots, n$ as variáveis ϵ_i são supostas independentes.

No caso sem censuras, a estimação de $\boldsymbol{\beta}$ em (2.1) pode ser feita através do método de mínimos quadrados, que consiste em encontrar para α e $\boldsymbol{\beta}$ valores a e $\mathbf{b}' = (b_1, \ldots, b_p)$ de modo que a expressão

$$\sum_{i=1}^{n}(T_i - a - \mathbf{x}_i\mathbf{b})^2$$

assuma o menor valor possível. Isto é equivalente a minimizar $n^{-1}\sum_{i=1}^{n}(T_i - a - \mathbf{x}_i\mathbf{b})^2 = n^{-1}\sum_{i=1}^{n}\epsilon_i^2$ que pode simbolicamente ser escrito como

$$\int \epsilon^2 d\widehat{F}(\epsilon),$$

onde $d\widehat{F}(\epsilon)$ correponde aos saltos de $\widehat{F}(\epsilon)$, a função de distribuição empírica, nos pontos de descontinuidade.

No caso com censura, Miller(1976) propõe substituir $\widehat{F}(\cdot)$ pelo estimador produto-limite da função de distribuição $\widehat{F}^{PL}(\cdot) = 1 - \widehat{S}(\cdot)$, calculado com base nos valores $e_i = t_i - a - \mathbf{x}_i\mathbf{b}$ devidamente ordenados. Desta forma, a integral anterior fica

$$\int \epsilon^2 d\widehat{F}^{PL}(\epsilon). \tag{2.2}$$

Para valores fixados de a e $\mathbf{b}$ (2.2) é a soma de quadrados ponderada

$$\sum_{i=1}^{n}(t_i - a - \mathbf{x}_i\mathbf{b})^2 w_i(\mathbf{b}),$$

onde $w_i(\mathbf{b})$ é igual a zero se t_i é censurado e igual ao salto dado pelo estimador PL nos pontos de descontinuidade, expresso por $d\widehat{F}^{PL}(e_i) = \{\prod_{e_{(j)}<e_i}(\frac{n-j}{n-j+1})\}/(n-j+1)$ se t_i não é censurado com $e_{(j)}$ correspondendo ao j-ésimo valor da amostra ordenada $e_{(1)} < \ldots < e_{(n)}$.

A obtenção dos estimadores consiste em se minimizar (2.2). Entretanto, procedimentos baseados em desenvolvimentos em séries de potências não podem ser utilizados. Analizando os pesos $w_i(\cdot)$ verifica-se que não dependem de a pois qualquer que seja seu valor a relação de ordem entre $e_1, e_2, \ldots, e_n$ será preservada; o mesmo não ocorre com $\mathbf{b}$ que conforme muda de valor pode alterar a ordenação dos e_i, $i = 1, \ldots, n$. Se esta alteração envolver um e_i calculado sobre um tempo censurado, $w_i(\cdot)$ causará descontinuidades em (2.2) quando esta é vista como função de $\mathbf{b}$. Portanto, a obtenção dos valores a e $\mathbf{b}$ terá de ser feita através de procedimentos denominados *shrinking grid*, que se baseiam na criação de reticulados formados através de partições nos intervalos de variação dos elementos de $\mathbf{b}$ e a partir dos nós formados, calcular (2.2). O processo é repetido sobre partições com um número crescente de elementos até a obtenção do mínimo. Note que o intercepto não necessita ser considerado nesta fase se for convencionado que $\widehat{F}^{PL}(\epsilon) = 1$ para $\epsilon > e_{(n)}$ quando $e_{(n)}$ é censurado, pois aí $\sum_{i=1}^{n} w_i(\mathbf{b}) = 1$ e denotando o estimador de a por $\widehat{a}^*$ segue que

$$\widehat{a}^* = \sum_{i=1}^{n} w_i(\mathbf{b})(t_i - \mathbf{x}_i\mathbf{b}).$$

Desta forma, assumindo um valor para $\widehat{a}^*$, os coeficientes $\boldsymbol{\beta}$ serão estimados pelo vetor $\mathbf{b}$ que minimiza

$$\sum_{i=1}^{n}(t_i - \widehat{a}^* - \mathbf{x}_i\mathbf{b})^2 w_i(\mathbf{b}). \tag{2.3}$$

O procedimento descrito apresenta algumas desvantagens. Dado que o ínfimo de (2.3) pode ocorrer em uma descontinuidade, não pode ser feito uso de expansões para estudos assintóticos. Além disso, Miller(1976) apresenta um exemplo em que as estimativas de $\boldsymbol{\beta}$ não convergem para os verdadeiros valores populacionais quando n tende a infinito. Outro agravante é que os procedimentos do tipo *shrinking grid* costumam se tornar extremamente laboriosos em dimensões maiores que 2 (mais que dois parâmetros

no modelo). Estes problemas sugerem a necessidade de se modificar o processo. Uma alteração proposta consiste em se utilizar um estimador inicial $\widehat{\beta}^0$, dado pelo estimador de mínimos quadrados usual, calculado apenas sobre os dados não censurados, que pode ser escrito como

$$\widehat{\beta}^0 = \left[(\mathbf{X} - \overline{\mathbf{X}}_{NC})'\Delta(\mathbf{X} - \overline{\mathbf{X}}_{NC})\right]^{-1}(\mathbf{X} - \overline{\mathbf{X}}_{NC})'\Delta\mathbf{t}, \tag{2.4}$$

onde $\mathbf{X}' = (\mathbf{x}_1', \ldots, \mathbf{x}_n')$, $\overline{\mathbf{X}}_{NC}$ é uma matriz de dimensão n x p cuja i-ésima coluna contém todos os elementos iguais a $\overline{x}_i$, a média da i-ésima covariada calculada para os dados não censurados, Δ é uma matriz diagonal cujo elemento (i,i) é igual a δ_i, $i = 1, \ldots, n$ e $\mathbf{t}' = (t_1, \ldots, t_n)$.

Uma vez obtido $\widehat{\beta}^0$, calculam-se os pesos $w_i(\widehat{\beta}^0)$ sobre os valores $e_i = t_i - \mathbf{x}_i\widehat{\beta}^0$, $i = 1, \ldots, n$ e ao invés de reclassificar um $e_{(n)}$ censurado como não censurado sugere-se utilizar os pesos normalizados

$$w_i^*(\widehat{\beta}^0) = \frac{w_i(\widehat{\beta}^0)}{\sum_{j=1}^n w_j(\widehat{\beta}^0)}. \tag{2.5}$$

Com estes novos pesos, o procedimento usual de mínimos quadrados é utilizado, de modo que o estimador de Miller para β denotado por $\widehat{\beta}^M$, será dado por

$$\widehat{\beta}^M = \left[(\mathbf{X} - \overline{\mathbf{X}}^*)'\mathbf{W}^*(\widehat{\beta}^0)(\mathbf{X} - \overline{\mathbf{X}}^*)\right]^{-1}(\mathbf{X} - \overline{\mathbf{X}}^*)'\mathbf{W}^*(\widehat{\beta}^0)\mathbf{t}, \tag{2.6}$$

onde $\mathbf{X}$ e $\mathbf{t}$ são definidos da mesma forma que em (2.4), $\overline{\mathbf{X}}^*$ é uma matriz de dimensão n x p cuja j-ésima coluna contém todos os elementos iguais à média ponderada $\overline{x}_j^* = \sum_{i=1}^n w_i^*(\widehat{\beta}^0)x_{ij}$ e $\mathbf{W}^*(\widehat{\beta}^0)$ é uma matriz diagonal culo elemento (i,i) é dado por $w_i^*(\widehat{\beta}^0)$. O passo seguinte é utilizar $\widehat{\beta}^M$ como novo $\widehat{\beta}^0$ no cálculo dos pesos (2.5) e assim dar início ao processo iterativo.

Miller(1976) comenta que em alguns casos simulados $\widehat{\beta}^M$ não convergiu, tomando sempre uma mesma sequência de dois ou mais valores após certo número de iterações. Neste caso, sugere-se tomar o valor médio destes valores caso não sejam muito discrepantes. As simulações também mostraram que a utilização de $w_i^*(\cdot)$ aumentam as possibilidades de convergência.

Uma vez obtido um valor para $\widehat{\beta}^M$, a estimativa para o intercepto é dada por

$$\widehat{\alpha}^M = \sum_{i=1}^n w_i(\widehat{\beta})(t_i - \mathbf{x}_i\widehat{\beta}^M). \tag{2.7}$$

Nesta expressão a ponderação é feita com os pesos não normalizados e sem se reclassificar uma observação $e_{(n)}$ censurada como não censurada. Desta forma, tende a ser atenuada a sub-estimação de α causada pelas censuras.

Resultados assintóticos para $\widehat{\boldsymbol{\beta}}^M$ são estudados heuristicamente. Entretanto, para assegurar a consistência assintótica, é feita a suposição crucial de que a distribuição das censuras não varia com $\mathbf{x}$, o vetor de covariadas, de modo que

$$H(t, \mathbf{x}) = H(t - \alpha - \mathbf{x}\boldsymbol{\beta}). \tag{2.8}$$

Sob esta condição, o estimador $\widehat{\boldsymbol{\beta}}^0$ calculado sob as observações não censuradas é consistente. Para se ter uma idéia dos passos dados por Miller(1976) na determinação da distribuição assintótica, considere o caso $p = 1$. Então, de (2.6)

$$\widehat{\beta}^M = \frac{\sum_{i=1}^n w_i^*(\widehat{\beta}^0)(x_i - \overline{x}^*)t_i}{\sum_{i=1}^n w_i^*(\widehat{\beta}^0)(x_i - \overline{x}^*)^2},$$

que simbólicamente pode ser escrita como

$$\widehat{\beta}^M = \frac{\int t(x - \overline{x}^*)d\widehat{F}^{PL}(t)}{\int (x - \overline{x}^*)^2 d\widehat{F}^{PL}(t)}.$$

Lembrando que $t = \alpha + \beta x + \epsilon$, segue que

$$n^{\frac{1}{2}}(\widehat{\beta}^M - \beta) = n^{\frac{1}{2}} \frac{\int \epsilon(x - \overline{x}^*)d\widehat{F}^{PL}(\epsilon)}{\int (x - \overline{x}^*)^2 d\widehat{F}^{PL}(\epsilon)}. \tag{2.9}$$

Supondo que $\widehat{\beta}^M$ converge em probabilidade para β e assumindo que a suposição (2.8) vale, segue pela lei fraca dos grandes números que $\lim_{n\to+\infty} \overline{x}^* = \lim_{n\to+\infty} \overline{x} = \eta$, e assim

$$\lim_{n\to+\infty} \int (x - \overline{x})^2 d\widehat{F}^{PL}(\epsilon) = \tau^2, \tag{2.10}$$

onde η e τ^2 são constantes.

Caso o verdadeiro valor de β seja utilizado no cálculo dos pesos segue do teorema 4 do capítulo 1 que a quantidade $n^{\frac{1}{2}}\{\widehat{F}^{PL}(\cdot) - F(\cdot)\}$ converge fracamente para um processo gaussiano de média zero. Desta forma, $\widehat{F}^{PL}(t)$ tem distribuição assintótica normal para t fixado. Encarando-se o numerador de (2.9) como uma combinação linear de variáveis normais, ele deverá convergir para uma distribuição normal com média zero. Para obter

a variância assintótica, após trabalhosos cálculos envolvendo esperança condicional e expansões, verifica-se que

$$\lim_{n\to+\infty} nE\{\int e(x-\overline{x})d\widehat{F}^{PL}(e)\}^2 = \tau^2 \int e^2 \lim_{n\to+\infty} nE\{d\widehat{F}^{PL}(e)\}^2,$$

e que

$$\lim_{n\to+\infty} nE\{d\widehat{F}^{PL}(e)\}^2 = \frac{dF(e)}{1-H(e)}, \tag{2.11}$$

de modo que a variância assintótica de $\widehat{\beta}^M$ será dada por

$$Var\{\widehat{\beta}^M\} = \frac{1}{n\tau^2}\int e^2 \frac{dF(e)}{1-H(e)}. \tag{2.12}$$

Para estimar (2.12), inspirando-se em (2.10), (2.1) e (2.11) são feitas as aproximações $\tau^2 \cong \sum_{i=1}^n w_i^*(\widehat{\beta}^M)(x_i-\overline{x}^*)^2$, $e_i^2 \cong (t_i - \widehat{\alpha}^M - \widehat{\beta}^M x_i)$ e $d\widehat{F}^{PL}(e_i)/\{n(1-H(e_i))\} \cong \{w_i^*(\widehat{\beta}^m)\}^2$. Portanto,

$$\widehat{Var}\left\{\widehat{\beta}^M\right\} = \frac{\sum_{i=1}^n \{w_i^*(\widehat{\beta}^M)\}^2(t_i - \widehat{\alpha}^M - \widehat{\beta}^M x_i)^2}{\sum_{i=1}^n w_i^*(\widehat{\beta}^M)(x_i-\overline{x}^*)^2}, \tag{2.13}$$

e para o caso de p covariadas,

$$\widehat{Var}\left\{\widehat{\boldsymbol{\beta}}^M\right\} = \sum_{i=1}^n \{w_i^*(\widehat{\boldsymbol{\beta}}^M)\}^2(t_i - \widehat{\alpha}^M - \mathbf{x}_i\widehat{\boldsymbol{\beta}}^M)^2 \left[(\mathbf{X}-\overline{\mathbf{X}}^*)'\mathbf{W}^*(\widehat{\boldsymbol{\beta}})(\mathbf{X}-\overline{\mathbf{X}}^*)\right]^{-1}. \tag{2.14}$$

Todo o desenvolvimento assintótico se baseia na suposição (2.8). Esta é a principal desvantagem do método, uma vez que em situações práticas este padrão de censuras nem sempre é observado.

2.2. Estimadores de Buckley–James

Uma abordagem diferente para estender a teoria de modelos lineares padrão para amostras censuradas foi desenvolvida por Buckley e James(1979) utilizando uma metodologia análoga à utilizada por Schmee e Hahn(1979) para a distribuição normal. Ao contrário dos estimadores de Miller, que são obtidos através de modificações no método de mínimos quadrados, é sugerido modificar as equações normais e então obter os estimadores de α e $\boldsymbol{\beta}$. Apesar de pouco ser conhecido para o caso assintótico, esta abordagem parece não exigir a condição (2.8).

No caso em que não existem censuras, o método de mínimos quadrados aplicado ao modelo (2.1) gera as equações normais

$$\mathbf{1}'\mathbf{t} - n\alpha - \mathbf{1}'\mathbf{X}\boldsymbol{\beta} = 0$$

e

$$[\mathbf{t} - \mathbf{X}\beta]'[\mathbf{X} - \overline{\mathbf{X}}]\mathbf{1}_p = 0, \tag{2.15}$$

onde $\overline{\mathbf{X}}$ é uma matriz n x p cujos elementos da i-ésima coluna são todos iguais à média da i-ésima covariada para as n observações, $\mathbf{1}$ é um vetor coluna de dimensão n e $\mathbf{1}_p$ é um vetor coluna de dimensão p, ambos com todos os elementos iguais a 1, $\mathbf{t}' = (t_1, \ldots, t_n)$, $\beta' = (\beta_1, \ldots, \beta_p)$ e $\mathbf{X}' = (\mathbf{x}'_1, \ldots, \mathbf{x}'_p)$. Para $p = 1$, as equações normais ficam

$$\sum_{i=1}^{n}(t_i - \alpha - \beta x_i) = 0$$

e

$$\sum_{i=1}^{n}(t_i - x_i\beta)(x_i - \overline{x}) = 0.$$

Para utilizar o método de mínimos quadrados em amostras censuradas, Buckley e James(1979) observam que $E(t_i) \neq \alpha + \mathbf{x}_i\boldsymbol{\beta}$ e propõem utilizar no lugar de t_i as variáveis *pseudo-aleatórias*

$$t_i^* = T_i\delta_i + E(T_i \mid T_i > L_i)(1 - \delta_i), \qquad i = 1, \ldots, n. \tag{2.16}$$

Para estas variáveis segue que

$$\begin{aligned}
E(t_i^*) &= E(T_i\delta_i) + E\{E(T_i \mid T_i > L_i)(1 - \delta_i)\} \\
&= E\left[E\left(T_i I_{\{T_i \leq L_i\}} | T_i\right)\right] + E\left[E\left\{E\left(T_i | T_i > L_i\right) I_{\{T_i > L_i\}} \mid L_i\right\}\right] \\
&= \int_0^\infty uE(I_{\{u \leq L_i\}})dF(u) + \int_0^\infty E(T_i | T_i > s)E(I_{\{T_i > s\}})dH(s) \\
&= \int_0^\infty u[1 - H(u)]dF(u) + \int_0^\infty \int_0^\infty ud\left[\frac{P\{T_i \leq u; T_i > s\}}{P\{T_i > s\}}\right][1 - F(s)]\,dH(s) \\
&= \int_0^\infty u[1 - H(u)]dF(u) + \int_0^\infty \int_s^\infty udF(u)dH(s) \\
&= \int_0^\infty u[1 - H(u)]dF(u) + \int_0^\infty H(u)dF(u) = E(T_i) \\
&= \alpha + \mathbf{x}_i\boldsymbol{\beta}.
\end{aligned}$$

Supondo $p = 1$, o uso de t_i^* na equação (2.15) faz com que $E\{\sum(t_i^* - \beta x_i)(x_i - \overline{x})\} = \sum\{(\alpha + \beta x_i - \beta x_i)(x_i - \overline{x})\} = 0$. Portanto um procedimento para obter estimativas para β é encontrar um valor b tal que $\sum(t_i^* - x_i b)(x_i - \overline{x})$ atinja o mínimo, ou no caso p-variado, encontrar um vetor $\mathbf{b}$ que minimiza $[\mathbf{t} - \mathbf{X}\mathbf{b}]'[\mathbf{X} - \overline{\mathbf{X}}]\mathbf{1}_p$.

