Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Modelos de Censura Tipo II Progressiva e Propriedades Assintóticas do Estimador de Máxima Verossimilhança

Éder Silva de Brito

Brasília
2014

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Modelos de Censura Tipo II Progressiva e Propriedades Assintóticas do Estimador de Máxima Verossimilhança

Éder Silva de Brito

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre em Matemática.

Orientadora
Prof^a. Dr^a. Daniele da Silva Baratela Martins Neto

Brasília
2014

À minha mãe Márcia Neide com todo carinho.

Agradecimentos

Aos meus pais e meus irmãos pelo amor e confiança que depositaram em mim por todos esses anos, sempre se esforçando pra que eu pudesse continuar lutando pela realização dos meus sonhos.

Às minhas queridas amigas Lais e Eliana por todo o apoio, incentivo, ajuda, compreensão e amizade. Devo a vocês tudo de bom que ocorreu na minha vida nesses últimos anos. Obrigado por terem sempre me confortado nos momentos de maiores dificuldades.

Aos grandes amigos e companheiros Jorge, Gustavo e Johnathan pela paciência nos momentos que me fiz ausente, pelas palavras sempre sábias nos momentos tortuosos e pela disposição em ajudar sempre que fosse possível. Tenho por vocês uma amizade e carinho imensuráveis.

A todos meus amigos, colegas e alunos do Instituto Federal de Goiás - Campus Formosa, pelo apoio e incentivo para a realização dessa etapa.

À Professora Daniele pelas orientações, dedicação e empenho pelo sucesso do nosso trabalho. Sua atenção e incentivo me fizeram aprender muito e crescer profissionalmente.

Às Professoras Viviane e Cira, componentes da banca examinadora pelas valiosíssimas correções e sugestões dadas.

A todos os Professores do Departamento de Matemática com os quais tive aula, em especial à Professora Cátia pelo exemplo de Docente que ama o que faz. Suas aulas são espelho para mim e me motivam na escolha por essa linda carreira.

A todos os funcionários do Departamento de Matemática, em especial à Bruna pela constante disposição e atenção com todos os alunos. Você é muito querida!

A toda a organização e todos os envolvidos com a OBMEP, programa que mudou minha vida me dando perspectivas de seguir o sonho de ser Professor de Matemática. Obrigado por mudar não só a minha vida, mas a de milhares de outros jovens de todo o país!

À Capes pelo apoio financeiro.

Resumo

Neste trabalho, estudamos métodos inferenciais baseados em amostras na presença de censura tipo II progressiva. Primeiramente, apresentamos três modelos envolvendo as distribuições: de Valor Extremo, por Ding e Yu (2012), Exponencial Generalizada, por Ismail (2012), e Lognormal de Três Parâmetros, por Basak et al. (2009). Num segundo momento, baseados no estudo de Lin e Balakrishnan (2011), investigamos as propriedades de consistência e normalidade assintótica de estimadores de máxima verossimilhança para modelos sob esquema de censura tipo-II progressiva.

Palavras-chave: censura tipo II progressiva, estimador de máxima verossimilhança, teoria assintótica, consistência.

Abstract

In this work, we study inferential methods based on samples in the presence of progressively Type-II censoring. First, we present three models involving distributions: Extreme-Value, by Ding and Yu (2012), Generalized Exponential, by Ismail (2012), and Three-Parameter Lognormal, by Basak et al. (2009). Secondly, based on the study of Lin and Balakrishnan (2011), we investigated the properties of consistency and asymptotic normality of maximum likelihood estimators for models under a progressive Type-II censoring scheme.

Keywords: progressive Type-II censoring, maximum likelihood estimation, asymptotic theory, consistency.

Sumário

Introdução

Ao realizar inferência estatística a partir de dados obtidos em testes de confiabilidade, muitas vezes nos deparamos com amostras onde nem todos os tempos de falha desejados são observados. Esses casos são denominados censuras, isto é, são observações parciais em um estudo interrompido por alguma razão, não permitindo que as observações completas do tempo de falha sejam obtidas. Censuras são recorrentes em processos de análise de sobrevivência, onde o tempo e o custo de tais experimentos são limitados, ou por diversos outros motivos alheios ao estudo e às condições impostas sobre o objeto de estudo.

Um caso amplamente aplicável em situações como essa é a Censura do Tipo II Progressiva, onde a censura é realizada em algumas etapas, como o próprio nome sugere. Mais precisamente, a cada falha observada, outras unidades em funcionamento são retiradas aleatoriamente do experimento, até que se obtenha um número pré-determinado de falhas observadas.

Cohen observa em seu trabalho [11] que era possível modelar satisfatoriamente, por exemplo, situações de perda de unidades por acidentes de manuseio ou de mau usos em testes de durabilidade de produtos eletrônicos, utilizando censuras progressivas. Daí, surgem na literatura trabalhos com diferentes métodos de estimação dos parâmetros para amostras censuradas tipo II progressivamente, dos quais podemos destacar o Princípio de Máxima Verossimilhança.

Introduzido formalmente por Fisher em [17], o princípio de Verossimilhança é um dos mais clássicos procedimentos utilizados para a obtenção dos estimadores para um modelo paramétrico, além de ser um dos mais importantes, do ponto de vista teórico. Mais precisamente, seja $X_1, \ldots, X_n$ uma amostra aleatória da densidade $f(x;\theta)$ onde θ é o parâmetro a ser estimado, então pretende-se encontrar o valor de θ que maximiza a função de verossimilhança $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$. Visto que a função logarítmica é uma função estritamente monótona, o Princípio de Máxima Verossimilhança sugere maximizar a função de log-verossimilhança por meio da resolução das equações de verossimilhança

$$\frac{\partial \log L(\theta)}{\partial \theta} = 0.$$

Tal máximo, caso exista e seja único, é denominado o Estimador de Máxima Verossimilhança para o parâmetro θ do modelo.