Para $\delta_i = 0$, t_i^* é desconhecido e precisa ser estimado. Considerando $\epsilon_i = T_i - \alpha - \mathbf{x}_i\boldsymbol{\beta}$, vem que

$$\begin{aligned} E\{T_i \mid T_i > L_i\} &= E\{\alpha + \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i \mid \alpha + \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i > L_i\} \\ &= \alpha + \mathbf{x}_i\boldsymbol{\beta} + E\{\epsilon_i \mid \epsilon_i > L_i - \alpha - \mathbf{x}_i\boldsymbol{\beta}\}. \end{aligned}$$

Mas,

$$\begin{aligned} E\{\epsilon_i | \epsilon_i > L_i - \alpha - \mathbf{x}_i\boldsymbol{\beta}\} &= \int_0^\infty e\, d\left[\frac{P\{\epsilon_i \leq e; \epsilon_i > L_i - \alpha - \mathbf{x}_i\boldsymbol{\beta}\}}{P\{\epsilon_i > L_i - \alpha - \mathbf{x}_i\boldsymbol{\beta}\}}\right] \\ = \int_{L_i-\alpha-\mathbf{x}_i\boldsymbol{\beta}}^\infty e\, d\left[\frac{P\{\epsilon_i \leq e\}}{P\{\epsilon_i > L_i - \alpha - \mathbf{x}_i\boldsymbol{\beta}\}}\right] &= \int_{L_i-\alpha-\mathbf{x}_i\boldsymbol{\beta}}^\infty e\frac{dF(e)}{1 - F(L_i - \alpha - \mathbf{x}_i\boldsymbol{\beta})}. \end{aligned}$$

Portanto,

$$E\{T_i \mid T_i > L_i\} = \alpha + \mathbf{x}_i\boldsymbol{\beta} + \frac{\int_{L_i-\alpha-\mathbf{x}_i\boldsymbol{\beta}}^\infty e\, dF(e)}{1 - F(L_i - \alpha - \mathbf{x}_i\boldsymbol{\beta})}.$$

Um estimador para a expressão acima é obtido substituindo-se as quantidades desconhecidas por estimadores. Uma vez que $F(\cdot)$ é desconhecida, utiliza-se o estimador produto-limite, de modo que

$$\widehat{E}\{T_i \mid T_i > L_i\} = \widehat{\alpha} + \mathbf{x}_i\widehat{\boldsymbol{\beta}} + \frac{\int_{L_i-\widehat{\alpha}-\mathbf{x}_i\widehat{\boldsymbol{\beta}}}^\infty e d\widehat{F}^{PL}(e)}{1 - \widehat{F}^{PL}(L_i - \widehat{\alpha} - \mathbf{x}_i\widehat{\boldsymbol{\beta}})}.$$

Escrevendo em termos de somatória, no lugar de $d\widehat{F}^{PL}(\cdot)$ são utilizados os pesos $w_i(\cdot)$ definidos na seção anterior . Agora estes pesos são calculados reclassificando-se a maior observação como não censurada quando for censurada. Como α não altera a ordem das observações, fazendo $r_i = t_i - \mathbf{x}_i\widehat{\boldsymbol{\beta}}$ segue

$$\begin{aligned} \widehat{E}\{T_i \mid T_i > L_i\} =& \widehat{\alpha} + \mathbf{x}_i\widehat{\boldsymbol{\beta}} + \frac{\sum_{k:r_k>r_i}(t_k - \widehat{\alpha} - \mathbf{x}_k\widehat{\boldsymbol{\beta}})w_k(\widehat{\boldsymbol{\beta}})}{1 - \widehat{F}^{PL}(r_i)} \\ =& \mathbf{x}_i\widehat{\boldsymbol{\beta}} + \sum_{k:r_k>r_i} r_k v_{ki}(\widehat{\boldsymbol{\beta}}), \end{aligned}$$

onde $v_{ki}(\widehat{\boldsymbol{\beta}}) = w_k(\widehat{\boldsymbol{\beta}})/(1 - \widehat{F}^{PL}(r_i))$. Desta forma, pode-se escrever $\widehat{t}_i^*(\widehat{\boldsymbol{\beta}}) = T_i\delta_i + \widehat{E}(T_i \mid T_i > L_i)(1 - \delta_i)$, e o estimador de Buckley-James para $\boldsymbol{\beta}$, denotado por $\widehat{\boldsymbol{\beta}}^{BJ}$ será dado por

$$\widehat{\boldsymbol{\beta}}^{BJ} = \left[(\mathbf{X} - \overline{\mathbf{X}})'(\mathbf{X} - \overline{\mathbf{X}})\right]^{-1}(\mathbf{X} - \overline{\mathbf{X}})'\widehat{t}^*(\widehat{\boldsymbol{\beta}}_0), \tag{2.17}$$

onde $\widehat{t}^*(\cdot) = \left(\widehat{t}_1^*(\cdot), \ldots, \widehat{t}_n^*(\cdot)\right)'$ e $\widehat{\boldsymbol{\beta}}_0$ é uma estimativa inicial para $\boldsymbol{\beta}$. Buckley e James(1979) sugerem tomar como valor inicial aquele atribuído ao estimador de mínimos quadrados

considerando todos os dados como não-censurados. Desta forma, a cada iteração utiliza-se o valor de $\widehat{\boldsymbol{\beta}}^{BJ}$ obtido na iteração anterior no lugar de $\widehat{\boldsymbol{\beta}}_0$ na expressão (2.17) até a obtenção da convergência.

Uma vez obtido $\widehat{\boldsymbol{\beta}}^{BJ}$, a estimativa de α será dada por

$$\widehat{\alpha}^{BJ} = \left(\frac{1}{n}\right) \mathbf{1}' \left[\widehat{t}^{*}(\widehat{\boldsymbol{\beta}}^{BJ}) - \mathbf{X}\widehat{\boldsymbol{\beta}}^{BJ}\right],$$

que devido às censuras será um estimador viciado para α.

Através de amostras simuladas, Buckley e James(1979) observam que, a exemplo dos estimadores de Miller, os valores obtidos no processo iterativo podem permanecer sempre sobre uma mesma sequência de números, não ocorrendo a convergência. No entanto, em todas as situações onde isto ocorreu, a dispersão dos valores observados foi menor do que aquela observada para os estimadores de Miller. Em tais casos também é sugerido tomar-se a média dos valores como estimativa.

As propriedades assintóticas de $\widehat{\boldsymbol{\beta}}^{BJ}$ são de difícil obtenção devido a possibilidade de não unicidade na solução de (2.17). Entretanto, baseando-se em simulações, conjectura-se que os estimadores de Buckley–James devem ser consistentes e assintoticamente obedecem à distribuição normal.

Algum trabalho foi feito, heuristicamente, para verificação de algumas propriedades. De modo a simplificar a notação, considere o caso em que $p = 1$. Então de (2.17) segue que

$$\begin{aligned}
\widehat{\beta}^{BJ} &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})[t_i\delta_i + \widehat{E}\{T_i \mid T_i > L_i\}(1-\delta_i)]}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \\
&= \frac{\sum_{i=1}^{n}(x_i - \overline{x})\left\{t_i\delta_i + \left(\widehat{\beta}^{BJ}x_i + \sum_{k:r_k > r_i} v_{ki}(\widehat{\beta}^{BJ})r_k\right)(1-\delta_i)\right\}}{\sum_{i=1}^{n}(x_i - \overline{x})^2}.
\end{aligned} \tag{2.18}$$

Não se sabe sob que condições $\widehat{E}(T_i|T_i > L_i)$ estima consistentemente $E(T_i|T_i > L_i)$. Supondo que isto ocorra e escrevendo $\widehat{E}(T_i|T_i > L_i) = E(T_i|T_i > L_i) + u_i$, tem-se que

$$\begin{aligned}
E\left\{\widehat{\beta}^{BJ}\right\} &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})E\left\{t_i\delta_i + E(T_i|T_I > L_i)(1-\delta_i)\right\}}{\sum_{i=1}^{n}(x_i - \overline{x})^2} + E\left\{\frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i(1-\delta_i)}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right\} \\
&= \beta + E\left\{\frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i(1-\delta_i)}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right\}.
\end{aligned} \tag{2.19}$$

Portanto, $\widehat{\beta}^{BJ}$ será não viciado para β se o segundo termo em (2.19) puder ser desprezado.

Trabalhando a expressão (2.18) convenientemente, pode-se escrever que

$$\widehat{\beta}^{BJ} = \frac{\sum_{k=1}^{n} t_k \left\{(x_k - \overline{x})\delta_k + \sum_{i:r_i<r_k} v_{ki}(\widehat{\beta}_1^{BJ})(x_i - \overline{x})(1-\delta_i)\right\}}{\sum_{i=1}^{n}\left\{(x_i-\overline{x})^2 - (x_i-\overline{x})(x_i - \tilde{x}_i(\widehat{\beta}_1^{BJ}))(1-\delta_i)\right\}}, \tag{2.20}$$

com $\tilde{x}_i(\cdot) = \sum_{k:r_k>r_i} v_{ki}(\cdot)x_k$. Para se obter uma estimativa para a variância de $\widehat{\beta}^{BJ}$ utiliza-se a relação $Var\{\widehat{\beta}^{BJ}\}=E\{Var(\widehat{\beta}^{BJ}|\boldsymbol{\delta})\} + Var\{E(\widehat{\beta}^{BJ}|\boldsymbol{\delta})\}$ onde $\boldsymbol{\delta}' = (\delta_1, \ldots, \delta_n)$. Assumindo que o estimador é não viciado, o segundo termo da soma é zero e se a variação dos pesos $v_{ki}(\cdot)$ puder ser desprezada, segue de (2.20) que

$$\widehat{Var}(\widehat{\beta}^{BJ}) = \sum_{i=1}^{n} Var\{t_i\}\left\{\frac{(x_i-\overline{x})\delta_i + \sum_{k:r_k<r_i} v_{ik}(\widehat{\beta}^{BJ})(x_i-\overline{x})(1-\delta_i)}{\sum_{k=1}^{n}\{(x_k-\overline{x})^2 - (x_k-\overline{x})(x_k-\tilde{x}_k(\widehat{\beta}^{BJ}))(1-\delta_i)}\right\}^2. \tag{2.21}$$

Buckley e James(1979) verificam que nas amostras simuladas, (2.21) se comporta de maneira muito instável e propõem, sem justificativa matemática, estimar a variância por

$$\widehat{Var}(\widehat{\beta}^{BJ}) = \frac{\sum_{k:\delta_k=1}\widehat{Var}\{t_k\}(x_k-\overline{x}_{NC})^2}{\left\{\sum_{k:\delta_k=1}(x_k-\overline{x}_{NC})^2\right\}^2},$$

onde $\widehat{Var}\{t_k\} = \sum_{k:\delta_k=1}(r_k - \overline{r}_{NC})^2/(n_{NC}-2)$, com $\overline{x}_{NC} = \sum_{i=1}^{n}\delta_i x_i/n_{NC}$, $\overline{r}_{NC} = \sum_{i=1}^{n}\delta_i r_i/n_{NC}$ e n_{NC} representa o número de observações não censuradas.

2.3. Método de Cox

A teoria desenvolvida por Cox(1972) assume para os dados um modelo de *riscos proporcionais* onde a função de risco, dada por $\lambda(t;\mathbf{x})$, pode ser escrita como

$$\lambda_0(t)\exp\{\mathbf{x}\boldsymbol{\beta}\}, \tag{2.22}$$

onde $\lambda_0(t)$ é uma função desconhecida que corresponde à função de risco de um indivíduo para o qual $\mathbf{x} = \mathbf{0}$. Baseando-se no fato de que $\lambda(t;\mathbf{x}) = f(t;\mathbf{x})/(1-F(t;\mathbf{x})) = \lim_{\Delta t\to 0^+} P\{t < T \leq t+\Delta t | T \geq t\}/\Delta t$ segue que

$$P\{t < T \leq t+\Delta t | T \geq t\} \cong \lambda(t;\mathbf{x})\Delta t. \tag{2.23}$$

Assumindo para $\lambda_0(t)$ uma função arbitrária, Cox(1972) comenta que possivelmente esta função pode assumir o valor zero em intervalos onde não ocorrem mortes. Desta forma convém considerar apenas a situação sobre o conjunto dos instantes em que

ocorrem falhas, ou seja, é conveniente trabalhar com distribuições condicionais. Considere a situação em que não ocorrem empates, isto é, serão observados n instantes $t_1, \ldots, t_n$ que são ordenados e representados por $t_{(1)} < \ldots < t_{(n)}$ acompanhados dos respectivos indicadores de censura $\delta_{(i)}$ e covariadas $\mathbf{x}_{(i)}$, $i = 1, \ldots, n$.

Seja o conjunto de indivíduos em risco (isto é, vivos e não censurados) no instante t, representado por $\Re_t = \{j : t_j \geq t\}$. Considerando-se apenas os indivíduos em risco, então, para cada tempo $t_{(i)}$ não censurado

$$P_{\Re}\{\text{ uma morte em } (t_{(i)}; t_{(i)} + \Delta t)\} \cong \sum_{j \in \Re_{t_{(i)}}} \lambda(t_{(i)}; \mathbf{x}_j)\Delta t = \sum_{j \in \Re_{t_{(i)}}} \lambda_0(t_{(i)}) \exp\{\mathbf{x}_j \boldsymbol{\beta}\}\Delta t.$$

Portando, a probabilidade de ocorrer a morte do indivíduo (i) no instante $t_{(i)}$ dado que um indivíduo de $\Re_{t(i)}$ morre é dada por

$$\frac{\lambda(t_{(i)}; \mathbf{x}_{(i)})}{\sum_{j \in \Re_{t_{(i)}}} \lambda(t_{(i)}; \mathbf{x}_{(j)})} = \frac{\lambda_0(t_{(i)}) \exp\{\mathbf{x}_{(i)}\boldsymbol{\beta}\}}{\sum_{j \in \Re_{t_{(i)}}} \lambda_0(t_{(i)}) \exp\{\mathbf{x}_{(j)}\boldsymbol{\beta}\}} = \frac{\exp\{\mathbf{x}_{(i)}\boldsymbol{\beta}\}}{\sum_{j \in \Re_{t_{(i)}}} \exp\{\mathbf{x}_{(j)}\boldsymbol{\beta}\}}. \qquad (2.24)$$

Cada instante de falha contribui como um fator para a chamada *verossimilhança condicional* dada por

$$\prod_{i:\delta_{(i)}=1} \frac{\exp\{\mathbf{x}_{(i)}\boldsymbol{\beta}\}}{\sum_{j \in \Re_{t_{(i)}}} \exp\{\mathbf{x}_j \boldsymbol{\beta}\}}. \qquad (2.25)$$

Tratando (2.25) como uma verossimilhança usual, os estimadores de Cox para $\boldsymbol{\beta}$, denotados por $\widehat{\boldsymbol{\beta}}^c$ são obtidos pelo método de máxima verossimilhança a partir da maximização do logaritmo da verossimilhança, dado por

$$L(\boldsymbol{\beta}) = \sum_{i:\delta_{(i)}=1} \mathbf{x}_{(i)}\boldsymbol{\beta} - \sum_{j:\delta_{(j)}=1} \log \left\{ \sum_{j \in \Re_{t_{(i)}}} \exp\{\mathbf{x}_{(j)}\boldsymbol{\beta}\} \right\}.$$

As equações normais serão dadas por

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_i} = 0 \qquad i = 1, \ldots, p, \qquad (2.26)$$

e têm que ser resolvidas iterativamente. Trabalhando-se com o método de Newton-Raphson (seção 3.1 - capítulo 1) segue a equação

$$\widehat{\boldsymbol{\beta}}^c = \widehat{\boldsymbol{\beta}}^0 - \mathbf{V}^{-1}(\widehat{\boldsymbol{\beta}}^0)\mathbf{U}(\widehat{\boldsymbol{\beta}}^0), \qquad (2.27)$$

onde $\widehat{\boldsymbol{\beta}}^0$ é uma estimativa inicial, $\mathbf{U}(\widehat{\boldsymbol{\beta}}^0)$ é um vetor coluna de dimensão p cujos elementos são dados pelas derivadas (2.26) calculadas em $\widehat{\boldsymbol{\beta}}^0$ e $\mathbf{V}(\widehat{\boldsymbol{\beta}^0})$ é uma matriz quadrada de dimensão p com elementos dados por

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial\beta_i\partial\beta_j}, \qquad i,j=1,\ldots,p, \tag{2.28}$$

calculados em $\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^0$. No caso em que $p=1$,

$$U(\widehat{\beta}^0)=\sum_{i:\delta_{(i)}=1} x_{(i)} - \sum_{i:\delta_{(i)}=1} \frac{\sum_{j\in\Re_{t_{(i)}}} x_{(j)}\exp\{x_{(j)}\widehat{\beta}^0\}}{\sum_{j\in\Re_{t_{(i)}}}\exp\{x_{(j)}\widehat{\beta}^0\}}$$

e

$$V(\widehat{\beta}^0)=\sum_{i:\delta_{(i)}=1}\left\{\frac{\sum_{j\in\Re_{t_{(i)}}} x_{(j)}^2\exp\{x_{(j)}\widehat{\beta}^0\}}{\sum_{j\in\Re_{t_{(i)}}}\exp\{x_{(j)}\widehat{\beta}^0\}} - \left(\frac{\sum_{j\in\Re_{t_{(i)}}} x_{(j)}\exp\{x_{(j)}\widehat{\beta}^0\}}{\sum_{j\in\Re_{t_{(i)}}}\exp\{x_{(j)}\widehat{\beta}^0\}}\right)^2\right\}$$

Das relações $S(t;\mathbf{x})=1-F(t;\mathbf{x})=\exp\{\int_0^t\lambda(u;\mathbf{x})du\}$, $f(t;\mathbf{x})=\lambda(t;\mathbf{x})S(t;\mathbf{x})$ e da expressão (3.3) do capítulo 1, segue que a verossimilhança completa pode ser escrita como

$$\begin{aligned}
&L(\boldsymbol{\beta},\lambda_0(\cdot))=\\
&=\prod_{i=1}^n\left[\lambda(t_{(i)};\mathbf{x}_{(i)})\exp\left\{-\int_0^{t_i}\lambda(u;\mathbf{x}_{(i)})du\right\}\right]^{\delta_{(i)}}\left[\exp\left\{-\int_0^{t_{(i)}}\lambda(u;\mathbf{x}_{(i)})du\right\}\right]^{1-\delta_{(i)}}\\
&=\prod_{i=1}^n\left[\lambda_0(t_{(i)})\epsilon^{\mathbf{x}_{(i)}\boldsymbol{\beta}}\exp\left\{-\epsilon^{\mathbf{x}_{(i)}\boldsymbol{\beta}}\int_0^{t_{(i)}}\lambda_0(u)du\right\}\right]^{\delta_{(i)}}\\
&\qquad\times\left[\exp\left\{-\epsilon^{\mathbf{x}_{(i)}\boldsymbol{\beta}}\int_0^{t_{(i)}}\lambda_0(u)du\right\}\right]^{1-\delta_{(i)}}\\
&=\prod_{i=1}^n\left[\frac{\epsilon^{\mathbf{x}_{(i)}\boldsymbol{\beta}}}{\sum_{j\in\Re_{t_{(i)}}}\epsilon^{\mathbf{x}_j\boldsymbol{\beta}}}\right]^{\delta_{(i)}}\prod_{i=1}^n\left[\lambda_0(t)\sum_{j\in\Re_{t_{(i)}}}\epsilon^{\mathbf{x}_j\boldsymbol{\beta}}\right]^{\delta_{(i)}}\prod_{i=1}^n\left[\exp\left\{-\epsilon^{\mathbf{x}_{(i)}\boldsymbol{\beta}}\int_0^{t_{(i)}}\lambda_0(u)du\right\}\right].
\end{aligned}$$

Portanto, a verossimilhança condicional dada por (2.25) é parte da verossimilhança completa e é tratada por Cox(1975) da seguinte forma: Defina a variável aleatória S_i especificando o particular indivíduo que falha em $t_{(i)}$ e a variável aleatória C_i denotando a censura no intervalo $[t_{(i-1)};t_{(i)})$ mais a informação de que a primeira falha nesse intervalo

ocorre no instante $t_{(i)}$. Então, simbólicamente, a verossimilhança completa pode ser escrita como

$$\begin{aligned} & f_{C_1,S_1,\ldots,C_n,S_n} = \\ & = f_{S_n|C_1,S_1,\ldots,C_{n-1},S_{n-1},C_n} \times f_{C_n|C_1,S_1,\ldots,C_{n-1},S_{n-1}} \times \cdots \times \cdots \times f_{C_2|C_1,S_1} \times f_{S_1|C_1} \times f_{C_1} \\ & = \prod_{i=1}^{n} f_{S_i|\mathbf{S}^{(i-1)},\mathbf{C}^{(i)}} \times \prod_{i=1}^{n} f_{C_i|\mathbf{S}^{(i)},\mathbf{C}^{(i)}}, \end{aligned} \tag{2.29}$$

onde $\mathbf{S}^{(i)} = (S_1, \ldots, S_i)$ e $\mathbf{C}^{(i)} = (C_1, \ldots, C_i)$. Desta forma, resta notar que o primeiro produtório em (2.29), denominado por Cox(1975) de *verossimilhança parcial*, corresponde à verossimilhança condicional dada em (2.25). Escrevendo a verossimilhança completa de maneira conveniente, Efron(1977) interpreta os componentes desprezados em (2.29) e um estudo detalhado sobre a perda de informação decorrente da utilização de parte da verossimilhança e em particular da verossimilhança parcial, pode ser encontrado em Mariotto(1983). As propriedades assintóticas podem então ser determinadas heuristicamente, iniciando-se com o desenvolvimento de $\mathbf{U}(\boldsymbol{\beta})$ em série de Taylor que fornece a igualdade

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}^c - \boldsymbol{\beta}^0) = n^{\frac{1}{2}}\mathbf{V}^{-1}(\boldsymbol{\beta}^*)\mathbf{U}(\boldsymbol{\beta}^0), \tag{2.30}$$

onde $\boldsymbol{\beta}^0$ é o "verdadeiro valor" de $\boldsymbol{\beta}$ e, $\boldsymbol{\beta}^*$ está entre $\widehat{\boldsymbol{\beta}}^c$ e $\boldsymbol{\beta}^0$. Portanto, a distribuição assintótica de $\mathbf{U}(\boldsymbol{\beta})$ fornece informação a respeito da distribuição de $\widehat{\boldsymbol{\beta}}^c$.

Considere que todo o histórico de censuras e falhas anteriores a um instante $t_{(i)}$ dada por $\mathbf{S}^{(i)}$ e $\mathbf{C}^{(i)}$ é representado pela variável aleatória H_i. Então,

$$P\{S_i \mid H_i\} = \frac{\exp\{\mathbf{x}_{(i)}\boldsymbol{\beta}^0\}}{\sum_{j \in \Re_{t_{(i)}}} \exp\{\mathbf{x}_{(j)}\boldsymbol{\beta}^0\}}. \tag{2.31}$$

O produto $\prod_{i:\delta_i=1} P\{S_i|H_i\}$ corresponde à verossimilhança parcial e a derivada do logaritmo desta verossimilhança com relação a β_i corresponde ao i-ésimo elemento do vetor $\mathbf{U}(\boldsymbol{\beta})$. Defina $\mathbf{U}_i = \mathbf{U}_i(\boldsymbol{\beta}^0)$ como sendo o vetor cujo k-ésimo elemento corresponde à derivada do logaritmo de (2.31) com relação a β_k. Segue que $\mathbf{U}(\boldsymbol{\beta}^0) = \sum_{i:\delta(i)=1} \mathbf{U}_i$.

Supondo que (2.31) satisfaz as condições de regularidade usuais (Rao, 1973, §5) segue que

$$\begin{aligned} E\{\mathbf{U}_i \mid H_i\} &= \int \frac{\partial \log P(s|H_i)}{\partial \boldsymbol{\beta}} P(s|H_i)ds = \int \frac{\partial P(s|H_i)}{\partial \boldsymbol{\beta}} ds \\ &= \frac{\partial \int P(s|H_i)ds}{\partial \boldsymbol{\beta}} = \mathbf{0}. \end{aligned} \tag{2.32}$$

Portanto, $E\{\mathbf{U}_i\} = E\{E(\mathbf{U}_i|H_i)\} = \mathbf{0}$ de modo que $E\{\mathbf{U}(\boldsymbol{\beta}^0)\} = \mathbf{0}$. Da mesma forma, $Var\{\mathbf{U}_i|H_i\}$ será representada pela matriz quadrada de dimensão p cujo elemento (i,j) será dado por

$$E\left\{-\frac{\partial^2 \log P\{S_i|H_i)}{\partial\beta_i\partial\beta_j} \mid H_i\right\}.$$

Através da relação $Var\{\mathbf{U}_i\} = E\{Var(\mathbf{U}_i|H_i)\} + Var\{E(\mathbf{U}_i|H_i)\}$ segue por (2.32) que o segundo termo é zero e o primeiro corresponde à matriz de dimensão p x p cujo elemento (i,j) é dado por

$$E\left\{E\left(-\frac{\partial^2 \log P\{S_i|H_i\}}{\partial\beta_i\partial\beta_j} \mid H_i\right)\right\} = E\left\{-\frac{\partial^2 \log P\{S_i|H_i\}}{\partial\beta_i\partial\beta_j}\right\},$$

isto é, corresponde à matriz de informação de Fisher para S_i, denotada por $\mathbf{I}_i(\boldsymbol{\beta}^0)$. Deve ser notado que $\mathbf{U}_i$ e $\mathbf{U}_j$ são não correlacionados pois supondo $i < j$ vem que

$$E\{\mathbf{U}_i\mathbf{U}_j{}'\} = E\{E(\mathbf{U}_i \mid H_i)\mathbf{U}_j'\} = \mathbf{0}. \tag{2.33}$$

Portanto, $Var\{\mathbf{U}(\boldsymbol{\beta}^0)\} = \sum_{i:\delta(i)=1} \mathbf{I}_i(\boldsymbol{\beta}^0)$. Assumindo-se que as quantidades $\mathbf{U}_i$ são independentes para $i = 1, \ldots, k$ e que os valores de $\mathbf{I}_i(\boldsymbol{\beta}^0)$ não são muito diferentes, $\mathbf{U}(\boldsymbol{\beta})$ deverá ter distribuição assintótica normal com média $\mathbf{0}$ e variância $\sum_{i:\delta(i)=1} \mathbf{I}_i(\boldsymbol{\beta}^0)$ e $n^{-1}\mathbf{V}(\boldsymbol{\beta}^*)$ deverá convergir em probabilidade para $n^{-1}\mathbf{I}_i(\boldsymbol{\beta}^0)$.

Através de (2.30) conclui-se que $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}^c - \boldsymbol{\beta}^0)$ terá distribuição assintótica normal com vetor de médias $\mathbf{0}$ e matriz de variância dada por $\mathbf{V}^{-1}(\boldsymbol{\beta}^0)$.

A consistência de $\widehat{\boldsymbol{\beta}}^c$ bem como os resultados anteriores foram provados formalmente por Tsiats(1981) inspirando-se nos desenvolvimentos feitos por Breslow e Crowley(1974) para o estimador P.L. .

3. Estimação Paramétrica

Em muitas situações é possível encontrar através dos dados observados leis probabilísticas que permitem a modelagem paramétrica, proporcionando um estudo mais estruturado para o problema em questão. A procura por esses modelos normalmente se baseia em uma abordagem inicial não paramétrica que fornece evidências de possíveis modelos para os dados (veja, por exemplo, Lawless, 1982, §6.2).

A pesquisa dos modelos paramétricos deve levar em conta que a característica de interesse é representada por uma variável aleatória não negativa. Esta restrição tem influência direta sobre a escolha da maneira pela qual se relacionam as variáveis independentes com as dependentes. Dois tipos de abordagens têm sido utilizados para tratar do problema. A primeira é a utilização do modelo de *riscos proporcionais* já mencionado na seção anterior. A segunda é a utilização de transformações, em particular a logarítmica dos tempos de vida, que fornece modelos de *locação-escala.*

Suponha que $\mathbf{x}_j$ corresponda a um vetor 1 x p contendo os valores de p variáveis independentes para o j-ésimo indivíduo. Os modelos de riscos proporcionais correspondem a uma classe de modelos com a propriedade de que diferentes indivíduos têm funções de risco que são proporcionais entre si, isto é, para dois indivíduos com vetores $\mathbf{x}_1$ e $\mathbf{x}_2$, a razão $\lambda(t;\mathbf{x}_1)/\lambda(t;\mathbf{x}_2)$ não varia com t. Conforme visto na seção anterior, isto implica que

$$\lambda(t;\mathbf{x}) = \lambda_0(t)g(\mathbf{x}),$$

onde normalmente $g(\mathbf{x}) = \exp\{\mathbf{x}\boldsymbol{\beta}\}$, porém outras funções podem ser utilizadas.

Suponha que os tempos de vida representados pela variável aleatória T são transformados em $Y = \log T$. A classe dos modelos de locação–escala corresponde àquela para qual, dado $\mathbf{x}$, a variável Y tem distribuição com um parâmetro de locação $\mu(\mathbf{x})$ que introduz as covariadas no modelo e um parâmetro de escala $\sigma > 0$, constante. Estas quantidades se relacionam através da expressão

$$Y = \mu(\mathbf{x}) + \sigma\epsilon, \tag{3.1}$$

onde ϵ tem uma distribuição conhecida, independente de $\mathbf{x}$ e normalmente $\mu(\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$.

Nos modelos de locação–escala as variáveis independentes atuam linearmente sobre Y e portanto multiplicativamente sobre T ao passo que nos modelos de riscos proporcionais $\mathbf{x}$ atua multiplicativamente sobre a função de risco para T.

3.1. Estimação de Máxima Verossimilhança

Entre os modelos mais utilizados, os únicos que pertencem às duas classes (de riscos proporcionais e de locação–escala) são os modelos exponencial e Weibull. Os outros modelos normalmente pertencem à classe de modelos de locação–escala. A fim de considerar a estimação nestes modelos, sejam n variáveis aleatórias $t_1, \ldots, t_n$ correspondendo a tempos de falha (T_i) ou instantes de censura (L_i) para o i-ésimo indivíduo, para qual também é observado indicador de censura δ_i e o vetor de covariadas $\mathbf{x}_i$ de dimensão 1 x p. Considere a transformação $Y_i = \log t_i$ e assuma o modelo (3.1) na sua forma linear com intercepto,

$$Y_i = \alpha + \mathbf{x}_i\boldsymbol{\beta} + \sigma\epsilon_i,$$

onde α é um parâmetro de locação, $\boldsymbol{\beta}$ é um vetor de dimensão 1 x p contendo os coeficientes de regressão e σ é um parâmetro de escala.

Suponha que ϵ_i tenha função de sobrevivência conhecida, dada por $G(\cdot)$ e densidade $g(\cdot)$. Para $\delta_i = 1$,

$$\begin{aligned} S(t;\mathbf{x}_i) &= P\{T_i > t\} = P\{Y_i > \log t\} \\ &= P\{\epsilon_i > \frac{\log t - \alpha - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\} = G\left(\frac{y_i - \alpha - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) \end{aligned}$$

e

$$f(t;\mathbf{x}_i) = \frac{1}{\sigma} g\left(\frac{y_i - \alpha - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right).$$

Seguindo a expressão (3.3) do capítulo 1, a verossimilhança pode ser escrita como

$$\begin{aligned} L(\boldsymbol{\theta}, \sigma) &= \prod_{i=1}^{n} \left\{\frac{1}{\sigma} g\left(\frac{y_i - \alpha - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)\right\}^{\delta_i} \left\{G\left(\frac{y_i - \alpha - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)\right\}^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left\{\frac{1}{\sigma} g(z_i)\right\}^{\delta_i} \{G(z_i)\}^{1-\delta_i}, \end{aligned} \tag{3.2}$$

onde $\boldsymbol{\theta}' = (\alpha, \boldsymbol{\beta}')$ é um vetor $(p+1)$ x 1 e $z_i = (y_i - \alpha - \mathbf{x}_i\boldsymbol{\beta})/\sigma$.

Os estimadores de máxima verossimilhança, denotados por $\widehat{\boldsymbol{\theta}}^{MV} = (\widehat{\alpha}^{MV}, \widehat{\boldsymbol{\beta}}^{MV})'$ e $\widehat{\sigma}^{MV}$, são obtidos através das equações normais. Segue que

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\theta}, \sigma)}{\partial \theta_j} &= -\sum_{i=1}^{n} \left[\delta_i \left(\frac{d \log g(z_i)}{dz_i} x_{ji}\right) + (1 - \delta_i)\left(\frac{d \log G(z_i)}{dz_i} x_{ji}\right)\right] \\ &= \sum_{i=1}^{n} x_{ji} a_i, \qquad j = 1, \ldots, p+1, \end{aligned} \tag{3.3}$$

$$\frac{\partial \log L(\boldsymbol{\theta},\sigma)}{\partial \sigma} = -\sum_{i=1}^{n}\left[\left(\frac{1}{\sigma}+\frac{d\log g(z_i)}{dz_i}\frac{z_i}{\sigma}\right)\delta_i + \frac{d\log G(z_i)}{dz_i}z_i(1-\delta_i)\right]$$

$$= \frac{1}{\sigma}\sum_{i=1}^{n}(-1+z_i a_i), \tag{3.4}$$

onde $a_i = -\frac{d\log g(z_i)}{dz_i}\delta_i - \frac{d\log G(z_i)}{dz_i}(1-\delta_i)$ e define-se $x_{1i}=1$, $i=1,\ldots,n$ (correspondente ao intercepto).

A resolução das equações normais tem de ser feita iterativamente. Novamente o método de Newton–Raphson pode ser utilizado. O vetor $\mathbf{U}(\boldsymbol{\beta},\sigma)$ terá dimensão $(p+2)$ x 1 com elementos dados por (3.3) e (3.4) e a matriz $\mathbf{V}(\boldsymbol{\beta},\sigma)$, de dimensão $(p+2)$ x $(p+2)$, correspondente à matriz de informação observada, terá como elementos

$$\frac{\partial^2 \log L(\boldsymbol{\theta},\sigma)}{\partial\theta_j\partial\theta_k} = \frac{1}{\sigma^2}\sum_{i=1}^{n}x_{ji}x_{ki}A_i, \qquad j,k=1,\ldots,p+1, \tag{3.5}$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta},\sigma)}{\partial\theta_j\partial\sigma} = \frac{1}{\sigma^2}\sum_{i=1}^{n}x_{ji}z_iA_i + \frac{1}{\sigma}\sum_{i=1}^{n}(z_ia_i-\delta_i), \qquad j=1,\ldots,p+1, \tag{3.6}$$

e

$$\frac{\partial^2 \log L(\boldsymbol{\theta},\sigma)}{\partial\sigma^2} = \frac{1}{\sigma}\sum_{i=1}^{n}(z_i^2A_i+\delta_i) + \frac{2}{\sigma}\sum_{i=1}^{n}(z_i-a_i-\delta_i), \tag{3.7}$$

onde $A_i = \frac{da_i}{dz_i}$.

A partir das expressões (3.3)—(3.7) e de valores iniciais $\widehat{\boldsymbol{\theta}}^0 = (\widehat{\theta}_1^0,\ldots,\widehat{\theta}_{p+1}^0) = (\widehat{\alpha}_1^0,\widehat{\beta}_1^0,\ldots,\widehat{\beta}_p^0)$ e $\widehat{\sigma}^0$, utilizando a expressão (3.6) do capítulo 1 tem início o processo iterativo que deve fornecer as estimativas de máxima verossimilhança.

Devido a complexidade da verossimilhança, a distribuição exata dos estimadores $\widehat{\boldsymbol{\theta}}$ e $\widehat{\sigma}$ não é conhecida e as inferências desenvolvidas baseiam-se em resultados assintóticos. Assim sendo, denotando-se a matriz de informação de Fisher como $\mathbf{I}(\boldsymbol{\theta},\sigma)$ e assumindo válidas as condições de regularidade (Rao, 1973, §5f), segue que os estimadores de máxima verossimilhança tem distribuição assintótica normal multivariada com média $(\boldsymbol{\theta},\sigma)$ e variância $\mathbf{I}(\boldsymbol{\theta},\sigma)^{-1}$. Uma discussão destes resultados é dada por Kalbfleisch e Prentice(1980), que obtêm a distribuição assintótica de $\mathbf{U}(\boldsymbol{\theta},\sigma)$ e a partir daí derivam a distribuição dos estimadores de máxima verossimilhança. Dada as complicações comumente encontradas no cálculo de $\mathbf{I}(\boldsymbol{\theta},\sigma)$ é comum utilizar-se em seu lugar a matriz de informação observada $\mathbf{V}(\boldsymbol{\theta},\sigma)$ cujo valor esperado corresponde à informação de Fisher. Portanto, assegurada a convergência de $(\widehat{\boldsymbol{\theta}}^{MV},\widehat{\sigma}^{MV})$, a matriz $\mathbf{I}(\boldsymbol{\theta},\sigma)$ pode ser substituída por $\mathbf{V}(\widehat{\boldsymbol{\theta}}^{MV},\widehat{\sigma}^{MV})$ sem

afetar as distribuições assintóticas. Outras abordagens têm sido utilizadas para se construir testes de hipóteses, utilizando por exemplo a estatística da razão de verossimilhança (Cox, 1962).

4. Estimação Não Paramétrica x Estimação Paramétrica

Os métodos de estimação abordados nas seções anteriores podem ser comparados através da eficiência dos estimadores. Para o caso sem censuras é possível obter expressões analíticas para a eficiência assintótica. Quando a amostra contém informações incompletas ou tem poucas observações, algumas indicações sobre o comportamento da eficiência são obtidas por simulações.

De modo a facilitar a notação e o cálculo das integrais envolvidas considere o caso em que $p = 1$ e $\sigma = 1$ é suposto conhecido. Então de (3.1),

$$Y_i = \alpha + \beta x_i + \epsilon_i, \qquad i = 1, \ldots, n, \tag{4.1}$$

onde α e β são parâmetros desconhecidos. Suponha também que $Var\{\epsilon_i\} < +\infty$.

Em geral o parâmetro α não é de interesse principal, portanto será considerada apenas a comparação dos estimadores de β. Para o caso sem censuras, alguns resultados assintóticos são obtidos para o estimador de mínimos quadrados com relação ao estimador de máxima verossimilhança. Com a presença de censuras, a comparação será feita entre os estimadores de Cox, máxima verossimilhança e a extensão do método de mínimos quadrados dada por Buckley e James, que fornece um estimador mais estável que a abordagem dada por Miller.

4.1. Amostras Sem Censuras

A eficiência assintótica do estimador de mínimos quadrados com relação ao estimador de máxima verossimilhança foi considerada por Cox e Hinkley(1968).

Segue como um resultado clássico da teoria de regressão linear, que o estimador de mínimos quadrados para β, dado por

$$\widetilde{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2},$$

é não viciado com distribuição assintótica normal cuja variância é dada por

$$Var\{\widetilde{\beta}\} = \frac{Var\{\epsilon_i\}}{\sum_{i=1}^{n} x_i^2}, \tag{4.2}$$

contanto que $Var\{\epsilon_i\} < +\infty$.