Ao obter o estimador de máxima verossimilhança para os parâmetros de um modelo, deseja-se saber sobre a qualidade desse estimador e seu comportamento em amostras suficientemente grandes. Cramer, em [14], exibe condições que garantem propriedades como consistência e normalidade assintótica dos estimadores de máxima verossimilhança. Isso significa que sob determinadas condições de regularidade, temos que qualquer sequência $\hat{\theta}_n$ de estimadores de θ converge em probabilidade para o verdadeiro valor do parâmetro quando $n \to \infty$, isto é, $\hat{\theta} = \hat{\theta}_n$ é dito um estimador consistente de θ. Além disso, $\sqrt{n}(\hat{\theta}_n - \theta)$ converge em distribuição para uma variável com distribuição normal quando $n \to \infty$, isto é, o estimador $\hat{\theta}_n$ apresenta normalidade assintótica. Estes resultados são clássicos e amplamente aplicados em estudos sobre comportamento assintóticos de estimadores em amostras completas.

Alguns anos após os primeiros estudos de Fisher, surgem na literatura novos modelos estatísticos envolvendo estimação de parâmetros, como por exemplo, no tratamento de amostras em que nem todas as unidades são observadas, ou seja, modelos envolvendo censuras. Dentre os primeiros trabalhos podemos citar Gupta [18] e Cohen [10] e [13]. A partir daí, diversos modelos baseados em diferentes esquemas de censura foram estudados. Por sua vasta aplicabilidade prática e teórica, destacam-se as censuras à direita e seus dois principais tipos: a Censura do Tipo I e a Censura do Tipo II. Posteriormente surgem generalizações desses dois tipos de censura, como é o caso da Censura Progressiva, e em particular, a Censura do Tipo II Progressiva, destacada nos estudos deste trabalho.

Herd apresenta em [20] o primeiro trabalho com modelo de Censura Progressiva. Desde então, vários autores estudaram diferentes modelos sob censura tipo II progressiva, baseados em diferentes distribuições e obtiveram os estimadores de máxima verossimilhança para os parâmetros desses modelos, o que pode ser visto, por exemplo em Balakrishnan [3], Basak [6], Ismail [22] e Ding [16].

A formalização dos modelos de censura trouxe naturalmente os mesmos questionamentos acerca do comportamento assintótico dos estimadores de máxima verossimilhança, agora obtidos a partir de amostras censuradas. Nesse sentido, Balakrishnan e Lin exibem em [4], as propriedades de consistência e normalidade assintótica para o estimador de máxima verossimilhança em modelos uniparamétricos, envolvendo amostras censuradas tipo II progressivamente. Eles utilizam o Princípio da Informação Perdida apresentado por Louis em

[27], para desmembrar a amostra sob censura em duas partes: uma considerada completa (sem censura) e a outra composta pelas unidades censuradas.

Parte do problema já estava resolvido, pois a convergência dos estimadores para uma amostra sem censura é um resultado clássico (apresentado, por exemplo, em [14]). Restava então verificar o comportamento dos parâmetros na parte censurada da amostra. Para isso, Balakrishnan e Lin propoem novas condições de regularidade para as funções de densidade e de distribuição do modelo, em relação às condições exigidas nas amostras completas. Tais condições permitem o uso de uma versão da Lei fraca dos Grandes Números apresentada por Hoadley em [21], o que garante a obtenção das convergências desejadas.

Baseados nesse histórico, temos como objetivos neste trabalho, apresentar o modelo de Censura Tipo II Progressiva, obter os estimadores de máxima verossimilhança dos parâmetros de alguns modelos e detalhar os estudos de Balakrishnan e Lin em [4], onde é mostrado que sob determinadas condições de regularidade as propriedades de consistência e normalidade assintótica são garantidas para o estimador de máxima verossimilhança do parâmetro em amostras sob esse tipo de censura.

No Capítulo 1 resumimos alguns conceitos e resultados necessários ao desenvolvimento do trabalho. Exibimos as propriedades de consistência e normalidade assintótica de estimadores de máxima verossimilhança em amostras sem censura e a garantia dessas propriedades sobre determinadas condições. Também neste capítulo apresentamos uma breve explanação sobre testes de confiabilidade e a definição de censura e seus principais tipos. Na Seção 1.2.1, damos atenção especial à censura do tipo II progressiva, nosso principal objeto de estudo. Apresentamos na Seção 1.3 o algoritmo EM, proposto por Desmpster et al. em [15], importante método iterativo para obter os estimadores de máxima verossimilhança de parâmetros em amostras onde os dados observados são incompletos, como por exemplo, nos casos de censura.

No Capítulo 2 apresentamos três modelos de censura do tipo II progressiva e obtemos os estimadores de máxima verossimilhança para seus parâmetros, por meio do princípio de máxima verossimilhança. Tais modelos, baseados nas distribuição de Valor Extremo Gumbel, Exponencial Generalizada e Lognormal de 3 parâmetros, foram escolhidos por serem amplamente utilizados na modelagem de testes de confiabilidade sobre tempo de falha de diversos produtos, além de caracterizarem inúmeras outras situações em análises de sobrevivência. Na Seção 2.2, exibimos a estimação de parâmetros, via máxima verossimilhança, considerando uma amostra com distribuição de Valor Extremo Gumbel, apresentada em [16]. Na Seção 2.3,tratamos de um modelo com distribuição Exponencial Generalizada, apresentado em [22], onde acrescenta-se a hipótese de realização do teste sob condições de aceleramento, procedimento comumente utilizado em teste de confiabilidade. Na Seção 2.4, utilizamos o algoritmo EM para estimar os parâmetros , via máxima verossimilhança de um modelo com distribuição Lognormal de 3 parâmetros, proposta apresentada em [6].