Por outro lado, se ϵ_i tem distribuição conhecida com densidade dada por $f(\cdot)$ para $i = 1, \ldots, n$, segue dos resultados da seção 3 que se $f(\cdot)$ satisfaz as condições de regularidade, o estimador de máxima verossimilhança para β, denotado por $\widehat{\beta}^{MV}$ tem distribuição assintótica normal. Para obter a variância assintótica, assuma sem perda de generalidade que $\sum x_i = 0$. O logaritmo da verossimilhança será dado por

$$\log L(\alpha, \beta) = \sum_{i=1}^{n} \log f(\epsilon_i) = \sum_{i=1}^{n} \log f(y_i - \alpha - \beta x_i).$$

Derivando esta expressão com relação a α e a β segue que

$$\frac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha \partial \beta} = \sum_{i=1}^{n} \frac{d^2 \log f(\epsilon_i)}{d\epsilon_i^2} \frac{\partial \epsilon_i}{\partial \alpha} \frac{\partial \epsilon_i}{\partial \beta} = \sum_{i=1}^{n} \frac{d^2 \log f(\epsilon_i)}{d\epsilon_i^2} x_i.$$

Logo,

$$E\left\{-\frac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha \partial \beta}\right\} = \sum_{i=1}^{n} x_i E\left\{-\frac{d^2 \log f(\epsilon_i)}{d\epsilon_i^2}\right\} = 0.$$

Analogamente,

$$E\left\{-\frac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha^2}\right\} = \sum_{i=1}^{n} E\left\{-\frac{d^2 \log f(\epsilon_i)}{d\epsilon_i^2}\right\} = n A_\epsilon$$

e

$$E\left\{-\frac{\partial^2 \log L(\alpha, \beta)}{\partial \beta^2}\right\} = \sum_{i=1}^{n} x_i^2 E\left\{-\frac{d^2 \log f(\epsilon_i)}{d\epsilon_i^2}\right\} = A_\epsilon \sum_{i=1}^{n} x_i^2,$$

onde

$$A_\epsilon = \int_{-\infty}^{+\infty} \left(-\frac{d^2 \log f(\epsilon_i)}{d\epsilon_i^2}\right) f(\epsilon_i) d\epsilon_i.$$

Portanto, a matriz de informação esperada para α e β será expressa por

$$\begin{pmatrix} n & 0 \\ 0 & A_\epsilon \sum_{i=1}^{n} x_i^2 \end{pmatrix},$$

de forma que a variância assintótica para $\widehat{\beta}^{MV}$ será dada por

$$Var\{\widehat{\beta}^{MV}\} = (A_\epsilon \sum_{i=1}^{n} x_i^2)^{-1}. \tag{4.3}$$

Segue de (4.2) e (4.3) que a eficiência assintótica do estimador de mínimos quadrados com relação ao estimador de máxima verossimilhança será

$$Ef_A = \frac{Var\{\widehat{\beta}^{MV}\}}{Var\{\widetilde{\beta}\}} = \frac{\{A_\epsilon \sum_{i=1}^n x_i^2)\}^{-1}}{Var\{\epsilon\}(\sum_{i=1}^n x_i^2)^{-1}} = \frac{1}{A_\epsilon Var\{\epsilon\}}. \tag{4.4}$$

Por exemplo, considere o caso em que ϵ tem distribuição log–gama, isto é,

$$f(u) = \frac{1}{\Gamma(k)} \exp\{ku - e^u\}, \quad -\infty < u < +\infty. \tag{4.5}$$

onde $\Gamma(k) = \int_0^{+\infty} x^{k-1}e^{-x}dx$ é a função gama. Para este modelo, a função geradora de momentos é $M(\theta) = E(e^{\theta\epsilon}) = \Gamma(k+\theta)/\Gamma(k)$. Derivando esta função em relação a θ pode-se mostrar que

$$Var\{\epsilon\} = \frac{d^2 \log \Gamma(k)}{dk^2} = \Psi'(k). \tag{4.6}$$

A função $\Psi'(k)$ é a função di–gama, para qual $\Psi'(1) = \pi^2/6$ e $\Psi'(k+1) = \Psi'(k) - 1/k^2$, $k > 0$ (Abramowitz e Stegun, 1965, Cap. 6). Aplicando o logaritmo a (4.5) e derivando, segue que

$$\begin{aligned} A_\epsilon &= \int_{-\infty}^{+\infty} e^u \frac{1}{\Gamma(k)} \exp\{ku - \epsilon^u\}du = \int_{-\infty}^{+\infty} \frac{1}{\Gamma(k)} \exp\{(k+1)u - \epsilon^u\}du \\ &= \frac{\Gamma(k+1)}{\Gamma(k)} = k, \end{aligned}$$

uma vez que $\Gamma(k+1) = k\Gamma(k)$. Portanto para o modelo log-gama, segue que

$$Ef_A = \frac{1}{k\Psi'(k)}. \tag{4.7}$$

Para verificar o comportamento da eficiência assintótica, Cox e Hinkley(1968) calculam (4.7) para diversos valores de k, conforme a tabela a seguir.

k	0.5	1.0	1.5	2	2.5	3	4	5	10
Ef_A	0.405	0.608	0.713	0.775	0.816	0.844	0.881	0.904	0.951

TABELA IV - Eficiência assintótica do estimador de minimos quadrados em relação ao estimador de máxima verossimilhança. Modelo com erro log–gama.

Pela tabela IV verifica-se que a perda de eficiência decresce a medida que k aumenta de valor. Nota-se que para $k \geq 5$, a distribuição log-gama é bastante próxima da distribuição normal (Bartlett e Kendall, 1946).

4.2. Amostras Censuradas

Nesta seção são comparados os estimadores de máxima verossimilhança, Buckley-James e Cox. Supõem-se que os tempos de censura correspondem a variáveis aleatórias L_i com distribuição $H(\cdot)$ independentes dos instantes de falha T_i.

A exemplo do capítulo 1, são obtidas indicações sobre o comportamento da eficiência dos estimadores, calculada em 2000 amostras simuladas sob diversas condições: com censura acentuada (50%), média (25%) e sem censura e, em amostras de tamanhos $n = 10, 20, 30, 50$ e 100. Em cada situação foi obtida uma estimativa do erro quadrático médio, dada por

$$\sum_{i=1}^{2000} \frac{(\widehat{\beta}_i - \beta)^2}{2000},$$

para cada um dos três estimadores. A razão destas quantidades forneceu estimativas das eficiências.

Foram considerados tempos de vida com distribuição exponencial (modelo de riscos proporcionais) e com distribuição gama (modelo de locação-escala). Em ambos os casos assumiu-se censura exponencial, independente dos tempos de falha. Para cada amostra os valores x_i foram obtidos de uma distribuição uniforme, independente de T_i e L_i.

4.2.1. Tempos de Vida com Distribuição Exponencial

Suponha que a variável aleatória T_i seja distribuída exponencialmente de tal forma que $E(T_i \mid x_i) = \gamma \exp\{-x_i\beta\}$. Isto é, a função densidade de probabilidade para T_i dado x_i é

$$f(t \mid x_i) = \theta_i \exp\{-\theta_i t\} \qquad , t > 0 \tag{4.8}$$

com $\theta_i = \exp\{-\alpha - \beta x_i\}$, $\alpha = \log\gamma$. A densidade (4.8) pertence à classe dos modelos de riscos proporcionais. Considerando $1 - H(t) = \exp\{-\mu_i t\}$, segue que a proporção de censura é dada por $p = P\{T_i > L_i\} = \mu_i/(\theta_i + \mu_i)$. Desta forma, uma vez fixados α e β, os valores para L_i são gerados da distribuição exponencial com

$$\mu_i = \frac{p}{1-p} \exp\{-\alpha - \beta x_i\}.$$

São então definidos os valores observados $t_i = \min\{T_i, L_i\}$ e os estimadores de Cox e Buckley-James são calculados com base nos ternos (t_i, x_i, δ_i), $i = 1, \ldots, n$.

Para os estimadores de máxima verossimilhança é utilizada a abordagem da seção 3, isto é, apesar de (4.8) corresponder à densidade de um modelo de riscos proporcionais, através da transformação $Y_i^* = \log T_i$ segue que

$$g(y_i^* \mid x_i) = \exp\{(y_i^* - \alpha - \beta x_i) - \exp(y_i^* - \alpha - \beta x_i)\}, \tag{4.9}$$

que equivale a um modelo de locação–escala $Y_i^* = \alpha + \beta x_i + z_i$ onde z_i tem distribuição valor extremo padrão com densidade por $g(z_i) = \exp\{z_i - e^{z_i}\}$, $-\infty < z_i < +\infty$.

Assim sendo, definindo $Y_i = \log t_i$, utiliza-se o método de máxima verossimilhança aplicado à verossimilhança (3.2) com $g(y_i \mid x_i)$ dado por (4.9).

Uma vez obtidas as estimativas de máxima verossimilhança, de Cox e de Buckley–James para as 2000 amostras, foram calculadas as estimativas das eficiências dadas por

$$\widehat{Ef}_1 = \frac{\widehat{EQM}(\hat{\beta}^{MV})}{\widehat{EQM}(\hat{\beta}^{BJ})}, \quad \widehat{Ef}_2 = \frac{\widehat{EQM}(\hat{\beta}^{MV})}{\widehat{EQM}(\hat{\beta}^{c})} \quad \text{e} \quad \widehat{Ef}_3 = \frac{\widehat{EQM}(\hat{\beta}^{BJ})}{\widehat{EQM}(\hat{\beta}^{c})}, \tag{4.10}$$

onde $\widehat{Ef}_1$ e $\widehat{Ef}_2$ correspondem às estimativas da eficiências dos estimadores de Buckley–James e de Cox com relação ao estimador de máxima verossimilhança, enquanto $\widehat{Ef}_3$ é a estimativa da eficiência do estimador de Cox com relação ao estimador de Buckley–James.

		Eficiência Estimada		
p	n	$\widehat{Ef}_1$	$\widehat{Ef}_2$	$\widehat{Ef}_3$
0,50	10	1.1049	0,4714	0,4266
	20	0,6643	0,8257	1,2429
	30	0,5666	0,8306	1,4658
	50	0,5801	0,9055	1,5611
	100	0,5817	0,9539	1,6398
0,25	10	0,6899	0,5872	0,8511
	20	0,5598	0,7937	1,3052
	30	0,5977	0,8037	1,3447
	50	0,5916	0,8903	1,5047
	100	0,6432	0,9431	1,4664
0	10	0,7465	0,6014	0,8056
	20	0,6605	0,7812	1,1826
	30	0,6655	0,8073	1,2131
	50	0,6100	0,8799	1,4424
	100	0,6581	0,9263	1,4076

TABELA V - Eficiência dos estimadores de Cox,
Buckley-James e de máxima verossimilhança.
Tempos de vida Exponenciais.

É importante ressaltar que no caso em que a amostra não contém censuras, o estimador de Buckley–James é obtido em uma única iteração, uma vez que corresponde ao estimador de mínimos quadrados. Os resultados obtidos estão na tabela V.

4.2.2. Tempos de Vida com Distribuição Gama

Com a finalidade de estudar o comportamento dos estimadores em um modelo que não seja de riscos proporcionais, considere que os tempos de vida T_i seguem a distribuição gama com densidade

$$f(t \mid x_i) = \frac{\theta_i^k}{\Gamma(k)} t^{k-1} e^{-\theta_i t}, \qquad k > 0, \quad \theta_i > 0, \quad t > 0. \tag{4.11}$$

As covariadas se relacionam com os instantes de falha através de $\theta_i = \exp\{-\alpha - \beta x_i\}$. Para efeitos de simulação será considerado o caso em que $k = 2$, que fornece um modelo conhecido como *two-hit* e que pode ser interpretado como a distribuição do instante de ocorrência do segundo evento em um processo de Poisson. Se os instantes de censura seguem um distribuição exponencial, a questão que se coloca é qual deve ser o valor de μ de tal forma que para α e β dados, a proporção de censurados seja igual a p.

Não é difícil verificar que para $k = 2$ a função de sobrevivência para (4.11) é dada por

$$S(t \mid x_i) = \theta_i t \epsilon^{-\theta_i t} + \epsilon^{-\theta_i t}. \tag{4.12}$$

Desta forma, se L_i representa o tempo de censura para o i-ésimo indivíduo, segue que

$$\begin{aligned} p = P\{T_i > L_i\} &= E[P\{T_i > L_i \mid L_i\}] = \int_0^{+\infty} P\{T_i > u\} \mu_i \epsilon^{-\mu_i u} du \\ &= \int_0^{+\infty} \theta_i \mu_i u \epsilon^{-(\theta_i + \mu_i)u} du + \int_0^{+\infty} \mu_i \epsilon^{-(\theta_i + \mu_i)u} du \\ &= \frac{\theta_i \mu_i}{(\theta_i + \mu_i)^2} + \frac{\mu_i}{\theta_i + \mu_i}. \end{aligned}$$

Trabalhando esta expressão chega-se à equação $(1-p)\mu_i^2 + 2\theta_i(1-p)\mu_i - \theta_i^2 p = 0$, cuja raiz positiva é dada por

$$\mu_i = \theta_i \left(\frac{1}{\sqrt{1-p}} - 1 \right) = e^{-\alpha - \beta x_i} \left(\frac{1}{\sqrt{1-p}} - 1 \right). \tag{4.13}$$

Os instantes de censura são então gerados considerando-se (4.13) e os ternos (t_i, x_i, δ_i) podem ser obtidos, sobre os quais são calculados os estimadores de Cox e Buckley–James.

Para obter as estimativas de máxima verossimilhança, considere novamente a variável transformada $Y_i^* = \log T_i$. Segue de (4.11) que a densidade de Y_i^* é dada por

$$g(y_i^* \mid x_i) = \frac{1}{\Gamma(k)} \exp\left\{k(y_i^* - \alpha - \beta x_i) - \exp\{y_i^* - \alpha - \beta x_i\}\right\}, \qquad (4.14)$$

que é equivalente ao modelo de locação–escala $Y_i = \alpha + \beta x_i + z_i$, onde z_i tem distribuição log-gama com densidade $g(z_i) = \exp\{kz_i - e^{z_i}\}/\Gamma(k)$, $-\infty < z_i < +\infty$. Desta forma, define-se a variável aleatória $Y_i = \log t_i$ e com $g(y_i \mid x_i)$ dado por (4.14) na expressão (3.2) calculam-se as estimativas de máxima verossimilhança. Uma vez calculadas estes valores para as 2000 amostras, as estimativas das eficiências podem ser obtidas. Os resultados estão na tabela VI, com a mesma notação definida em (4.10).

		Eficiência Estimada		
p	n	$\widehat{Ef}_1$	$\widehat{Ef}_2$	$\widehat{Ef}_3$
0,50	10	1,2218	0,1163	0,0952
	20	0,8803	0,2606	0,2960
	30	0,7791	0,4416	0,5668
	50	0,8120	0,4969	0,6120
	100	0,9145	0,4722	0,5160
0,25	10	0,8923	0,2152	0,2412
	20	0,7722	0,3775	0,4888
	30	0,7720	0,3959	0,5129
	50	0,7996	0,4307	0,5386
	100	0,8388	0,3845	0,4583
0	10	0,8888	0,2421	0,2724
	20	0,8157	0,3330	0,4082
	30	0,7812	0,3458	0,4426
	50	0,7751	0,3679	0,4747
	100	0,7574	0,3163	0,4176

TABELA VI - Eficiência dos estimadores de Cox,
Buckley–James e de máxima verossimilhança.
Tempos de vida com distribuição gama - k=2, conhecido.

4.3. Comentários

De um modo geral, os estimadores não paramétricos são menos eficientes quando vistos em relação ao estimador de máxima verossimilhança, o que é esperado. Para o estimador

de Buckley–James, seu melhor desempenho ocorre para amostras pequenas e diminue conforme n aumenta. O inverso se dá para o estimador de Cox, que tem sua eficiência melhorada quando n aumenta, talvez pelo fato de considerar parte da verossimilhança em seu desenvolvimento.

Para tempos de vida distribuídos exponencialmente (Tabela V), as estimativas da eficiência do estimador de Buckley–James com relação ao estimador de máxima verossimilhança indicam que conforme aumenta o tamanho da amostra, melhor é o comportamento do estimador paramétrico. Para amostras pequenas (n=10, 20), a medida que aumenta a proporção de observações censuradas, ocorre um acréscimo na eficiência do estimador de Buckley–James, ao passo que para amostras maiores se dá o inverso, isto é, as estimativas da eficiência têm seu valor diminuído. Nota-se também que para amostras de tamanho 10 fortemente censuradas (p = 0,50) o estimador não paramétrico mostra-se um pouco mais eficiente que o estimador de máxima verossimilhança, única situação onde isto ocorre. Com relação ao estimador de Cox, nota-se que sempre é menos eficiente que o estimador de máxima verossimilhança, melhorando a medida que a amostra aumenta de tamanho. Comparado ao estimador de Buckley–James, o estimador de Cox mostra-se bem superior quando a amostra é maior que 10, independente da proporção de censuras considerada.

Para tempos de vida com distribuição gama, houve um acréscimo nos valores das estimativas das eficiências do estimador de Buckley–James quando comparado ao estimador de máxima verossimilhança. Tais eficiências se comportam de maneira semelhante ao caso anterior, aumentando de valor a medida que o tamanho de amostra diminue. Nota-se que houve uma queda significativa na eficiência do estimador de Cox, o que era esperado uma vez que este modelo não pertence à classe dos modelos de riscos proporcionais.

Na presença de censura, em ambos os modelos o estimador de Buckley–James sempre custou mais a convergir que os outros dois estimadores, sendo este fato acentuado a medida que o tamanho da amostra aumentava. Também é importante ressaltar que este foi o estimador que mais vezes não convergiu. Considerando estes dois fatores, o estimador de máxima verossimilhança sempre apresentou melhor desempenho.

Para amostras não censuradas nota-se que os resultados numéricos obtidos na simulação para tamanhos de amostra grandes se aproximam daqueles verificados teoricamente (tabela IV), tanto para o modelo exponencial (k=1) quanto para o modelo gama (k=2).

Novamente é importante ressaltar que, a exemplo dos resultados obtidos no capítulo 1, foi utilizado o modelo correto para a estimação paramétrica, o que nem sempre

ocorre na prática, particularmente para amostras de tamanho reduzido. Quando o modelo é incorretamente especificado, seguramente deve haver um acréscimo na eficiência das estimativas não paramétricas, que assim podem ser consideradas uma importante ferramenta na fase inicial da estimação.

Capítulo 3

Estimação Bayesiana em Modelos de Sobrevivência

1. Introdução

Nos capítulos anteriores foi apresentada uma abordagem clássica para alguns aspectos da estimação em modelos de análise de sobrevivência. Alternativamente, este capítulo trata de algumas questões relacionadas a modelos de sobrevivência, sob o ponto de vista Bayesiano.

Com o intuito de introduzir a notação que será utilizada, considere que $\mathbf{y} = (y_1, \ldots, y_n)$ é um vetor representando n observações associadas a uma distribuição de probabilidades indexada por p parâmetros representados pelo vetor $\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_p) \in \Theta^p$. Supõe-se que $\boldsymbol{\theta}$ tem uma densidade $\Pi(\boldsymbol{\theta})$ denominada densidade *a priori*, que representa todo o conhecimento subjetivo do pesquisador sobre o fenomeno em estudo. Uma vez que os dados $\mathbf{y}$ são observados, a maneira natural de se combinar o conhecimento inicial com a informação dos dados é através da distribuição condicional de $\boldsymbol{\theta}$ dado $\mathbf{y}$, que através do teorema de Bayes pode ser escrita como

$$\Pi(\boldsymbol{\theta}|\mathbf{y}) = c\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})\Pi(\boldsymbol{\theta}), \tag{1.1}$$

que é denominada densidade *a posteriori* de $\boldsymbol{\theta}$ onde $\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})$ é a função de verossimilhança de $\boldsymbol{\theta}$ dado o vetor $\mathbf{y}$ e $c^{-1} = \int \mathcal{L}(\boldsymbol{\theta}|\mathbf{y})\Pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ é a constante normalizadora. Nota-se que a função de verossimilhança $\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})$ representa a informação a respeito de $\boldsymbol{\theta}$ contido nos dados.

Um ponto importante é a determinação da distribuição a priori. Assumindo que pouco se sabe a respeito de $\boldsymbol{\theta}$, a escolha mais natural é a de uma priori *localmente uniforme*, isto é, aquela que não muda muito sob a região na qual a verossimilhança concentra massa e não se altera significativamente fora desse intervalo (verossimilhança dominante).

Nas seções seguintes será considerada a utilização de (1.1) em algumas situações no contexto de análise de sobrevivência. Como normalmente as integrais envolvidas no processo têm resolução analítica muito difícil, na seção 2 é descrito o método de Laplace, que possibilita a obtenção de aproximações acuradas para tais integrais. Posteriormente são discutidos a utilização de um supermodelo para dados referentes a tempos de vida, a questão da transformação dos dados para uma distribuição de valor extremo e o uso de densidades preditivas em análise de sobrevivência.