Por fim, no Capítulo 3 estudamos as propriedades assintóticas do estimador de máxima verossimilhança em amostras censuradas tipo II progressivamente. Exibimos as condições de regularidade apresentadas por Balakrishnan e Lin, e exibimos detalhadamente a demonstração feita por esses autores verificando a consistência e a normalidade assintótica do estimador.

Capítulo 1

Preliminares

1.1 - Inrodução

Nesta seção introduzimos as definições e os resultados teóricos necessários ao desenvolvimento dos estudos deste trabalho, tendo como referências básicas Cramér [14], Lehmann [25], Casella [8], Lawless [24], Klein [23], Dempster [15], Wu [33] e Tanner [32].

Iniciamos na Seção 1.2 apresentando as definições de amostra aleatória e estimadores de máxima verossimilhança e verificamos as propriedades de consistência e normalidade assintótica para esses estimadores. Na Seção 1.3 tratamos das análises de sobrevivência e testes de confiabilidade, exibindo os diferentes tipos de censura, em especial a censura do tipo II progressiva, principal interesse deste trabalho. Terminamos na Seção 1.4 abordando o algoritmo EM, ferramenta clássica para estimações de parâmetros em amostras com dados incompletos.

1.2 - Princípio de Máxima Verossimilhança

Definição 1.1. Uma *amostra aleatória* de tamanho n de uma função de distribuição F é dada por $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ onde $X_1, X_2, \ldots, X_n$ são variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com função de distribuição comum F.

Se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias contínuas i.i.d. com densidade comum f, podemos dizer que $\mathbf{X}$ é uma *amostra aleatória de* f.

Para introduzirmos as definições abaixo, consideramos inicialmente $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma densidade $f(x;\theta)$ com respeito a medida σ-finita μ sobre $\mathbb{R}$, tal que θ é um parâmetro pertencente ao espaço $\Theta \subset \mathbb{R}^r$, $r \in \mathbb{N}, r \geq 1$. Assumimos que θ é um parâmetro r-dimensional desconhecido e, portanto, temos como objetivo estimá-lo.

Definição 1.2. Um *estimador* de θ é uma estatística $\hat{\theta}_n = T(X_1, \ldots, X_n)$, onde T é uma função de $\mathbb{R}^n$ em Θ.

Definição 1.3. A *função de verossimilhança* de um parâmetro θ baseada nas observações $x_1, \ldots, x_n$ da amostra $X_1, \ldots X_n$ é dada por

$$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta).$$

Definição 1.4. O *estimador de máxima verossimilhança* (EMV) de θ é o valor $\hat{\theta}$ que maximiza a função de verossimilhança $L(\theta) = L(\theta; x_1, \ldots, x_n)$ caso exista, isto é, é um valor $\hat{\theta} = \hat{\theta}_n = \hat{\theta}(x_1, \ldots, x_n) \in \Theta$ tal que

$$L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta; x_1, \ldots, x_n) = \sup_{\theta \in \Theta} \prod_{i=1}^{n} f(x_i; \theta),$$

onde $(x_1, \ldots, x_n)$ é uma amostra observada de $(X_1, \ldots, X_n)$.

Observação 1.5. **(i)** Pode ocorrer de $\hat{\theta}$ não existir ou existir e não ser único.

(ii) A função logarítmica é uma função monótona estritamente crescente e, sendo assim, afim de facilitar o processo de obtenção dos EMV, podemos substituir a função de verossimilhança pela função log-verossimilhança de θ, dada por:

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta).$$

Se existir, o valor de θ que maximiza a função $l(\theta)$ será também o valor que maximiza a função $L(\theta)$, isto é, será o estimador de máxima verossimilhança desejado.

(iii) Para obter o valor do EMV $\hat{\theta}$, resolvemos as equações de verossimilhança

$$\frac{\partial l(\theta)}{\partial \theta_i} = 0, \quad i = 1, \ldots, r,$$

desde que $l(\theta)$ seja diferenciável em Θ e $\hat{\theta}$ seja um máximo local.

A grande importância dos EMV dos parâmetros de uma amostra se dá pelas suas propriedades matemáticas quando o tamanho n da amostra é suficientemente grande. Tais propriedades são enunciadas abaixo.

Definição 1.6. Um estimador $\hat{\theta}_n$ é dito *estimador consistente de* θ se a sequência $\{\hat{\theta}_n\}_n$ converge em probabilidade para θ, isto é, se para todo $\epsilon > 0$ temos que

$$\lim_{n \to \infty} P(|\hat{\theta}_n - \theta| < \epsilon) = 1.$$

Definição 1.7. A matriz quadrada de ordem r, $I(\theta) = [I_{jk}(\theta)]_{r\times r}$ cujas entradas são dadas por:
$$I_{jk}(\theta) = cov\left[\frac{\partial}{\partial\theta_j}\log f(X_1;\theta), \frac{\partial}{\partial\theta_k}\log f(X_1;\theta)\right], \quad j,k = 1,\ldots,r,$$

e nos casos em que
$$E\left(\frac{\partial}{\partial\theta_j}\log f(X_1;\theta)\right) = 0, \ \forall\ j = 1,\ldots,r;$$
e
$$E\left(\frac{\partial}{\partial\theta_j}\log f(X_1;\theta).\frac{\partial}{\partial\theta_k}\log f(X_1;\theta)\right) = -E\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log f(X_1;\theta)\right),$$

então $I_{jk} = -E\left(\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log f(X_1;\theta)\right)$ é chamada de *Matriz de Informação de Fisher de* X_1.

Definição 1.8. Dizemos que um estimador $\hat{\theta}_n$ tem distribuição *assintoticamente normal* se satisfaz a condição
$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N\left(0, \frac{1}{I(\theta)}\right),$$
isto é, se $\sqrt{n}(\hat{\theta}_n - \theta)$ converge em distribuição para uma variável aleatória com distribuição normal de média zero e matriz de covariância $[I(\theta)]^{-1}$.

No que segue, vemos alguns resultados acerca da existência e qualidade do EMV de um parâmetro θ, ou seja, apresentamos condições de regularidade, sob as quais, garantimos a consistência e a normalidade assintótica do estimador de máxima verossimilhança de θ. Assumimos que θ é um parâmetro unidimensional com valores possíveis em $\Theta \subset \mathbb{R}$. A referência básica utilizada aqui é Cramér [14] e Lehmann [25].

Seja $\mathscr{F} = \{f(x;\theta), \theta \in \Theta\}$ uma família de densidades relativas a uma medida σ-finita μ. Considere um conjunto de dados observados $x_1,\ldots,x_n$ que são realizações i.i.d. de uma variável aleatória X com densidade $f(x;\theta_0) \in \mathscr{F}$, onde $\theta_0 \in \Theta \subset \mathbb{R}$ é o parâmetro verdadeiro a ser estimado.

Seja $l(\theta) = \sum_{i=1}^{n}\log f(x_i;\theta)$ a função de log-verossimilhança de θ_0. Vejamos abaixo, então, uma listagem das condições de regularidade para os resultados que vêm a seguir.

(C1) As distribuições das observações são identificáveis, isto é, para $\theta_1, \theta_2 \in \Theta \subset \mathbb{R}$, se $\theta_1 \neq \theta_2$ então $f(x;\theta_1) \neq f(x;\theta_2)$.

(C2) O espaço paramétrico Θ é um intervalo aberto não degenerado $I = (a,b)$, tal que
$$\Theta : -\infty \leq a < \theta < b \leq \infty.$$

(C3) O conjunto suporte de $f(x;\theta)$, isto é, *supp*$(f) = \{x; f(x;\theta) > 0\}$ é independente de θ.

(C4) Para quase todo x, as derivadas $\frac{\partial}{\partial\theta}\log f(x;\theta)$, $\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)$ e $\frac{\partial^3}{\partial\theta^3}\log f(x;\theta)$ existem para todo $\theta \in I$.

(C5) Para todo $\theta \in I$, $\left|\frac{\partial}{\partial\theta}f(x;\theta)\right| < G_1(x)$, $\left|\frac{\partial^2}{\partial\theta^2}f(x;\theta)\right| < G_2(x)$ e $\left|\frac{\partial^3 \log f(x;\theta)}{\partial\theta^3}\right| < G_3^*(x)$, onde G_1 e G_2 são integráveis sobre $(-\infty,\infty)$, enquanto $\displaystyle\int_{-\infty}^{\infty} G_3^*(x)f(x;\theta)dx < K$, onde K é independente de θ.

(C6) Para todo $\theta \in I$, a integral $\gamma^2 = \displaystyle\int_{-\infty}^{\infty}\left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right]^2 f(x;\theta)dx$ é finita e positiva.

Lema 1.9. *Seja* $\boldsymbol{X} = (X_1,\ldots,X_n)$ *uma amostra aleatória da densidade* $f(x;\theta_0)$, *que satisfaz as condições* $(C1)$, $(C2)$ *e* $(C3)$. *Então, para todo* $\theta \in \Theta, \theta \neq \theta_0$,

$$P\left(L(\theta) < L(\theta_0)\right) \overset{n\to\infty}{\longrightarrow} 1. \tag{1.1}$$

Demonstração. Podemos reescrever a desigualdade

$$L(\theta) < L(\theta_0)$$

como

$$\log L(\theta) < \log L(\theta_0),$$

donde

$$\frac{1}{n}\sum_{i=1}^{n}\left[\log f(X_i;\theta) - \log f(X_i;\theta_0)\right] < 0$$

e portanto

$$\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right) < 0.$$

Como a função $(-\log)$ é estritamente convexa, pela Desigualdade de Jensen e pelas condições $(C1)$, $(C2)$ e $(C3)$ segue que

$$E_{\theta_0}\left[\log\left(\frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right)\right] < \log\left(E_{\theta_0}\left[\frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right]\right) = 0. \tag{1.2}$$

Como as variáveis X_i, $i = 1,\ldots,n$ são i.i.d., temos que $\log\left(\frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right)$ também o são e possuem esperança finita, por (1.2). Daí, pela Lei Fraca dos Grandes Números (LfGN) de

Khintchine segue que

$$\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right)\overset{n\to\infty}{\longrightarrow} E_{\theta_0}\left[\log\left(\frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right)\right], \tag{1.3}$$

e então por (1.2) e (1.3) segue que

$$P\left(\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right)<0\right)\overset{n\to\infty}{\longrightarrow} 1.$$

■

Mesmo que não se conheça o valor verdadeiro θ_0, pode-se estimar o valor $\hat{\theta}$ de θ_0 que maximiza a densidade conjunta de $X_1,\ldots,X_n$, isto é, que maximiza a função de verossimilhança dadas as observações $x_1,\ldots,x_n$. O Lema 1.9 mostra que, com probabilidade tendendo a 1 quando $n\to\infty$, a função de verossimilhança em θ_0 assume valor maior do que em qualquer outro θ fixado.