2. Método de Aproximação de Laplace

Algumas situações requerem o cálculo de integrais que não podem ser calculadas analiticamente. Pode ser o caso, por exemplo, da média a posteriori. Suponha uma função $g(\boldsymbol{\theta})$ a valores reais. A média a posteriori para esta função será dada por

$$\begin{aligned} E\{g(\boldsymbol{\theta})|\mathbf{y}\} &= \int_{\boldsymbol{\theta}\in\Theta^p} g(\boldsymbol{\theta})\Pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \\ &= \frac{\int_{\boldsymbol{\theta}\in\Theta^p} g(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})\Pi(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}\in\Theta^p} \mathcal{L}(\boldsymbol{\theta}|\mathbf{y})\Pi(\boldsymbol{\theta})d\boldsymbol{\theta}}. \end{aligned} \tag{2.1}$$

Tierney e Kadane(1986) propõem uma aproximação para (2.1) no caso em que $Pi(\boldsymbol{\theta}|\mathbf{y})$ é unimodal, baseada no método de Laplace, onde

$$\int_{\boldsymbol{\theta}\in\Theta^p} e^{-nh(\boldsymbol{\theta})}d\boldsymbol{\theta} \cong (2\pi)^{p/2}|n\mathbf{H}(\hat{\boldsymbol{\theta}})|^{-1/2}\exp\{-nh(\hat{\boldsymbol{\theta}})\},$$

onde $\hat{\boldsymbol{\theta}}$ é o vetor que maximiza $-h(\boldsymbol{\theta})$, $\mathbf{H}(\hat{\boldsymbol{\theta}})$ é a matriz Hessiana de h calculada em $\hat{\boldsymbol{\theta}}$, cujos elementos são dados por

$$\frac{\partial^2 h(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad , \qquad i,j=1,\ldots,p,$$

e $|\mathbf{H}(\cdot)|$ representa o determinante da matriz $\mathbf{H}(\cdot)$. Escrevendo (2.1) adequadamente, segue que para $g(\cdot)$ positiva,

$$\begin{aligned} E\{g(\boldsymbol{\theta})|\mathbf{y}\} &= \frac{\int_{\boldsymbol{\theta}\in\Theta^p}\exp\left[\log\left\{g(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})\Pi(\boldsymbol{\theta})\right\}\right]d\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}\in\Theta^p}\exp\left[\log\left\{\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})\Pi(\boldsymbol{\theta})\right\}\right]d\boldsymbol{\theta}} \\ &\cong \left\{\frac{|\mathbf{H}^*(\hat{\boldsymbol{\theta}})|}{|\mathbf{H}(\widetilde{\boldsymbol{\theta}})|}\right\}\frac{g(\hat{\boldsymbol{\theta}})\mathcal{L}(\hat{\boldsymbol{\theta}}|\mathbf{y})\Pi(\hat{\boldsymbol{\theta}})}{\mathcal{L}(\widetilde{\boldsymbol{\theta}}|\mathbf{y})\Pi(\widetilde{\boldsymbol{\theta}})}, \end{aligned} \tag{2.2}$$

onde $\hat{\boldsymbol{\theta}}$ é o vetor que maximiza $-h^*(\boldsymbol{\theta}) = \log\{g(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})\Pi(\boldsymbol{\theta})\}$, $\widetilde{\boldsymbol{\theta}}$ maximiza $-h(\boldsymbol{\theta}) = \log\{\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})\Pi(\boldsymbol{\theta})\}$ e $\mathbf{H}^*(\cdot)$ e $\mathbf{H}(\cdot)$ são matrizes de dimensão p x p cujos elementos $(i,j), i,j = 1,\ldots,n$ são dados respectivamente por

$$-\frac{\partial^2 h^*(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \qquad \text{e} \qquad -\frac{\partial^2 h(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}.$$

Se o vetor de parâmetros puder ser escrito como $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ e se o interesse reside apenas sobre os parâmetros em $\boldsymbol{\theta}_1 \in \Theta^q$ sendo $\boldsymbol{\theta}_2 \in \Theta^{p-q}$ parâmetros *nuisance*, podem

ser utilizados os mesmos passos dados anteriormente, de modo que a densidade marginal a posteriori para $\boldsymbol{\theta}_1$ pode ser escrita como

$$\begin{aligned}\Pi\{\boldsymbol{\theta}_1|\mathbf{y}\} &= \int_{\boldsymbol{\theta}\in\Theta^{p-q}} \Pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}_2 \propto \int_{\boldsymbol{\theta}\in\Theta^{p-q}} \mathcal{L}(\boldsymbol{\theta}|\mathbf{y})\Pi(\boldsymbol{\theta})d\boldsymbol{\theta}_2 \\ &= \int_{\boldsymbol{\theta}\in\Theta^{p-q}} \exp\left[\log\left\{\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})\Pi(\boldsymbol{\theta})\right\}\right] d\boldsymbol{\theta}_2 \\ &\cong \left\{|\mathbf{H}_{\boldsymbol{\theta}_1}(\hat{\boldsymbol{\theta}}_2)|\right\}^{-1/2} \mathcal{L}(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2|\mathbf{y})\Pi(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2), \end{aligned} \tag{2.3}$$

onde $\hat{\boldsymbol{\theta}}_2$ é o vetor que maximiza $-h_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_2) = \log\{\mathcal{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y})\Pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\}$, esta considerada como uma função de $\boldsymbol{\theta}_2$ para $\boldsymbol{\theta}_1$ fixo e $\mathbf{H}_{\boldsymbol{\theta}_1}(\hat{\boldsymbol{\theta}}_2)$ é uma matriz de dimensão q x q cujo elemento (i, j) é dado por

$$-\frac{\partial^2 h_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_2)}{\partial\theta_{i2}\partial\theta_{j2}}\Big|_{\boldsymbol{\theta}_2=\hat{\boldsymbol{\theta}}_2}.$$

As aproximações em geral são muito boas e um estudo da precisão pode ser visto em Tierney e Kadane(1986).

3. Modelo Log-Linear com Erro Gama Generalizado

A idéia básica desta seção é apresentar a utilização de um *supermodelo*, desenvolvida por Achcar e Bolfarine(1986) para estudar os tempos de sobrevivência. Entende-se por supermodelo ao modelo que generaliza diversas densidades particulares. É considerado o caso em os tempos de vida, representados pela variável aleatória T não negativa, têm distribuição gama generalizada, cuja função densidade de probabilidades é dada por

$$f(t) = \frac{\theta\lambda(\lambda t)^{\theta k-1}}{\Gamma(k)} \exp\left\{-(\lambda t)^{\theta}\right\}, \quad \lambda > 0, \theta > 0 \text{ e } k > 0. \tag{3.1}$$

Nota-se que $f(t)$ generaliza algumas distribuições, a saber,

i. a distribuição exponencial, se $\theta = k = 1$;

ii. a distribuição de Weibull com parâmetros θ e λ, se $k = 1$;

iii. a distribuição gama com parâmetros λ e k, se $\theta = 1$ e

iv. a distribuição log-normal, se $k \to +\infty$.

Analogamente ao capítulo anterior, suponha que aos tempos de vida é aplicada a transformação logarítmica, de modo que

$$Y = \log T = \mu + \sigma W,$$

onde $\mu = -\log\lambda$, $\sigma = 1/\theta$ e W tem densidade dada por

$$f(w) = \frac{1}{\Gamma(k)} \exp\{kw - e^w\}, \tag{3.2}$$

com $-\infty < w < +\infty$. A fim de se incluir uma variável independente, o modelo pode ser escrito como

$$Y = \alpha + \beta X + \sigma W, \tag{3.3}$$

onde α é um parâmetro de locação, β é o coeficiente associado à covariada X e σ é um parâmetro de escala.

Dos parâmetros envolvidos em (3.2) e (3.3), σ e k determinam uma particular distribuição para T que posteriormente pode ser utilizada para fornecer informações a respeito de α e β. Desta forma, a estratégia a ser adotada é encontrar a densidade conjunta a posteriori para k e σ, através da qual é selecionado um modelo adequado. Uma vez obtido o modelo, é possível obter a densidade marginal a posteriori para α e β utilizando, se necessário, o método de Laplace. Para tanto, suponha que uma amostra aleatória de tamanho n representada pelas variáveis aleatórias $T_1, T_2, \ldots, T_n$ é observada. Assumindo-se válido o modelo (3.1) e aplicando-se a transformação logarítmica em T_i, $i = 1, \ldots, n$, a função de verossimilhança para α, β, σ e k é dada por

$$\mathcal{L}\{\alpha, \beta, \sigma, k|\mathbf{y}\} = \prod_{i=1}^{n} \frac{1}{\sigma} f(w_i) = \left\{\frac{1}{\sigma\Gamma(k)}\right\}^n \exp\left\{k\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} e^{w_i}\right\}. \tag{3.4}$$

onde $w_i = (y_i - \alpha - \beta x_i)/\sigma$, $i = 1, \ldots, n$.

3.1. Distribuição a Posteriori para α, β, σ e k

Segue de (1.1) que a distribuição a posteriori para os parâmetros será dada por

$$\Pi(\alpha, \beta, \sigma, k|\mathbf{y}) = c\mathcal{L}(\alpha, \beta, \sigma, k|\mathbf{y})\Pi(\alpha, \beta, \sigma, k). \tag{3.5}$$

Como normalmente não é razoável assumir independência entre os parâmetros, Achcar e Bolfarine(1986), partindo da expressão $\Pi(\alpha, \beta, \sigma, k) = \Pi(\alpha, \beta, \sigma|k)\Pi_0(k)$, sugerem utilizar a priori de Jeffreys , dada por

$$\Pi(\alpha, \beta, \sigma|k) \propto \left|\mathbf{I}_k(\alpha, \beta, \sigma)\right|^{\frac{1}{2}},$$

onde $\mathbf{I}_k(\alpha, \beta, \sigma)$ é a matriz de informação de Fisher considerando k fixo, cujos elementos são:

$$E\left\{-\frac{\partial^2 \log \mathcal{L}}{\partial \alpha^2}\right\} = E\left\{\sigma^{-2}\sum_{i=1}^{n} e^{w_i}\right\} = n\sigma^{-2}k,$$

$$E\left\{-\frac{\partial^2 \log \mathcal{L}}{\partial \alpha \partial \beta}\right\} = E\left\{\sigma^{-2}\sum_{i=1}^{n} x_i e^{w_i}\right\} = \sigma^{-2}k\sum_{i=1}^{n} x_i,$$

$$E\left\{-\frac{\partial^2 \log \mathcal{L}}{\partial \alpha \partial \sigma}\right\} = E\left\{\sigma^{-2}\left(\sum_{i=1}^{n} w_i e^{w_i} + \sum_{i=1}^{n} e^{w_i} - nk\right)\right\} = n\sigma^{-2}(1 + k\Psi(k)),$$

$$E\left\{-\frac{\partial^2 \log \mathcal{L}}{\partial \beta^2}\right\} = E\left\{\sigma^{-2}\sum_{i=1}^{n} x_i^2 e^{w_i}\right\} = \sigma^{-2}k\sum_{i=1}^{n} x_i^2,$$

$$E\left\{-\frac{\partial^2 \log \mathcal{L}}{\partial \beta \partial \sigma}\right\} = E\left\{\sigma^{-2}\left(\sum_{i=1}^{n} x_i w_i e^{w_i} + \sum_{i=1}^{n} x_i e^{w_i} - k\sum_{i=1}^{n} x_i\right)\right\}$$

$$= \sigma^{-2}\sum_{i=1}^{n} x_i(1 + k\Psi(k))$$

e

$$E\left\{-\frac{\partial^2 \log \mathcal{L}}{\partial \sigma^2}\right\} = E\left\{\sigma^{-2}\left(\sum_{i=1}^{n} w_i^2 e^{w_i} + 2\sum_{i=1}^{n} w_i e^{w_i} - 2k\sum_{i=1}^{n} w_i - n\right)\right\}$$

$$= n\sigma^{-2}\{1 + k\Psi'(k+1) + k(\Psi(k+1))^2\},$$

onde $\Psi(k)$, já mencionada no capítulo 2, é a função di-gama e $\Psi'(k)$é a função tri-gama. Os resultados anteriores são obtidos utilizando-se a função geradora de momentos para W e a igualdade $\Psi(k+1) = \Psi(k) + \frac{1}{k}$ (Abramowitz e Stegun, 1965). A matriz de informação de Fisher pode então ser escrita como

$$\sigma^{-2}\begin{pmatrix} nk & k\sum_{i=1}^{n} x_i & n(1 + k\Psi(k)) \\ k\sum_{i=1}^{n} x_i & k\sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i(1 + k\Psi(k)) \\ n(1 + k\Psi(k)) & \sum_{i=1}^{n} x_i(1 + k\Psi(k)) & n\{1 + k\Psi'(k+1) + k(\Psi(k+1))^2\} \end{pmatrix},$$

cujo determinante, aplicando-se a aproximação $\Psi'(k+1) \cong \frac{1}{k+1}$, é dado por

$$\left|\mathbf{I}_k(\alpha, \beta, \sigma)\right| = \sigma^{-6}\left(1 + \frac{k}{k+1}\right)nk^2\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum)^n x_i\right)^2\right].$$

Encarando-se o parâmetro k de forma semelhante à utilizada por Box e Tiao(1973) para a variância do modelo normal, segue que ao se tomar uma densidade a priori local-

mente uniforme para $\log k$ tem-se $\Pi_0(k) \propto 1/k$. Portanto,

$$\Pi(\alpha,\beta,\sigma,k) = \Pi(\alpha,\beta,\sigma|k)\Pi_0(k) \propto \left|\mathbf{I}_k(\alpha,\beta,\sigma)\right|^{\frac{1}{2}} \frac{1}{k}$$
$$= \frac{\sqrt{n}}{\sigma^3}\sqrt{1+\frac{k}{k+1}}\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2},$$

e de (3.5) pode ser obtida a densidade a posteriori, dada por

$$\Pi(\alpha,\beta,\sigma,k|\mathbf{y}) = \frac{c\sqrt{1+\frac{k}{k+1}}\exp\{k\sum_{i=1}^{n} w_i\}}{\sigma^{n+3}\{\Gamma(k)\}^n \exp\{\sum_{i=1}^{n} e^{w_i}\}},$$
$$= \frac{c\xi(k)\exp\left\{\frac{k}{\sigma}\sum_{i=1}^{n} y_i - \frac{nk\alpha}{\sigma} - \frac{k\beta}{\sigma}\sum_{i=1}^{n} x_i\right\}}{\sigma^{n+3}\{\Gamma(k)\}^n \exp\left\{\epsilon^{-\alpha/\sigma}\sum_{i=1}^{n} e^{(y_i-\beta x_i)/\sigma}\right\}}, \tag{3.6}$$

onde $\xi(k) = (1+\frac{k}{k+1})^{1/2}$, $k>0$, $\sigma>0$, $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$ e c é a constante normalizadora.

A fim de se obter a densidade marginal a posteriori para σ e k, inicialmente integra-se (3.6) em relação a z utilizando-se a transformação $z = e^{-\alpha}$, de modo que

$$\Pi(\sigma,\beta,k|\mathbf{y}) =$$
$$= \frac{c\xi(k)\exp\{\sum_{i=1}^{n} y_i - \beta\sum_{i=1}^{n} x_i\}}{\sigma^{n+3}\{\Gamma(k)\}^n}\int_0^{+\infty} z^{\frac{nk}{\sigma}-1}\exp\left\{-z^{1/\sigma}\sum_{i=1}^{n} e^{(y_i-\beta x_i)/\sigma}\right\}dz.$$

Desde que $\int_0^{+\infty} x^{p-1}\epsilon^{-ax^q}dx = (1/q)a^{-p/q}\Gamma(p/q)$ e considerando a aproximação $\Gamma(k) \cong \sqrt{2\pi}\epsilon^{-k}k^{k-1/2}$, segue que

$$\Pi(\sigma,\beta,k|\mathbf{y}) \propto \frac{\xi(k)k^{\frac{n-1}{2}}n^{nk-1/2}\exp\left\{\frac{k}{\sigma}(\sum_{i=1}^{n} y_i - \beta\sum_{i=1}^{n} x_i)\right\}}{\sigma^{n+2}\left\{\sum_{i=1}^{n}\exp\left\{\frac{y_i-\beta x_i}{\sigma}\right\}\right\}^{kn}}, \tag{3.7}$$

onde $k>0$, $\sigma>0$ e $-\infty<\beta<+\infty$. Integrando (3.7) em relação a β, segue que a distribuição marginal a posteriori para σ e k é dada por

$$\Pi(\sigma,k|\mathbf{y}) = \frac{c\xi(k)k^{(n+1)/2}n^{nk-1/2}\exp\left\{\frac{k}{\sigma}\sum_{i=1}^{n} y_i\right\}}{\sigma^{n+2}}\int_{-\infty}^{+\infty}\frac{\exp\left\{-\frac{k\beta}{\sigma}\sum_{i=1}^{n} x_i\right\}}{\left\{\sum_{i=1}^{n}\exp\left\{\frac{y_i-\beta x_i}{\sigma}\right\}\right\}^{kn}}d\beta.$$

Fazendo $v = \epsilon^{-\beta}$, segue que

$$\Pi(\sigma,k|\mathbf{y}) = \frac{c\xi(k)k^{(n-1)/2}n^{nk-1/2}\exp\left\{\frac{k}{\sigma}\sum_{i=1}^{n} y_i\right\}}{\sigma^{n+2}}\int_0^{+\infty}\frac{v^{k/\sigma\sum_{i=1}^{n} x_i-1}}{\left\{\sum_{i=1}^{n} e^{y_i/\sigma}v^{x_i/\sigma}\right\}^{nk}}dv$$
$$= c\int_0^{+\infty} e^{-nh_{k,\sigma}(v)}dv, \tag{3.8}$$

onde $-h_{k,\sigma}(v) = (1/n)\{\log\xi(k) + (n-1)/2\log k + (nk-1/2)\log n + (k/\sigma)\sum_{i=1}^{n} y_i - (n+2)\log\sigma + ((k/\sigma)\sum_{i=1}^{n} x_i - 1)\log v - nk\log\{\sum_{i=1}^{n} e^{y_i/\sigma} v^{x_i/\sigma}\}\}$.

A integral em (3.8) pode ser aproximada através do método de Laplace. Para tanto, consideram-se as derivadas

$$\frac{dh_{k,\sigma}(v)}{dv} = \frac{1}{n}\left\{\left(\frac{k}{\sigma}\sum_{i=1}^{n} x_i - 1\right)\frac{1}{v} - \frac{nk}{\sigma}\frac{\sum_{k=1}^{n} x_i e^{y_i/\sigma} v^{x_i/\sigma-1}}{\sum_{i=1}^{n} e^{y_i/\sigma} v^{x_i/\sigma}}\right\} \tag{3.9}$$

e

$$\frac{d^2h_{k,\sigma}(v)}{dv^2} = \frac{1}{n}\left\{\left(1 - \frac{k}{\sigma}\sum_{i=1}^{n} x_i\right)\frac{1}{v^2} - \frac{nk}{\left(\sum_{i=1}^{n} e^{y_i/\sigma} v^{x_i/\sigma}\right)^2}\times\right.$$
$$\left.\times\left[\left(\sum_{i=1}^{n}\frac{x_i}{\sigma}\left(\frac{x_i}{\sigma}-1\right)e^{\frac{y_i}{\sigma}} v^{\frac{x_i}{\sigma}-2}\right)\left(\sum_{i=1}^{n} e^{\frac{y_i}{\sigma}} v^{\frac{x_i}{\sigma}}\right) - \left(\sum_{i=1}^{n}\frac{x_i}{\sigma} e^{\frac{y_i}{\sigma}} v^{\frac{x_i}{\sigma}-1}\right)^2\right]\right\}. \tag{3.10}$$

A densidade marginal a posteriori será dada por

$$\Pi(\sigma, k|\mathbf{y}) \propto \frac{\hat{H}^{\frac{1}{2}}\xi(k)k^{(n-1)/2}n^{nk-1/2}e^{k/\sigma\sum_{i=1}^{n} y_i}\hat{v}^{k/\sigma\sum_{i=1}^{n} x_i - 1}}{\sigma^{n+2}\left\{\sum_{i=1}^{n} e^{y_i/\sigma}\hat{v}^{x_i/\sigma}\right\}^{nk}}, \tag{3.11}$$

onde $\hat{H} = -\left(\frac{d^2h_{k,\sigma}(v)}{dv^2}\right)^{-1}\Big|_{v=\hat{v}}$ e $\hat{v}$ é obtido igualando-se (3.9) a zero.