Lema 1.10. *Seja $\boldsymbol{X}=(X_1,\ldots,X_n)$ uma amostra aleatória da densidade $f(x;\theta_0)$ que satisfaz as condições $(C1)-(C6)$. Então, para todo $a>0$ suficientemente pequeno, com probabilidade tendendo a 1 quando $n\to\infty$, existem soluções $\hat{\theta}_n(a)=\hat{\theta}_n(X_1,\ldots,X_n)(a)$ da equação de verossimilhança*

$$\frac{\partial}{\partial\theta}l(\theta)=0, \tag{1.4}$$

tais que $\hat{\theta}_n(a)$ é um máximo local de $L(\theta)$ e $\hat{\theta}_n(a)\in I_a=(\theta_0-a,\theta_0+a)$. Ainda,

$$\lim_{n\to\infty}P(L(\hat{\theta}_n(a))<L(\theta_0))=1.$$

Demonstração. Consideremos, primeiramente, a expansão da função $\frac{\partial}{\partial\theta}\log f(x;\theta)$ pela fórmula de Taylor com resto de Lagrange em torno de θ_0, ou seja,

$$\frac{\partial}{\partial\theta}\log f(x;\theta)=\frac{\partial}{\partial\theta}\log f(x;\theta_0)+(\theta-\theta_0)\frac{\partial^2}{\partial\theta^2}\log f(x;\theta_0)+\frac{(\theta-\theta_0)^2}{2}\frac{\partial^3}{\partial\theta^3}\log f(x;\bar{\theta}),$$

onde $\bar{\theta}$ está entre θ e θ_0.

Daí podemos escrever

$$\frac{\partial}{\partial\theta}\log L(\theta)=\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i;\theta)=$$

$$\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i;\theta_0)+(\theta-\theta_0)\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(x_i;\theta_0)+\frac{(\theta-\theta_0)^2}{2}\sum_{i=1}^{n}\frac{\partial^3}{\partial\theta^3}\log f(x_i;\bar{\theta}).$$

Pela condição (C5) podemos definir

$$\Delta(x_i) = \begin{cases} 0, & \text{se } G_3^*(x_i) = 0 \\ \frac{\partial^3}{\partial\theta^3} \log f(x_i; \bar{\theta}) \frac{1}{G_3^*(x_i)}, & \text{se } G_3^*(x_i) > 0, \end{cases}$$

o que nos dá $0 \leq |\Delta(x_i)| \leq 1$, para todo $i = 1, \ldots, n$, e então segue que

$$\frac{1}{n}\frac{\partial}{\partial\theta} \log L(\theta) = B_0 + (\theta - \theta_0)B_1 + \frac{(\theta-\theta_0)^2}{2}\bar{B}_2 \tag{1.5}$$

onde

$$\begin{aligned} B_0 &= \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(x_i;\theta_0); \\ B_1 &= \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2}{\partial\theta^2} \log f(x_i;\theta_0); \\ \bar{B}_2 &= \frac{1}{n}\sum_{i=1}^{n} G_3^*(x_i)\Delta(x_i). \end{aligned}$$

Observe que, como $|\Delta(x_i)| \leq 1$, para todo $i = 1, \ldots n$,

$$\begin{aligned} \left|\bar{B}_2\right| &= \left|\frac{1}{n}\sum_{i=1}^{n} G_3^*(x_i)\Delta(x_i)\right| \\ &\leq \frac{1}{n}\sum_{i=1}^{n} |\Delta(x_i)|\, G_3^*(x_i) \\ &\leq \frac{1}{n}\sum_{i=1}^{n} G_3^*(x_i) := B_2. \end{aligned}$$

Assim, temos que $\bar{B}_2 = \Delta B_2$, onde $|\Delta| \leq 1$ e podemos reescrever a igualdade (1.5) como

$$\frac{1}{n}\frac{\partial}{\partial\theta} \log L(\theta) = B_0 + (\theta - \theta_0)B_1 + \frac{(\theta-\theta_0)^2}{2}\Delta B_2. \tag{1.6}$$

Note que as B_i's são funções das variáveis aleatórias $X_1, \ldots, X_n$.

Mostramos primeiramente que, com probabilidade tendendo a 1 quando $n \to \infty$, a equação (1.4) tem uma raiz em $I_a = (\theta_0 - a, \theta_0 + a)$ para $a > 0$ (suficientemente pequeno). Fazemos isso em três etapas. São elas:

(I) Pelo Teorema da Convergência Dominada e pelas condições $(C4)$ e $(C5)$, temos que

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(x;\theta_0)dx = \int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta^2} f(x;\theta_0)dx = 0,$$

para todo $\theta \in I$, e portanto

$$\begin{aligned}
E\left(\frac{\partial}{\partial\theta}\log f(X;\theta_0)\right) &= \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta}\log f(x;\theta_0).f(x;\theta_0)dx \\
&= \int_{-\infty}^{\infty} \frac{1}{f(x;\theta_0)}\frac{\partial}{\partial\theta} f(x;\theta_0).f(x;\theta_0)dx = 0
\end{aligned}$$

e

$$\begin{aligned}
E\left(\frac{\partial^2}{\partial\theta^2}\log f(x;\theta_0)\right) &= \int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta^2}\log f(x;\theta_0).f(x;\theta_0)dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta}\left[\frac{\frac{\partial}{\partial\theta} f(x;\theta_0)}{f(x;\theta_0)}\right] f(x;\theta_0)dx \\
&= \int_{-\infty}^{\infty} \frac{f(x;\theta_0)\frac{\partial^2}{\partial\theta^2} f(x;\theta_0) - \left[\frac{\partial}{\partial\theta} f(x;\theta_0)\right]^2}{(f(x;\theta_0))^2} f(x;\theta_0)dx \\
&= \int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta^2} f(x;\theta_0)dx - \int_{-\infty}^{\infty}\left[\frac{\partial}{\partial\theta}\log f(x;\theta_0)\right]^2 f(x;\theta_0)dx \\
&= -E\left[\frac{\partial}{\partial\theta}\log f(X;\theta_0)\right]^2 = -\gamma^2,
\end{aligned}$$

onde a última igualdade sai da condição (C6). Dessa forma, obtemos

$$E(B_0) = 0 \ \text{ e } \ E(B_1) = -\gamma^2.$$

(II) Analisamos agora o comportamento das B_i's.

$(II.i)$ Como $X_1, \ldots, X_n$ são variáveis i.i.d., as variáveis $\frac{\partial}{\partial\theta}\log f(X_i;\theta)$, $i = 1, \ldots, n$, também o são. De (I) e pela LfGN de Khintchine, temos que

$$B_0 = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log f(X_i;\theta) \xrightarrow{P} E\left(\frac{\partial}{\partial\theta}\log f(X;\theta_0)\right) = 0.$$

Então para qualquer $a > 0$ temos que

$$P(|B_0| < a^2) = P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i;\theta_0)\right| < a^2\right) \stackrel{n\to\infty}{\longrightarrow} 1.$$

$(II.ii)$ Pela mesma ideia de $(II.i)$, as variáveis $\frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta)$ são i.i.d. e, pela LfGN de Khintchine, temos que

$$B_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta_0) \xrightarrow{P} E\left(\frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta_0)\right) = -\gamma^2.$$

Assim, para todo $\epsilon > 0$,

$$P(|B_1 - (-\gamma^2)| < \epsilon) \overset{n\to\infty}{\longrightarrow} 1.$$

Tomando $\epsilon = \gamma^2/2$ obtemos

$$P(B_1 < -\gamma^2/2) \overset{n\to\infty}{\longrightarrow} 1.$$

$(II.iii)$ Temos que $G_3^*(X_i)$, $i = 1,\ldots,n$, são variáveis aleatórias i.i.d. com média $E(G_3^*(X_i)) = E(G_3^*(X_1)) := \bar{x}$, onde, pela condição $(C5)$, $0 < \bar{x} < K < \infty$, $i = 1,\ldots,n$. Assim, pela Lei dos Grandes Números, segue que

$$B_2 = \frac{1}{n}\sum_{i=1}^{n} G_3^*(X_i) \xrightarrow{P} \bar{x} < K.$$

Então, para todo $\epsilon > 0$, temos que

$$P(|B_2 - \bar{x}| < \epsilon) \overset{n\to\infty}{\longrightarrow} 1. \tag{1.7}$$

Tomando $\epsilon = \bar{x} > 0$ em (1.7), segue que

$$P(|B_2| < 2\bar{x}) \overset{n\to\infty}{\longrightarrow} 1,$$

e como $0 < \bar{x} < K$, temos também que

$$P(|B_2| < 2K) \overset{n\to\infty}{\longrightarrow} 1,$$

ou seja, B_2 é limitado.

Dos resultados obtidos em $(II.i)$, $(II.ii)$ e $(II.iii)$, temos que

$$P(|B_0| < a^2, B_1 < -\gamma^2/2, |B_2| < 2K) \overset{n\to\infty}{\longrightarrow} 1.$$

Portanto, para $a > 0$, com probabilidade tendendo a 1 quando $n \to \infty$, temos:
(1º)

$$\frac{1}{n}\frac{\partial}{\partial\theta}\log L(\theta)\bigg|_{\theta=\theta_0+a} = B_0 + aB_1 + \frac{\Delta}{2}a^2B_2 < a^2 - a\gamma^2/2 + a^2K = a[-\gamma^2/2 + a(1+K)].$$

Fazendo $0 < a < \frac{\gamma^2/2}{1+K}$, temos

$$\frac{\partial}{\partial\theta}\log L(\theta) < 0.$$

(2º)

$$\frac{1}{n}\frac{\partial}{\partial\theta}\log L(\theta)\bigg|_{\theta=\theta_0-a} = B_0 + aB_1 + \frac{\Delta}{2}a^2B_2 > -a^2 + a\gamma^2/2 - a^2K = -a[-\gamma^2/2 + a(1+K)],$$

e, analogamente, fazendo $0 < a < \frac{\gamma^2/2}{1+K}$, seque que

$$\frac{\partial}{\partial\theta}\log L(\theta) > 0.$$

Pela condição (C4), a função $\frac{\partial}{\partial\theta}\log L(\theta)$ é contínua para quase todo $\mathbf{x} = (x_1, \ldots, x_n)$, então, pelo que acabamos de mostrar, temos que para $a > 0$ suficientemente pequeno, com probabilidade tendendo a 1 quando $n \to \infty$, existe $\hat{\theta}_n(a) \in I_a = (\theta_0 - a, \theta_0 + a)$ tal que $\hat{\theta}_n(a)$ é uma solução da equação de verossimilhança (1.4) e ponto de máximo local de $L(\theta)$.