3.2. Comparação de Dois Tratamentos

As expressões desenvolvidas anteriormente podem ser convenientemente escritas quando se está interessado em comparar dois tratamentos. Para tanto, suponha que os instantes $T_{11}, T_{12}, \ldots, T_{1n_1}$ representam tempos de sobrevivência de n_1 indivíduos submetidos ao tratamento 1 e $T_{21}, T_{22}, \ldots, T_{2n_2}$ correspondam a tempos de sobrevivência de n_2 indivíduos submetidos ao tratamento 2. Associe a cada indivíduo a covariada x_i tal que

$$x_i = \begin{cases} 0, & \text{se o } i\text{-ésimo indivíduo recebe tratamento 1;} \\ 1, & \text{se o } i\text{-ésimo indivíduo recebe tratamento 2.} \end{cases}$$

Suponha também que os dois grupos são independentes. Assumindo válido o modelo log-linear dado em (3.3), segue que $\hat{v}$ é obtido igualando-se a expressão (3.9) a zero. Portanto, fazendo $Y_{ij} = \log T_{ij}$, $j = 1, \ldots, n_i$, $i = 1, 2$ vem que

$$\left(\frac{kn_2}{\sigma} - 1\right)\frac{1}{\hat{v}} - \frac{nk}{\sigma}\frac{\sum_{i=1}^{n_2} e^{y_{2i}/\sigma}\hat{v}^{1/\sigma-1}}{\sum_{i=1}^{n_1} e^{y_{1i}/\sigma} + \sum_{i=1}^{n_2} e^{y_{2i}/\sigma}\hat{v}^{1/\sigma}} = 0 \quad \Rightarrow$$
$$\Rightarrow \hat{v}^{1/\sigma} = \frac{(kn_2 - \sigma)\sum_{i=1}^{n_1} T_{1i}^{1\sigma}}{(kn_2 + \sigma)\sum_{i=1}^{n_2} T_{2i}^{1/\sigma}}. \tag{3.12}$$

De (3.10) vem que

$$\hat{H} = \frac{n\sigma\sqrt{k}\hat{v}}{\sqrt{(n_2k-\sigma)(n_1k-\sigma)}},$$

de modo que a densidade marginal a posteriori para σ e k será dada por

$$\Pi(k,\sigma|\mathbf{y}) \dot{\propto} \frac{\xi(k)e^{k/\sigma\sum_{i=1}^{n} y_i}(n_1k+\sigma)^{n_1k-1/2}(n_2k-\sigma)^{n_2k-1/2}}{k^{n(k-1/2)}\sigma^{n+1}\left(\sum_{i=1}^{n_1}T_{1i}^{1/\sigma}\right)^{n_1k}\left(\sum_{i=1}^{n_2}T_{2i}^{1/\sigma}\right)^{n_2k}}, \tag{3.13}$$

onde $n = n_1 + n_2$, $k > 0$ e $\sigma > 0$ e $\dot{\propto}$ significa *aproximadamente proporcional.*

A expressão (3.13) fornece informações sobre k e σ. Dados estes valores, especifica-se um modelo para os dados através do qual podem ser feitas inferências a respeito de α e β. Portanto, escrevendo (3.4) em função destes parâmetros, vem que

$$\mathcal{L}(\alpha,\beta) \propto \exp\left\{-\frac{nk\alpha}{\sigma} - \frac{\beta k}{\sigma}\sum_{i=1}^{n}x_i - e^{\alpha/\sigma}\sum_{i=1}^{n}e^{(y_i-\beta x_i)/\sigma}\right\}.$$

Assumindo uma distribuição localmente uniforme para α e β segue que $\Pi(\alpha,\beta) \propto$ constante. Consequentemente,

$$\begin{aligned}\Pi(\alpha,\beta|\mathbf{y}) &\propto \exp\left\{-\frac{nk\alpha}{\sigma} - \frac{\beta k}{\sigma}\sum_{i=1}^{n}x_i - e^{-\alpha/\sigma}\sum_{i=1}^{n}e^{(y_i-\beta x_i/\sigma)}\right\} \\ &= \exp\left\{-\frac{kn\alpha}{\sigma} - \frac{n_2\beta k}{\sigma} - e^{-\alpha/\sigma}\left(\sum_{i=1}^{n_1}T_{1i}^{1/\sigma} + e^{-\beta/\sigma}\sum_{i=1}^{n_2}T_{2i}^{1/\sigma}\right)\right\}. \end{aligned} \tag{3.14}$$

Desta forma a densidade marginal a posteriori para β será dada por

$$\begin{aligned}\Pi(\beta|\mathbf{y}) &= \int_{-\infty}^{+\infty}\Pi(\alpha,\beta|\mathbf{y})d\alpha \\ &\propto e^{\frac{-n_2k\beta}{\sigma}}\int_{-\infty}^{+\infty}\exp\left\{-\frac{nk\alpha}{\sigma} - e^{-\alpha/\sigma}\left(\sum_{i=1}^{n_1}T_{1i}^{1/\sigma} + e^{-\beta/\sigma}\sum_{i=1}^{n_2}T_{2i}^{1/\sigma}\right)\right\}d\alpha.\end{aligned}$$

Fazendo $z = e^{-\alpha/\sigma}$ e utlizando novamente o fato de que $\int_0^{+\infty}x^{p-1}e^{-ax^q}dx = \frac{1}{p}a^{-\frac{p}{q}}\Gamma(\frac{p}{q})$, segue que

$$\Pi(\beta|\mathbf{y}) \propto \frac{e^{-n_2k\beta/\sigma}}{\left\{\sum_{i=1}^{n_1}T_{1i}^{1/\sigma} + e^{-\beta/\sigma}\sum_{i=1}^{n_2}T_{2i}^{1/\sigma}\right\}^{nk}}, \tag{3.15}$$

onde $-\infty < \beta < +\infty$.

A estimativa Bayesiana para β será o valor que maximiza (3.15) (isto é, a moda a posteriori), dada por

$$\widehat{\beta} = \sigma \log \left\{ \frac{n_1 \sum_{i=1}^{n_2} T_{2i}^{1/\sigma}}{n_2 \sum_{i=1}^{n_1} T_{1i}^{1/\sigma}} \right\}. \tag{3.16}$$

De modo análogo, é encontrada a densidade marginal a posteriori para α, de modo que

$$\Pi(\alpha|\mathbf{y}) \propto \frac{\exp\left\{ -\frac{nk\alpha}{\sigma} - e^{-\alpha/\sigma} \sum_{i=1}^{n_1} T_{1i}^{1/\sigma} \right\}}{\left\{ e^{-\alpha/\sigma} \sum_{i=1}^{n_1} T_{1i}^{1/\sigma} \right\}^{n_2 k}}, \tag{3.17}$$

cuja moda a posteriori pode ser escrita como

$$\widehat{\alpha} = \sigma \log \left\{ \frac{\sum_{i=1}^{n_1} T_{1i}^{1/\sigma}}{k n_1} \right\}. \tag{3.18}$$

3.3. Exemplo

Considere os tempos de sobrevivência para dois grupos (tratamento 1 e tratamento 2) dados em Lee(1980, pág.294)

Tratamento 1	5,0	10,0	17.0	32,0	32.0	33.0	34,0	36.0
	43.0	44.0	44.0	48.0	48.0	61.0	64.0	65.0
	65,0	66.0	67.0	68.0	82.0	85.0	90,0	92.0
	92,0	102.0	103,0	106.0	107.0	114.0	114,0	116.0
	117,0	124,0	139,0	142,0	143,0	151,0	158,0	195,0
Tratamento 2	20,9	32.2	33.2	39,4	40.0	46,8	54,3	57,3
	58.0	59,7	61.1	61,4	66.0	66,3	67,4	68.5
	69,9	72.4	73.0	73.2	88.7	89,3	91,6	93.1
	94,2	97,7	101,6	101,9	107.6	108,0	109,7	110.8
	114,1	117,5	119,2	120,3	133,0	133,8	163,3	165,1

TABELA VII - Tempos de Sobrevivência de pacientes Submetidos a Dois Tratamentos.

Para os dados da tabela, segue que $n_1 = n_2 = 40$ e $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n_1} \log T_{1i} + \sum_{i=1}^{n_2} \log T_{2i} = 341,40$. Da expressão (3.13), a densidade marginal a posteriori para k e σ é dada por

$$\Pi(k, \sigma|\mathbf{y}) \dot{\propto} \frac{\sqrt{2k+1}(40k+\sigma)^{40k-1/2}(40k-\sigma)^{40k-1/2} e^{341,40k/\sigma}}{\sqrt{k+1}\, k^{80(k-1/2)} \sigma^{81} \left(\sum_{i=1}^{40} T_{1i}^{1/\sigma}\right)^{40k} \left(\sum_{i=1}^{40} T_{2i}^{1/\sigma}\right)^{40k}}. \tag{3.19}$$

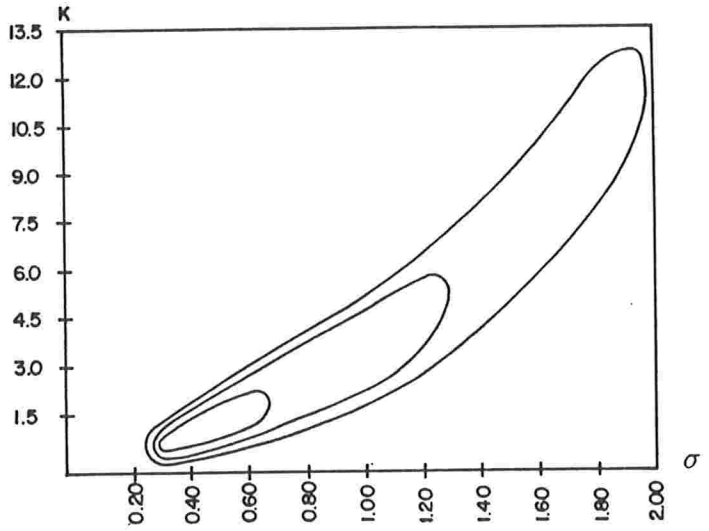

FIGURA 1 - Contornos para a Densidade a Posteriori de σ e k.

Alguns contornos de (3.19) estão na figura 1, de onde se conclui que $\widehat{k} = 1,0$ e $\widehat{\sigma} = 0,4$. Assumindo estes valores, as estimativas para β e α são dadas por (3.16) e (3.18) respectivamente, de modo que $\widehat{\alpha} = 4,578$ e $\widehat{\beta} = -0,034$. A densidade marginal a posteriori para β será dada por

$$\Pi(\beta|\mathbf{y}) \propto \frac{\exp\{-100\beta\}}{\{3736745 + 3435657e^{-2,5\beta}\}^{80}}. \tag{3.20}$$

que está representada na figura 2 a seguir.

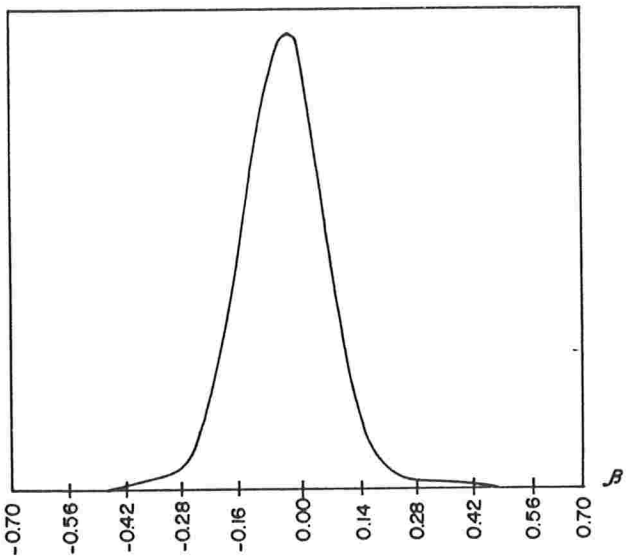

FIGURA 2 - Densidade Marginal a Posteriori para β.

Através do gráfico de (3.20) retratado na figura 2, nota-se imediatamente que o

intervalo HPD* para β deverá conter o valor zero. Portanto, conclui-se que ambos os tratamentos têm o mesmo efeito sobre o tempo de sobrevivência dos pacientes.

4. Transformação para a Distribuição de Valor Extremo

Achcar, Bolfarine e Pericchi(1987) utilizam a família de transformação potência desenvolvida por Box e Cox(1964) e definida como

$$Y^{(\lambda)} = \begin{cases} (T^{\lambda} - 1)/\lambda, & \text{se } \lambda \neq 0; \\ \log T & \text{se } \lambda = 0, \end{cases} \tag{4.1}$$

para encontrar uma transformação adequada para os tempos de sobrevivência, de tal forma que $Y^{(\lambda)}$ pode ser escrito na forma

$$Y^{(\lambda)} = \mu + \sigma W, \tag{4.2}$$

onde W tem distribuição de valor extremo com densidade dada por $f(w) = \exp\{w - e^{w}\}$, $-\infty < w < +\infty$. Note que para $\lambda = 0$ a variável T tem distribuição de Weibull. Este enfoque pode ser considerado como uma alternativa ao supermodelo da seção 3.

Inferências sobre a transformação mais adequada são baseadas na distribuição a posteriori de λ, considerando uma densidade a priori não informativa dada pela priori de Jeffreys.

4.1. Densidade a Posteriori para μ, σ e λ

Seja $T_1, T_2, \ldots, T_n$ uma amostra aletória representando os tempos de sobrevivência de n indivíduos. Assumindo (4.1) e (4.2), segue que

$$P\{T_i \leq t_i\} = P\left\{Y_i^{(\lambda)} \leq \frac{t_i^{\lambda} - 1}{\lambda}\right\} = P\left\{W_i \leq \frac{\left(\frac{t_i^{\lambda}-1}{\lambda}\right) - \mu}{\sigma}\right\}$$

$$= 1 - \exp\left\{-\exp\left(\frac{\frac{t_i^{\lambda}-1}{\lambda} - \mu}{\sigma}\right)\right\},$$

de modo que T tem densidade dada por

$$f(t_i) = \frac{t_i^{\lambda-1}}{\sigma} \exp\{w_i - e^{w_i}\} \tag{4.3}$$

---

* *Highest Posterior Density*,veja Box e Tiao(1973), §2.2

onde $w_i = (y_i^{(\lambda)} - \mu)/\sigma$ e $y_i^{(\lambda)} = (t_i^{\lambda} - 1)/\lambda$.

Segue de (4.3) que a verossimilhança para μ, σ e λ será dada por

$$\mathcal{L}(\mu,\sigma,\lambda) = \prod_{i=1}^{n} \frac{t_i^{\lambda-1}}{\sigma^n} \exp\{w_i - e^{w_i}\} = \frac{1}{\sigma^n} \prod_{i=1}^{n} t_i^{\lambda-1} \exp\left\{\sum_{i=1}^{n} w_i - \sum_{i=1}^{n} e^{w_i}\right\}$$
$$= \sigma^{-n} \prod_{i=1}^{n} t_i^{\lambda-1} \exp\left\{\frac{1}{\sigma}\sum_{i=1}^{n} y_i^{(\lambda)} - \frac{n\mu}{\sigma} - e^{-\mu/\sigma}\sum_{i=1}^{n} e^{y_i^{(\lambda)}/n}\right\}. \tag{4.4}$$

A densidade a priori para μ, σ e λ pode ser escrita como

$$\Pi(\mu,\sigma,\lambda) = \Pi(\mu,\sigma|\lambda)\Pi_0(\lambda), \tag{4.5}$$

onde $\Pi(\mu,\sigma|\lambda) \propto |\mathbf{I}_\lambda(\mu,\sigma)|^{\frac{1}{2}}$ é a priori de Jeffreys com $\mathbf{I}_\lambda(\mu,\sigma)$ representando a matriz de informação de Fisher para λ fixado, obtida a seguir.

Aplicando a transformação logarítmica a (4.4) e considerando as derivadas parciais de segunda ordem, segue que

$$E\left\{-\frac{\partial^2 \log \mathcal{L}}{\partial\mu^2}\right\} = E\left\{\sigma^{-2}\sum_{i=1}^{n} e^{w_i}\right\} = n\sigma^{-2},$$
$$E\left\{-\frac{\partial^2 \log \mathcal{L}}{\partial\mu\partial\sigma}\right\} = E\left\{\sigma^{-2}\left(n + \sum_{i=1}^{n} e^{w_i} - \sum_{i=1}^{n} w_i e^{w_i}\right)\right\} = n\sigma^{-2}\left(1 + \Psi(1)\right),$$
$$E\left\{-\frac{\partial^2 \log \mathcal{L}}{\partial\sigma^2}\right\} = E\left\{\sigma^{-2}\left(-2\sum_{i=1}^{n} w_i + 2\sum_{i=1}^{n} w_i e^{w_i} + \sum_{i=1}^{n} w_i^2 e^{w_i}\right)\right\}$$
$$= n\sigma^{-2}\left(1 + \Psi'(2) + (\Psi(2))^2\right).$$

Estes resultados podem ser obtidos de modo análogo ao da seção 3. Portanto, a matriz de informação de Fisher será dada por

$$n\sigma^{-2}\begin{pmatrix} 1 & 1+\Psi(1) \\ 1+\Psi(1) & 1+\Psi'(2)+(\Psi(2))^2 \end{pmatrix},$$

cujo determinante é dado por $n^2\sigma^{-4}\left[\left(1 + \Psi'(2) + +(\Psi(2))^2\right) - (1+\Psi(1))^2\right]$. Uma vez que $\Psi(2) = 1 + \Psi(1)$, segue que

$$|\mathbf{I}_\lambda(\mu,\sigma)| = \frac{n^2}{\sigma^4}\left\{1 + \Psi'(2)\right\}.$$

Assumindo uma priori localmente uniforme para λ, segue de (4.5) que

$$\Pi(\mu,\sigma,\lambda) \propto \frac{n}{\sigma^2}. \tag{4.6}$$

Logo a densidade conjunta a posteriori será dada por

$$\begin{aligned}\Pi(\mu,\sigma,\lambda|\mathbf{y}) &= c\mathcal{L}(\mu,\sigma,\lambda|\mathbf{y})\Pi(\mu,\sigma,\lambda)\\ &\propto \frac{1}{\sigma^{n+2}}\prod_{i=1}^{n} t_i^{\lambda-1}\exp\left\{\frac{1}{\sigma}\sum_{i=1}^{n} y_i^{(\lambda)} - \frac{n\mu}{\sigma} - e^{-\mu/\sigma}\sum_{i=1}^{n} e^{y_i^{(\lambda)}/\sigma}\right\}. \qquad (4.7)\end{aligned}$$

Integrando (4.7) em relação a μ, obtem-se a densidade marginal a posteriori para σ e λ, isto é,

$$\begin{aligned}\Pi(\sigma,\lambda|\mathbf{y}) &= \int_{-\infty}^{+\infty}\Pi(\mu,\sigma,\lambda|\mathbf{y})d\mu\\ &\propto \frac{\prod_{i=1}^{n} t_i^{\lambda-1}}{\sigma^{n+2}}\exp\left\{\frac{1}{\sigma}\sum_{i=1}^{n} y_i^{(\lambda)}\right\}\int_{-\infty}^{+\infty}\exp\left\{-\frac{n\mu}{\sigma} - e^{-\mu/\sigma}\sum_{i=1}^{n} e^{y_i^{(\lambda)}/\sigma}\right\}d\mu\\ &= \frac{\prod_{i=1}^{n} t_i^{\lambda-1}}{\sigma^{n+1}}\exp\left\{\frac{1}{\sigma}\sum_{i=1}^{n} y_i^{(\lambda)}\right\}\int_{0}^{+\infty} z^{n-1}e^{-z\sum_{i=1}^{n} e^{y_i^{(\lambda)}/\sigma}}dz\\ &\propto \frac{\left\{\prod_{i=1}^{n} t_i^{\lambda-1}\right\}\exp\left\{\frac{1}{\sigma}\sum_{i=1}^{n} y_i^{(\lambda)}\right\}}{\sigma^{n+1}\left\{\sum_{i=1}^{n} e^{y_i^{(\lambda)}/\sigma}\right\}^{n}}. \qquad (4.8)\end{aligned}$$

Através do método de Laplace e utilizando a notação da seção 2, segue que

$$\begin{aligned}\Pi(\lambda|\mathbf{y}) &\propto \int_{0}^{+\infty}\Pi(\sigma,\lambda|\mathbf{y})d\sigma\\ &= \int_{0}^{+\infty}\epsilon^{-nh_\lambda(\sigma)}d\sigma,\end{aligned}$$

onde $-h_\lambda(\sigma)=\frac{1}{n}\{(\lambda-1)\sum_{i=1}^{n}\log t_i - (n+1)\log\sigma - n\log\{\sum_{i=1}^{n}\exp(\frac{y_i^{(\lambda)}}{\sigma})\} + \frac{1}{\sigma}\sum_{i=1}^{n} y_i^{(\lambda)}\}$. O valor $\hat{\sigma}$ que maximiza $-h_\lambda(\sigma)$ é obtido através da equação $h'_\lambda(\sigma)=0$. Assim procedendo, pode ser verificado que

$$\hat{v} = \frac{n\sum_{i=1}^{n} y_i^{(\lambda)}e^{y_i^{(\lambda)}/\dot{\sigma}}}{(n+1)\sum_{i=1}^{n} e^{y_i^{(\lambda)}/\dot{\sigma}}} - \frac{\sum_{i=1}^{n} y_i^{(\lambda)}}{n+1}. \qquad (4.9)$$

Portanto, utilizando a expressão (2.3), a densidade marginal a posteriori para λ é dada por

$$\Pi(\lambda|\mathbf{y}) \dot{\propto} \frac{\hat{H}^{\frac{1}{2}}\left(\prod_{i=1}^{n} t_i^{\lambda-1}\right)\exp\left\{\frac{1}{\hat{\sigma}}\sum_{i=1}^{n} y_i^{(\lambda)}\right\}}{\hat{\sigma}\left\{\sum_{i=1}^{n} e^{\frac{y_i^{(\lambda)}}{\hat{\sigma}}}\right\}^{n}}, \qquad (4.10)$$

onde $\hat{H} = \left(\frac{d^2 h_\lambda(\sigma)}{d\sigma^2}\right)^{-1}$, calculada em $\sigma = \hat{\sigma}$.