Daí, e de (1.1), segue que

$$\lim_{n\to\infty} P(L(\hat{\theta}_n(a)) < L(\theta_0)) = 1,$$

onde θ_0 é o verdadeiro valor de θ.

■

O Lema 1.10 garante que para cada $a > 0$ fixado quando $n \to \infty$ existem, com probabilidade tendendo a 1, máximos locais da função de verossimilhança no intervalo $(\theta_0 - a, \theta_0 + a)$, isto é, numa vizinhança do verdadeiro valor θ_0.

A partir daí, o Teorema 1.11 a seguir mostra que essas soluções convergem para o verdadeiro valor θ_0 quando $n \to \infty$.

Teorema 1.11. *Seja $\boldsymbol{X} = (X_1, \ldots, X_n)$ uma amostra aleatória da densidade $f(x;\theta_0)$ satisfazendo **(C1) - (C6)**. Então, com probabilidade tendendo a 1 quando $n \to \infty$, existem soluções $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ das equações de verossimilhança (1.4) tais que $\hat{\theta}_n$ é um estimador consistente de θ.*

Demonstração. Dado $a > 0$ suficientemente pequeno, pelo Lema 1.10, temos que existe uma sequência $\hat{\theta}_n(a)$ de soluções da equação de verossimilhança (1.4), onde $\hat{\theta}_n \in (\theta_0 - a, \theta_0 + a)$ é um máximo local de $L(\theta)$ e que satisfaz

$$P(|\hat{\theta}_n(a) - \theta_0| < a) \overset{n\to\infty}{\longrightarrow} 1.$$

Agora, fazendo $a \to 0$, seja $\hat{\theta}_n$ a solução de (1.4) mais próxima de θ_0 e, dessa forma, segue que para todo $\epsilon > 0$,

$$P(|\hat{\theta}_n - \theta_0| < \epsilon) \overset{n\to\infty}{\longrightarrow} 1.$$

■

Provamos assim, que sob condições de regularidade, o EMV de um parâmetro é consistente. A seguir verificamos que além da consistência, as condições de regularidade $(C1)-(C6)$ garantem também a normalidade assintótica do EMV.

Teorema 1.12. *Seja $\boldsymbol{X} = (X_1, \ldots, X_n)$ uma amostra aleatória da densidade $f(x; \theta_0)$ satisfazendo as condições (C1)-(C6). Se $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ são soluções consistentes da equação de verossimilhança (1.4), então*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{D}{\longrightarrow} N\left(0, \frac{1}{I(\theta_0)}\right)$$

onde $I(\theta_0) = E\left[\frac{\partial}{\partial\theta} \log f(X;\theta)\Big|_{\theta=\theta_0}\right]^2 = \gamma^2 < \infty.$

Demonstração. Como $\hat{\theta} = \hat{\theta}_n$ é solução consistente da equação de verossimilhança (1.4), da igualdade (1.6), temos que

$$B_0 + (\hat{\theta} - \theta_0)B_1 + \frac{(\hat{\theta} - \theta_0)^2}{2}\Delta B_2 = 0,$$

disto segue que

$$(\hat{\theta} - \theta_0) = \frac{B_0}{-B_1 - \frac{(\hat{\theta}-\theta_0)}{2}\Delta B_2}$$

e portanto

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{\frac{1}{\gamma^2\sqrt{n}}\displaystyle\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log f(X_i;\theta_0)}{\frac{-B_1}{\gamma^2} - \frac{(\hat{\theta}-\theta_0)}{2\gamma^2}\Delta B_2}. \tag{1.8}$$

Na demonstração do Lema 1.10, vimos que $B_1 \overset{P}{\longrightarrow} -\gamma^2$ e B_2 é limitado. Daí segue que o denominador da fração (1.8) converge em probabilidade para 1. Sendo assim, para verificar a normalidade assintótica de $\hat{\theta}_n$ basta verificarmos a convergência do numerador da fração.

Observe que $\frac{\partial}{\partial\theta}\log f(X_i;\theta_0)$ são variáveis i.i.d. com média 0 e variância $\gamma^2 < \infty$, então, pelo Teorema de Lindeberg-Levy,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log f(X_i;\theta_0) \overset{D}{\longrightarrow} N(0, \gamma^2),$$

ou seja,

$$\frac{1}{\gamma^2\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i;\theta_0) \xrightarrow{D} N\left(0,\frac{1}{\gamma^2}\right).$$

Portanto

$$\sqrt{n}(\hat{\theta}_n-\theta_0) \xrightarrow{D} N\left(0,\frac{1}{I(\theta_0)}\right),$$

onde $I(\theta_0)=\gamma^2$. ■

Os Teoremas 1.11 e 1.12 garantem que em amostras completas podemos obter boas estimações para os parâmetros desconhecidos, assegurando importantes propriedades quando essas amostras são suficientemente grandes. Entretanto, ao realizar estudos práticos para obtenção de estimadores, podemos nos deparar com situações de perda de informação, isto é, amostras que geram informações incompletas por algum motivo. Ao se estudar esse tipo de problema, surgiram na literatura novos modelos estatísticos que visavam assegurar à essas amostras com informações perdidas as mesmas propriedades das amostras completas. Nas próximas seções discutiremos alguns desses modelos.

1.3 - Esquemas de Censura

Nesta seção apresentamos esquemas de censura, características de dados observados em testes de estudos de análise de sobrevivência e confiabilidade. Para isso, iniciamos abordando alguns conceitos.