Achcar, Bolfarine e Pericchi(1987) também consideram o desenvolvimento da densidade marginal a posteriori para λ quando a amostra contém observações censuradas, num esquema semelhante ao aqui desenvolvido.

4.2. Exemplo

Considere os tempos de sobrevivência de 21 pacientes submetidos a um dado tratamento (dados em Achcar, Bolfarine e Pericchi, 1987): 1, 1, 4, 4, 9, 16, 16, 25, 25, 64, 64, 64, 64, 121, 121, 144, 144, 225, 289, 484, 529.

A figura 3 representa a densidade marginal a posteriori para λ dada pela expressão (4.10) e avaliada numericamente. Observa-se que a moda a posteriori para λ é $\hat{\lambda} = 0$, o que implica em uma transformação logarítmica para os dados.

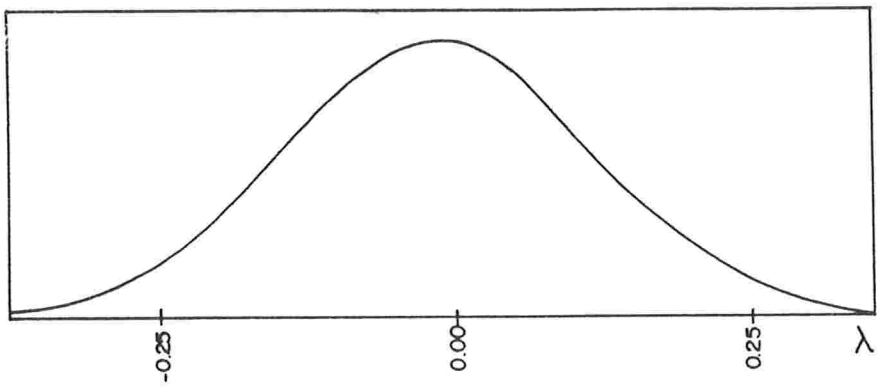

FIGURA 3 - Distribuição Marginal a Posteriori para λ.

O valor de λ também pode ser determinado empiricamente. Para tanto, assuma que os dados transformados seguem uma distribuição de valor extremo. Segue de (4.2) que

$$S(w) = P\{W > w\} = \exp\left\{-e^{(y^{(\lambda)}-\mu)/\sigma}\right\},$$

de modo que

$$\log\{-\log S(w)\} = \frac{1}{\sigma} y^{(\lambda)} - \frac{\mu}{\sigma} = \beta_0 - \beta_1 y^{(\lambda)},$$

onde $\beta_0 = -\mu/\sigma$ e $\beta_1 = 1/\sigma$.

Considerando $\widehat{S}(w)$, a função de distribuição empírica para $S(w)$ (ou o estimador produto limite, no caso censurado), segue que

$$\log\left\{-\log \widehat{S}(w)\right\} = \widehat{\beta}_0 + \widehat{\beta}_1 y^{(\lambda)},$$

onde $\widehat{\beta}_0$ e $\widehat{\beta}_1$ são os estimadores de mínimos quadrados, dados por

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(y_i^{(\lambda)} - \overline{y}^{(\lambda)}\right)(v_i - \overline{v})}{\left(y_i^{(\lambda)} - \overline{y}^{(\lambda)}\right)^2}$$

e

$$\widehat{\beta}_0 = \overline{v} - \widehat{\beta}_1 \overline{y}^{(\lambda)},$$

com $\overline{y}^{(\lambda)} = \sum_{i=1}^{n} y_i^{(\lambda)}/n$, $v_i = \log\{-\log \widehat{S}(w_i)\}$ e $\overline{v} = \sum_{i=1}^{n} v_i/n$. Uma vez encontrados os valores para $\widehat{\beta}_0$ e $\widehat{\beta}_1$, segue que $\hat{\mu}_\lambda = \widehat{\beta}_0/\widehat{\beta}_1$ e $\hat{\sigma}_\lambda = 1/\widehat{\beta}_1$.

Aplicando estes resultados aos dados, obtem-se a tabela VIII, com valores para $\hat{\mu}_\lambda$ e $\hat{\sigma}_\lambda$ calculados sob diferentes valores de λ e $L(\hat{\mu}_\lambda, \hat{\sigma}_\lambda, \lambda)$ correspondendo ao logaritmo da verossimilhança dada em (4.4).

λ	$\hat{\mu}_\lambda$	$\hat{\sigma}_\lambda$	$L(\hat{\mu}_\lambda, \hat{\sigma}_\lambda, \lambda)$
-1,0	1,04	0,35	-157,16
-0,6	1,56	0,50	-134,92
-0,2	2,90	1,00	-120,60
0	4,40	1,66	-118,08
0,2	7,42	3,27	-119,38
0,6	29,44	18,52	-130,11
1,0	167,14	142,86	-145,53

TABELA VIII - Variação do logaritmo da verossimilhança para alguns valores de λ.

Da tabela VIII conclui-se que $\lambda = 0$ maximiza a verossimilhança, confirmando o resultado obtido pelo método Bayesiano. Assim, com os dados transformados pode ser realizada uma análise semelhante à desenvolvida na seção 3.

5. Densidade Preditiva com um Modelo Gama Generalizado

A densidade preditiva pode ser usada para verificar se um modelo proposto para os dados é adequado, conforme considerado por Achcar e Bolfarine(1988). Normalmente a adequabilidade de um modelo é verificada através de análises gráficas com o uso de distribuições empíricas (veja por exemplo Lawless, 1982). No entanto algumas vezes ocorrem dúvidas com relação a possíveis modelos para os dados. Nesta seção é discutida a utilização da densidade preditiva para verificação do modelo de regressão com erro gama generalizado. Supõe-se que $\mathbf{y} = (y_1, \ldots, y_n)$ é um vetor aleatório e $\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_p)$ corresponde ao vetor de parâmetros. A densidade preditiva para um modelo especificado Υ (Aitchison e Dunsmore, 1975) é dada por

$$\Pi(\mathbf{y}|\Upsilon) = \int_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{y}|\Upsilon)\Pi(\boldsymbol{\theta}|\Upsilon)d\boldsymbol{\theta}. \tag{5.1}$$

Suponha o caso $p = 1$, em que uma covariada x está associada ao tempo de sobrevivência T através do modelo log-linear

$$Y = \log T = \alpha + \beta x + \sigma W, \tag{5.2}$$

onde α é o parâmetro de locação, β é o coeficiente associado à covariada, σ é o parâmetro de escala e W é uma variável aleatória com distribuição log-gama generalizada cuja densidade é dada por

$$f(w) = \frac{1}{\Gamma(k)} \exp\left\{kw - e^w\right\}, \tag{5.3}$$

onde $-\infty < w < +\infty$ e $k > 0$.

5.1. Densidade Preditiva para σ e k Conhecidos

Suponha que os valores transformados $y_1, \ldots, y_n$ sejam dados, bem como as covariadas $x_1, \ldots, x_n$ conhecidas. De (5.2) segue que $w_i = (y_i - \alpha - \beta x_i)/\sigma$, e a verossimilhança para α e β pode ser escrita como

$$\begin{aligned}
\mathcal{L}(\alpha, \beta|\mathbf{y}) &= \prod_{i=1}^{n} \left[\frac{1}{\sigma\Gamma(k)} \exp\left\{k\left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right) - e^{(y_i-\alpha-\beta x_i)/\sigma}\right\}\right] \\
&= \left\{\frac{1}{\sigma\Gamma(k)}\right\}^n \exp\left\{\frac{k}{\sigma}\sum_{i=1}^{n} y_i\right\} \exp\left\{-\frac{nk\alpha}{\sigma} - \frac{\beta}{\sigma}k\sum_{i=1}^{n} x_i\right\} \exp\left\{-e^{-\alpha/\sigma}\sum_{i=1}^{n} e^{(y_i-\beta x_i)/\sigma}\right\} \\
&= \xi(k, \sigma) \exp\left\{-\frac{nk\alpha}{\sigma} - \frac{\beta k}{\sigma}\sum_{i=1}^{n} x_i\right\} \exp\left\{-e^{-\alpha/\sigma}\sum_{i=1}^{n} e^{(y_i-\beta x_i)/\sigma}\right\},
\end{aligned} \tag{5.4}$$

onde $\xi(k,\sigma) = \{\sigma\Gamma(k)\}^{-n}\exp\{\frac{k}{\sigma}\sum_{i=1}^{n} y_i\}$. Considerando uma densidade localmente uniforme para α e β, segue que $\Pi(\alpha,\beta|\Upsilon) \propto$ constante. Logo, a densidade preditiva (5.1) pode ser escrita como

$$\begin{aligned}
&\Pi(\mathbf{y}|\Upsilon) = \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \xi(k,\sigma)\exp\left\{-\frac{nk\alpha}{\sigma} - \frac{\beta k}{\sigma}\sum_{i=1}^{n} x_i\right\}\exp\left\{e^{-\alpha/\sigma}\sum_{i=1}^{n} e^{(y_i-\beta x_i)/\sigma}\right\} d\alpha d\beta \\
&= \xi(k,\sigma)\int_{-\infty}^{+\infty}\exp\left\{-\frac{\beta k}{\sigma}\right\}\int_{-\infty}^{+\infty}(e^{-\alpha})^{nk/\sigma}\exp\left\{-(e^{-\alpha})^{1/\sigma}\sum_{i=1}^{n} e^{\frac{(y_i-\beta x_i}{\sigma)}}\right\} d\alpha d\beta \\
&= \xi(k,\sigma)\int_{-\infty}^{+\infty}\exp\left\{-\frac{\beta k}{\sigma}\sum_{i=1}^{n} x_i\right\}\int_{0}^{+\infty} z^{nk/\sigma-1}\exp\left\{-z^{1/\sigma}\sum_{i=1}^{n} e^{(y_i-\beta x_i)/\sigma}\right\} d\alpha d\beta.
\end{aligned}$$

Lembrando que $\int_0^{+\infty} x^{p-1}e^{-ax^q}dx = \frac{1}{q}a^{-\frac{p}{q}}\Gamma(p/q)$, segue que

$$\Pi(\mathbf{y}|\Upsilon) = \xi\int_{-\infty}^{+\infty}\exp\left\{-\frac{\beta k}{\sigma}\sum_{i=1}^{n} x_i\right\}\frac{\sigma\Gamma(kn)}{\left\{\sum_{i=1}^{n} e^{(y_i-\beta x_i)/\sigma}\right\}^{kn}} d\beta. \tag{5.5}$$

Através do método de Laplace obtem-se uma aproximação para a expressão (5.5), que pode ser escrita convenientemente como

$$\Pi(\mathbf{y}|\Upsilon) = \xi(k,\sigma)\int_{-\infty}^{+\infty} e^{-nh(\beta)}d\beta,$$

onde $h(\beta) = -\frac{1}{n}\{\log\sigma\Gamma(nk) - \frac{\beta k}{\sigma}\sum_{i=1}^{n} x_i - nk\log\sum_{i=1}^{n} e^{(y_i-\beta x_i)/\sigma}\}$. Derivando h em relação a β e igualando-se a zero obtem-se que o valor $\hat{\beta}$ que maximiza $h(\beta)$ é aquele que satisfaz a relação

$$\frac{\sum_{i=1}^{n} x_i e^{(y_i-\hat{\beta}x_i)/\sigma}}{\sum_{i=1}^{n} e^{(y_i-\hat{\beta}x_i)/\sigma}} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{5.6}$$

Segue de (2.3) que

$$\begin{aligned}
\Pi(\mathbf{y}|\Upsilon) \cong\ & \frac{\sqrt{2\pi}\sigma^2\Gamma(kn)\exp\left\{-\frac{\hat{\beta}k\sum_{i=1}^{n} x_i}{\sigma}\right\}\xi(k,\sigma)}{\sqrt{nk}\left\{\sum_{i=1}^{n} e^{\frac{y_i-\hat{\beta}x_i}{\sigma}}\right\}^{kn-1}}\times \\
&\times\left\{\left(\sum_{i=1}^{n} x_i^2 e^{(y_i-\hat{\beta}x_i)/\sigma}\right)\left(\sum_{i=1}^{n} e^{(y_i-\hat{\beta}x_i)/\sigma}\right) - \left(\sum_{i=1}^{n} x_i e^{(y_i-\hat{\beta}x_i)/\sigma}\right)^2\right\}^{-\frac{1}{2}}
\end{aligned} \tag{5.7}$$

Considere agora a situação em que é desejado verificar se um determinado modelo Υ é adequado aos dados observados. Box(1980) sugere utilizar o valor α dado por

$$\alpha = P\left\{\Pi(\mathbf{y}|\Upsilon) < \Pi(\mathbf{y}_d|\Upsilon)\right\}, \tag{5.8}$$

onde $\mathbf{y}_d$ é o vetor de dados observados. Se α é muito pequeno, exitem evidências contra o uso do modelo Υ. Alternativamente, se a situação consiste em decidir entre dois modelos Υ_1 e Υ_2, Achcar e Bolfarine(1988) sugerem utilizar a razão

$$\frac{\Pi(\mathbf{y}|\Upsilon_1)}{\Pi(\mathbf{y}|\Upsilon_2)}, \tag{5.9}$$

a fim de selecionar qual dos dois modelos é o mais adequado. Note que a expressão (5.9) representa a ordenação relativa dos modelos, de modo que quanto maior for o seu valor, maiores são as indicações a favor do uso do modelo Υ_1.

5.2. Comparação de Dois Tratamentos

Supondo a mesma situação da seção 3.2, considere uma amostra aleatória de tamanho n_1 dada por $T_{11}, T_{12}, \ldots, T_{1n_1}$ correspondendo ao tratamento 1 e uma amostra aleatória de tamanho n_2 dada por $T_{21}, \ldots, T_{2n_2}$ representando os tempos de sobrevivência para o tratamento n_2. Assumindo que a covariada assume apenas os valores $x = 0$ (tratamento 1) e $x = 1$ (tratamento 2), segue de (5.6) que

$$\frac{\sum_{i=1}^{n} x_i \epsilon^{(y_i - \hat{\beta}x_i/\sigma)}}{\sum_{i=1}^{n} \epsilon^{(y_i - \hat{\beta}x_i)/\sigma}} = \frac{\sum_{i=1}^{n} x_i}{n} \Rightarrow \frac{\epsilon^{-\hat{\beta}/\sigma} \sum_{i=1}^{n_2} T_{2i}^{1/\sigma}}{\sum_{i=1}^{n_1} T_{1i}^{1/\sigma} + \epsilon^{-\hat{\beta}/\sigma} \sum_{i=1}^{n_2} T_{2i}^{1/\sigma}} = \frac{n_2}{n}$$

$$\Rightarrow \hat{\beta} = -\sigma \log\left\{\frac{n_2 \sum_{i=1}^{n_1} T_{1i}^{1/\sigma}}{n_1 \sum_{i=1}^{n_2} T_{2i}^{1/\sigma}}\right\}.$$

Representando a amostra combinada por $\mathbf{t} = (t_1, \ldots, t_n)$, onde $n = n_1 + n_2$, segue de (5.6) e de (5.7) que a densidade preditiva é dada por

$$\Pi(\mathbf{t}|\Upsilon) = \frac{\sqrt{2\pi}\sigma^2(kn) n_1^{kn_1 - 1/2} n_2^{kn_2 - 1/2} \xi(k, \sigma)}{\sqrt{nk} n^{kn-1} \left(\sum_{i=1}^{n_1} T_{1i}^{1/\sigma}\right)^{kn_1} \left(\sum_{i=1}^{n_2} T_{2i}^{1/\sigma}\right)^{n_2 k}}. \tag{5.10}$$

5.3. Exemplo

Considere os dados da tabela VIII analisados na seção 3 onde $n_1 = n_2 = 40$. Verifica-se que dois modelos que poderiam ser escolhidos são o modelo de Weibull, denotado por Υ_1,

onde $k_1 = 1$ e $\sigma_1 = 0,4$ e o modelo gama, denotado por Υ_2, onde $k_2 = 3$ e $\sigma_2 = 1$. Note que ambos os pares (k_i, σ_i), $i = 1,2$ determinam pontos que se encontram dentro dos contornos da figura 1. A razão das densidades preditivas dadas por (5.9) será

$$\frac{\Pi(\mathbf{t}|\Upsilon_1)}{\Pi(\mathbf{t}|\Upsilon_2)} = \frac{\sqrt{3}\Gamma(80)80^{160}(0,4)^2e^{-170,7}2^{80}(3381,5)^{120}(3254)^{120}}{(0,4)^{80}\Gamma(240)40^{160}(373645)^{40}(3435657)^{40}} = 57,103.$$

Portanto pode-se concluir que o modelo Υ_1 é o mais indicado para os dados (como era de se esperar da figura 1).

Assumindo o modelo de Weibull, pode ser mostrado (Achcar e Bolfarine, 1988, Apêndice 1) que o α de Box (expressão 5.8) é aproximadamente dado por

$$\alpha = P\left\{\chi^2_{2n-1} > \epsilon\right\},$$

onde $\epsilon = 4\exp\{-\frac{1}{\sigma}\alpha + \beta/2\}(\sum_{i=1}^{40} t_{d1i}^{1/\sigma})^{1/2}(\sum_{i=1}^{40} t_{d2i}^{1/\sigma})^{1/2}$.

Para $\alpha = 4,578$, $\beta = -0,034$ e $k = 1$ (conforme seção 3.3), segue que $\epsilon = 161,24$, de modo que

$$\alpha = P\{\chi^2_{159} > 161,24\} \cong 0,56.$$

Portanto, também pelo método α de Box conclui-se que o modelo de Weibull é apropriado.

Apêndice 1

O Uso do GLIM em Modelos de Sobrevivência

1. Introdução

Neste apêndice são abordados aspectos a respeito da utilização do pacote estatístico GLIM no ajuste de alguns modelos estudados no capítulo 2.

O GLIM é um projeto patrocinado pela Royal Statistical Society e desenvolvido por R.J.Baker, M.R.B.Clarke e J.A.Nelder. Foi desenvolvido para ajustar *modelos lineares generalizados* (MLG's) utilizando o método de máxima verossimilhança. Uma discussão detalhada destes modelos, assim como do pacote é feita por Cordeiro(1986). Aqui apenas será mencionado que os MLG's compreendem duas componentes: uma *aleatória* e outra *sistemática.*

A componente sistemática envolve uma função monótona e diferenciável, denominada função de ligação, que relaciona o preditor *linear* $\eta = \mathbf{X}\beta$ ao valor esperado μ de uma observação. Note que no caso da distribuição normal esta função equivale à função identidade, ao passo que no modelo de Poisson a função de ligação é dada pela função logaritmo.

2. Ajuste da Distribuição Exponencial e Weibull

Aitkin e Clayton(1980) discutem a utilização do GLIM para as distribuições exponencial e Weibull. A idéia central é escrever a verossimilhança convenientemente, de forma a permitir o aparecimento de uma verossimilhança de Poisson.

Considere que os tempos de sobrevivência observados são representados por t_1, $t_2,\ldots,\ t_n,\ t_{n+1},\ \ldots,\ t_{n+m}$, sendo os últimos m censurados e, que a cada observação i, $i = 1,\ldots,m+n$, estão associadas p variáveis independentes x_{ij}, $j = 1,\ldots,p$. Assuma um modelo pertencente à classe dos modelos de riscos proporcionais de tal forma que

$$\lambda(t;\mathbf{x}) = \lambda_0(t)g(\mathbf{x}) = \lambda_0(t)\exp(\mathbf{x}\beta), \tag{2.1}$$

onde β é um vetor de parâmetros. Uma vez que a densidade pode ser escrita como $f(t;\mathbf{x}) = \lambda(t;\mathbf{x})\exp\{-\int_{-\infty}^{t}\lambda(u;\mathbf{x})du\}$ e a função de sobrevivência pode ser escrita como $S(t;\mathbf{x}) = \exp\{-\int_{-\infty}^{t}\lambda(u;\mathbf{x})du\}$, segue que a função de verossimilhança pode ser expressa como

$$L = \prod_{i=1}^{n+m}[\mu_i^{\delta_i}e^{-\mu_i}]\left\{\frac{\lambda_0(t_i)}{\Lambda_0(t_i)}\right\} \tag{2.2}$$

onde $\mu_i = \Lambda_0(t_i)\exp\{\mathbf{x}_i\beta\}$, com $\Lambda_0(t_i) = \int_{-\infty}^{t}\lambda_0(u)du$.

Analisando a expressão (2.2) é fácil identificar o núcleo da função de verossimilhança de $n+m$ "variáveis de Poisson". Note que o segundo termo, apesar de não envolver parâmetros de interesse, pode conter algum parâmetro desconhecido. Desta forma, o GLIM pode ser utilizado na estimação de $\boldsymbol{\beta}$. Para tanto, utiliza-se o modelo log-linear para a "média Poisson" aplicado à função de risco, isto é,

$$\log \mu_i = \log \Lambda_0(t_i) + \mathbf{x}_i\boldsymbol{\beta} \tag{2.3}$$

Dadas estimativas iniciais dos parâmetros desconhecidos em $\Lambda_0(t_i)$, estimativas de máxima verossimilhança para $\boldsymbol{\beta}$ podem ser obtidas, com $\log \Lambda_0(t_i)$ tomada como covariada com coeficiente conhecido e igual a 1*. Uma vez tendo as estimativas $\hat{\beta}$, novos valores para os parâmetros desconhecidos em $\Lambda_0(t_i)$ podem ser obtidas (utilizando-se uma MACRO) e o processo repetido, até a convergência.

3. Distribuição Exponencial

Considere o caso em que os tempos de sobrevivência são distribuídos exponencialmente de modo que

$$f(t) = \begin{cases} \exp\{\mathbf{x}\boldsymbol{\beta} - t\exp\{\mathbf{x}\boldsymbol{\beta}\}\}, & t \geq 0; \\ 0, & t < 0. \end{cases}$$

É fácil verificar que para este caso, $\lambda_o(t)/\Lambda_0(t) = t^{-1}$. Portanto, o processo consiste em se ajustar o modelo Poisson com

$$\log \mu_i = \log t_i + \mathbf{x}_i\boldsymbol{\beta},$$

onde o *offset* é representado por $\log t_i$.

Como exemplo, considere os dados a seguir, dados por Freireich et al(1963), no qual o efeito da droga 6–MP é comparado relativamente a placebo, com relação à eficiência no tempo de remissão em pacientes com leucemia aguda. Os valores se referem aos tempos, em semanas, de 42 pacientes igualmente divididos em 2 grupos:

6–MP: 6, 6, 6, 6*, 7, 9*, 10, 10*, 11*, 13, 16, 17*, 19*, 20, 22, 23, 25*, 32*, 32*, 34*, 35*.

Placebo: 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23.

Os dados assinalados com * correspondem a observações censuradas. As variáveis a serem consideradas serão denotadas por:

---

* Este tipo de covariada é denominada *offset*. Isto é, se for *fixado* um subconjunto dos parâmetros, a contribuição ao $\log \mu_i$ denomina-se OFFSET.

T : tempos de remissão;

DELTA : indica se o valor em T é um tempo de censura (0) ou morte (1);

X : covariada assumindo os valores -0.5 se o paciente pertence ao grupo 6–MP, e 0.5 se pertence ao grupo placebo.

3.1. Programa GLIM

```
GLIM 3.77 Royal Statistical Society, London
?  $UNITS 42$                  $C Comando que especifica total
                                  de dados a serem lidos
?  $DATA T DELTA X $           $C Declaracao das  variaveis  a
                                  serem lidas
?  $DINPUT 4$                  $C Ler da unidade 4, arquivo de
                                  nome LEUCEMIA
File name?  LEUCEMIA
 6 1 -.5  6 1 -.5  6 1 -.5  6 0 -.5  7 1 -.5  9 0 -.5
10 1 -.5 10 0 -.5 11 0 -.5 13 1 -.5 16 1 -.5 17 0 -.5
19 0 -.5 20 0 -.5 22 1 -.5 23 1 -.5 25 0 -.5 32 0 -.5
32 0 -.5 34 0 -.5 35 0 -.5  1 1  .5  1 1  .5  2 1  .5
 2 1  .5  3 1  .5  4 1  .5  4 1  .5  5 1  .5  5 1  .5
 8 1  .5  8 1  .5  8 1  .5  8 1  .5 11 1  .5 11 1  .5
12 1  .5 12 1  .5 15 1  .5 17 1  .5 22 1  .5 23 1  .5
?  $YVARIATE DELTA$           $C Especificacao  da   variavel
                                  dependente
?  $ERR P $                    $C Especificacao do erro; neste
                                  caso sera Poisson
?  $LINK L $                   $C Funcao de ligacao; neste ca-
                                  so logaritmo
?  $CALC OFF = %LOG(T) $       $C Operacao  de  calculo  envol-
                                   vendo a funcao log
?  $OFFSET OFF$                $C Declaracao do OFFSET
?  $FIT X $                    $C Este comando indica que  Glim
                                  deve  ajustar  o  modelo  com
                                  variavel dependente  DELTA  e
                                  covariada X
 scaled deviance = 38.017 at cycle  4
           d.f.  = 40
?  $DISPLAY E $                $C comando para impressao das es-
```

```
                                    timativas ( opcao E ). Existem
                                    outras opcoes disponiveis.
          estimate        s.e.          parameter
   1       -2.923      0.1979      1
   2        1.526      0.3958      X
   scale parameter taken as  1.000
?  $STOP                      $C termina o programa e volta ao
                                sistema operacional.
```

3.2. Resultados

Conforme obtido pelo programa, pode-se concluir que para o modelo exponencial as estimativas são $\hat{\alpha} = -2,92$ e $\hat{\beta} = 1,53$. Note que o GLIM cria uma variável com todos os valores iguais a 1 e estima seu coeficiente, fornecendo assim uma estimativa para o intercepto.

4. Distribuição de Weibull.

Suponha que os tempos de sobrevivência têm distribuição de Weibull com densidade dada por

$$f(t) = \begin{cases} kt^{k-1}\exp\{\mathbf{x}\beta - t^k \exp\{\mathbf{x}\beta\}\}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

Nesta situação, $\lambda_0(t)/\Lambda_0(t) = t^k$, que depende do parâmetro desconhecido k. O processo se baseia no modelo

$$\log \mu_i = k \log t_i + \mathbf{x}_i\beta.$$

onde $k \log t_i$ irá corresponder ao *offset*.

Da expressão (2.2) segue que o logaritmo da função de verossimilhança é

$$\log L = n_u \log k + \sum_{i=1}^{n} (\delta_i \log \mu_i - \mu_i),$$

onde n_u é o total de observações não censuradas.Derivando em relação a k e igualando a zero, segue que

$$\hat{k} = \left\{ \sum_{i=1}^{n} (\hat{\mu}_i - \delta_i) \log t_i / n_u \right\}^{-1}. \tag{4.1}$$

Desta forma, o processo se inicia com $\hat{k} = 1$, que corresponde ao modelo inicial. São então obtidas estimativas para α e β, que são utilizadas para atualizar $\hat{k}$ através da expressão (4.1), que deve ser programada como uma *macro*.

Considerando os mesmos dados da seção anterior, foi realizada uma análise pelo GLIM, cuja listagem vem a seguir.

4.1. Programa GLIM

```
GLIM 3.77 Royal Statistical Society, London
?$INPUT 8$                    $C Atraves deste comando pode-se
                                 carregar macros previamente
                                 digitadas usando a unidade 8
                                 que sera associada ao arquivo
                                 de nome MACRO.
File name?  MACRO              $C Segue, entao, o conteudo do
                                 arquivo MACRO.
$MACRO AJUSTE                  $C Designacao da macro de nome
                                 AJUSTE, que atualiza o valor
                                 do offset a cada iteracao. E'
                                 importante notar que ela uti-
                                 liza para isto o ultimo valor
                                 obtido de k.
$CALC OFF=%K*%LOG(T)          $C Calculo do OFFSET.
$PRINT %K                      $C Comando para imprimir o valor
                                 do parametro %K a cada itera-
                                 cao.
$OFFSET OFF                    $C Especificacao do OFFSET.
$CYCLE 5                       $C especificacao do numero maxi-
                                 mo de iteracoes.
$FIT X                         $C Ajustar o modelo com a cova-
                                 riada X.
$USE WEIBULL                   $C Chamada da macro WEIBULL, que
                                 atualizara o valor do parame-
                                 tro k.
$CALC %L = %L - 1             $C Feita uma iteracao, decremen-
                                 em um o valor do contador.
$ENDMAC                        $C Comando indicativo do final
                                 da macro.
$MACRO WEIBULL                 $C Inicio da macro weibull , que
                                  sera usada para atualizar o
                                  valor de k.
```

```
$CALC %B=%K : Z=%LOG(T)/30 :
   A=%FV*Z : B = DELTA*Z :
   C=A-B : C = %CU(C) :
   %K=C(42) :  %K = 1./%K :
   %K = (%B + %K)/2.          $C Serie de calculos necessarios
                               para obtencao do novo valor
                               de k.
$ENDMAC                       $C Final da macro
$RETURN                       $C Comando para indicar final do
                                  arquivo MACRO. Voltar para  a
                                  unidade 5 (default).
$C*****************************************************************
?  $UNITS 42$                 $C Total de dados a serem lidos.
?  $DATA T DELTA X $          $C Variaveis a serem lidas.
?  $DINPUT 4$                 $C Os dados sao entao lidos da
                                  unidade 4, que esta associada
                                  ao arquivo com os dados,de no-
                                  me LEUCEMIA.
File name?  LEUCEMIA
 6 1 -.5  6 1 -.5  6 1 -.5  6 0 -.5  7 1 -.5  9 0 -.5
10 1 -.5 10 0 -.5 11 0 -.5 13 1 -.5 16 1 -.5 17 0 -.5
19 0 -.5 20 0 -.5 22 1 -.5 23 1 -.5 25 0 -.5 32 0 -.5
32 0 -.5 34 0 -.5 35 0 -.5  1 1  .5  1 1  .5  2 1  .5
 2 1  .5  3 1  .5  4 1  .5  4 1  .5  5 1  .5  5 1  .5
 8 1  .5  8 1  .5  8 1  .5  8 1  .5 11 1  .5 11 1  .5
12 1  .5 12 1  .5 15 1  .5 17 1  .5 22 1  .5 23 1  .5
?$YVARIATE DELTA $            $C Especificacao da variavel
                                  dependente:  delta.
?$ERR P$                      $C Comando que especifica o erro
                                  do modelo:  Poisson.
?$LINK L$                     $C Funcao de ligacao:  LOG.
?$CALC %K = 1.0 :                      $C Valor inicial de k (1.0) e
      %L = 6 $                  maximo de iteracoes (6).
?$WHILE %L AJUSTE$            $C Aqui tem inicio o processo itera-
                                  tivo.  Por este comando, enquanto
                                  L for diferente de zero, o valor
                                  das estimativas sao recalculadas.
                                  Note que neste ponto pode ser de-
                                  finida outra regra de parada.
```

```
  1.000
scaled deviance = 38.017 at cycle  4
         d.f.  = 40
-- change to data affects model
  1.316
scaled deviance = 50.649 at cycle  5
         d.f.  = 40
-- change to data affects model
  1.355
scaled deviance = 52.343 at cycle  5
         d.f.  = 40
-- change to data affects model
  1.363
scaled deviance = 52.719 at cycle  5
         d.f.  = 40
-- change to data affects model
    1.365
scaled deviance = 52.805 at cycle  5
         d.f.  = 40
-- change to data affects model
  1.366
scaled deviance = 52.825 at cycle  5
         d.f.  = 40
-- change to data affects model
  1.366
scaled deviance = 52.829 at cycle  5
         d.f.  = 40
?$DISPLAY E$                    $C Uma vez ajustado o modelo,
                                  imprimir as estimativas.
        estimate        s.e.          parameter
   1      -3.936      0.1991      1
   2       1.731      0.3983      X
   scale parameter taken as  1.000
?$STOP                           $C Fim de programa.
```

4.2 Resultados

Conforme os resultados obtidos, verifica-se que as estimativas se estabilizam em $\hat{k} = 1,36$, $\hat{\alpha} = -3,94$ e $\hat{\beta} = 1,73$. Um intervalo de confiança para β seria $1,73 \pm 2 \times 0,40$, que não inclui o valor zero, indicando diferença significante entre os dois grupos. É sabido que para

estes dados o modelo de Weibull é mais indicado, o que se reflete no valor de $\hat{k}$, diferente de 1.

5. Discussão

O GLIM se mostra bastante útil na modelagem de dados de sobrevivência. O uso interativo do programa fornece uma maleabilidade considerável, principalmente quando existem muitas covariadas a serem consideradas. Análises de resíduos também podem ser feitas pelo pacote, que apresenta opções para gráficos e estatísticas descritivas.

Além dos modelos descritos, Aitkin e Clayton(1980) também consideram a estimação pelo GLIM dos parâmetros nos modelos de valor extremo e valor extremo generalizado. A estimação no modelo de Cox também pode ser realizada através do GLIM, inclusive no caso em que existem observações empatadas (veja Whitehead,1980).

REFERÊNCIAS BIBLIOGRÁFICAS

Abramowitz,M. and Stegun,I.A.,Eds. (1965). *Handbook of Mathematical Functions.* New York: Dover.

Achcar,J.A. and Bolfarine,H. (1986). The log-linear model with a generalized gamma distribution for the error: a Bayesian approach, *Statistics & Probability Letters*, **4**, 325-332, North-Holland.

Achcar,J.A. and Bolfarine,H. (1988). Predictive densities in survival analysis with a generalized gamma regression model, *Revista Brasileira de Probabilidade e Estatística*, **2**, 23-31.

Achcar,J.A., Bolfarine,H. and Pericchi,L.R. (1987). Transformation of survival data to an extreme value distribution, *The Statistician*, **36**, 229-234.

Aitkin,M. and Clayton,D.G. (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM, *Appl. Stat.*, **29**, 156-163.

Bartlett,M.S. and Kendall,D.G. (1946). The statistical analysis of variance heterogeneity and the logarithmic transformation, *J. R. Stat. Soc.*, **Suppl 8**, 128-138.

Billingsley,P. (1968). *Weak Convergence of Probability Measures.* New York: Wiley.

Böhmer,P.E. (1912). Theorie der unabhängigen Wahrscheinlichkeiten, *Rapports, Mémoires et Procès-verbaux de Septième Congrès International d'Actuaires, Amsterdam*, **2**, 327-343.

Box,G.E. (1980). Sampling and Bayes inference in scientific modelling and robustness, *J. R. Stat. Soc.*, **A143**, 383-430.

Box,G.E. and Cox,D.R. (1964). An analysis of transformations, *J. R. Stat. Soc.*, **B26**, 211-250.

Box,G.E. and Tiao,G.C. (1973). *Bayesian Inference in Statistical Analysis.* New York: Adison-Wesley.

Breslow,N.E. and Crowley,J. (1974). A large sample study of the life table and product limit estimates under random censorship, *Ann. Stat.*, **2**, 437-453.

Buckley,T. and James,I. (1979). Linear regression with censored data, *Biometrika*, **66**, 429-436.

Chambers,J.M. (1977). *Computational Methods for Data Analysis.* New York: Wiley.

Cordeiro,G.M. (1986). *Modelos Lineares Generalizados.* Campinas: IMECC-UNICAMP.

Cox,D.R. (1962). Further results on tests of separate families of hypotheses, *J. R. Stat. Soc.*, **B24**, 406-423.

Cox,D.R. (1972). Regression models and life tables (with discussion), *J. R. Stat. Soc.*, **B34**, 187-202.

Cox,D.R. (1975). Partial likelihood, *Biometrika*, **62**, 269-276.

Cox,D.R. and Hinkley,D.V. (1968). A note on the efficiency of least squares estimates, *J. R. Stat. Soc.*, **B30**, 284-289.

Crowley,J. (1970). A comparison of several life table methods. Unpublished M. S. Thesis, University of Washington.

Efron,B. (1977). The efficiency of Cox's likelihood function for censored data, *J. Am. Stat. Assoc.*, **72**, 555-565.

Freireich,E.O. et al. (1963). The efecct of 6-mercaptopurine on the duration of steroid induced remission in acute leukemia, *Blood*, **21**, 699-716.

Greenwood,M. (1926). The natural duration of cancer, *Reports of Public Health and Medical Subjects*, Vol. 33. London: Her Majesty's Stationery Office.

Kalbfleisch,J.D. and Prentice,R.L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Kaplan,E.L. and Meier,P. (1958). Nonparametric estimation from incomplete observations, *J. Am. Stat. Assoc.*, **53**, 457-481.

Kiefer,J. and Wolfowitz,I. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *Ann. Math. Stat.*, **27**, 887-906.

Koul,H., Susarla,V. and van Ryzin,J. (1981). Regression analysis with randomly right-censored data, *Ann. Stat.*, **9**, 1276-1288.

Lawless,J.F. (1982). *Statistical Models and Methods for Lifetime Data.* New York: Wiley.

Lee,E.T. (1980). *Statistical Methods for Survival Data Analysis.* Belmont: Lifetime Learning Publication.

Mariotto,A.B. (1983). Inferência Parcial. Dissertação de Mestrado, Universidade de São Paulo.

Miller,R.G. (1976). Least squares regression with censored data, *Biometrika*, **63**, 449-464.

Miller,R.G. (1983). What price Kaplan-Meier?, *Biometrics*, **39**, 1077-1081.

Peterson,A.V. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions, *J. Am. Stat. Assoc.*, **72**, 854-858.

Rao,C.R. (1973). *Linear Statistical Inference and Its Applications*, second edition. New York: Wiley.

Schmee,J. and Hahn,G.J. (1979). A simple method for regression analysis with censored data, *Technometrics*, **21**, 417-432.

Serfling,J.S. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

Tierney,L. and Kadane,J.B. (1986). Accurate approximations for posterior moments and marginal densities, *J. R. Stat. Soc.*, **A81**, 82-86.

Tsiatis,A.A. (1981). A large sample study of Cox's regression model, *Ann. Stat.*, **9**, 93-108.

Whitehead,J. (1980). Fitting Cox's regression model to survival data using GLIM, *Appl. Stat.*, **20**, 268-275.