Quando uma indústria fabrica algum produto, é importante que se tenha ideia de quanto tempo aquele produto irá funcionar, sob condições normais de uso. Essa informação se torna necessária para que se saiba, por exemplo, qual o tempo de garantia será oferecido sobre o produto e aproximadamente quantas unidades irão apresentar problemas até esse tempo. Essas ideias estão ligadas ao conceito de confiabilidade. *Confiabilidade* é uma medida da capacidade de um produto funcionar bem durante um período de tempo especificado, sob condições de uso pré-estabelecidas.

As situações estudadas em confiabilidade envolvem o tempo até a ocorrência de um evento de interesse. Na maioria dos casos, esses eventos são indesejáveis, o que nos faz denominá-los como *falhas*.

O tempo decorrido do início do experimento até o evento de interesse (falha) será analisado para responder as questões acerca da confiabilidade do produto, podendo ser denominado por “tempo de falha”, “tempo de sobrevivência”, “tempo de vida”, “tempo até a ocorrência do evento” ou “tempo até falha”. Neste trabalho iremos defini-lo como *tempo de falha*. Observe que esse tempo de falha não é necessariamente uma medida usual de tempo, podendo ser medido em outras escalas, de acordo com o objeto de estudo. Ao estudo acerca do tempo de falha se dá o nome de *Análise de Confiabilidade ou Sobrevivência*.

O interesse em estudos de análise de confiabilidade é observar os tempos de falha das unidades ou itens colocados em teste, e, a partir desses dados, realizar a inferência estatística. Porém, os testes realizados para obter as medidas de durabilidade de produtos, por exemplo, podem ser demorados e caros. Por esses motivos, muitas vezes são terminados antes que todos os itens falhem, gerando observações incompletas para estudo. Nesse caso, dizemos que ocorreu uma *censura*. Mais formalmente, uma *censura* é a observação parcial da resposta do estudo que foi interrompida por alguma razão, não permitindo a observação completa do tempo de falha.

Por exemplo, suponhamos que n itens eletrônicos sejam colocados em teste de vida, de modo que sejam deixados em funcionamento até que falhem, durante um determinado tempo. Se até esse tempo apenas $m < n$ itens deixarem de funcionar, apenas m tempos de falha serão observados, sendo os outros $n - m$ ditos *censurados*.

Apesar de não gerar informações do tempo de falha, as censuras carregam a informação de que esse tempo é maior que o tempo observado no estudo, isto é, a informação da sobrevivência das unidades em teste. Por esse motivo, mesmo que sejam parciais, essas observações não devem ser desconsideradas na análise estatística.

Na prática, temos alguns tipos de censuras. Destacamos aqui as censuras *à direita* e *à esquerda*.

Censura à direita: É o tipo mais comum em testes de análise de confiabilidade. Na censura à direita não se observa o tempo exato da falha, só se sabe que o tempo de falha é maior do que o tempo observado.

Por exemplo, suponhamos que um estudo acompanhe o tempo entre o diagnóstico de AIDS e o óbito, em 200 pessoas, entre os anos de 1996 e 2000. Denominemos como falha a morte da pessoa. Suponhamos ainda que 90 dessas pessoas morreram até o ano 2000 e as outras 110 não. Nesse caso, tivemos 90 falhas observadas e 110 censuras (à direita).

Entre censura à direita existem alguns outros tipos de censura, como a do Tipo I, Tipo II e a Tipo II progressiva, que serão estudadas nas próximas seções. Cada um desses tipos de censura gera uma diferente função de verossimilhança.

Censura à esquerda: Esse tipo de censura ocorre quando não conhecemos o momento da ocorrência da falha, mas sabemos que ela ocorreu antes do tempo registrado. Em outras palavras, o tempo de falha é menor que o tempo observado.

Por exemplo, o estudo do tempo decorrido entre a infecção pelo vírus HIV e o diagnóstico imunológico de AIDS. Consideremos como falha o diagnóstico imunológico. Não é possível saber o momento da falha, apenas que ela ocorreu após a infecção.

No que segue, concentramos nossos estudos nos casos de censura à direita.

1.3.1 Censuras do Tipo I e Tipo II

Apresentamos nesta Seção os dois tipos clássicos de esquemas de censura à direita: a do tipo I e a do tipo II. Esses modelos foram os primeiros modelos de censura que surgiram na literatura, sendo posteriormente generalizados para novos modelos. Motivados pela sua vasta aplicabilidade e pela base para o entendimento da censura tipo II progressiva, seguem detalhes de cada um.

Censura do tipo I

Damos início à apresentação dos esquemas de censura, com a *Censura do Tipo I*, onde a falha é observada apenas se ocorrer antes de um determinado tempo pré-fixado. Estudos sobre comportamentos de animais ou exames clínicos, por exemplo, podem facilmente gerar amostras com esse tipo de censura. Um estudo pode iniciar com um número fixo de animais ou pacientes, onde um tratamento é aplicado e, por questões financeiras ou temporais, o investigador termina seu estudo antes que todos os eventos sejam observados, ou seja, após um determinado tempo. Outros exemplos a serem considerados são testes de durabilidade de itens eletrônicos que demoram muito tempo até que apresentem mau funcionamento. Nesse caso, estipula-se um tempo limite para o teste. Observe que nesse tipo de censura o tempo máximo de observação é fixo, enquanto a quantidade e o tempo de falha dos itens observados são aleatórios.

Definamos formalmente censura do Tipo I.

Definição 1.13. Sejam $T_1, \ldots, T_n$ variáveis aleatórias i.i.d. que caracterizam tempos de falha, com função densidade de probabilidade e função de distribuição comum dadas por $f(\cdot;\theta)$ e $F(\cdot;\theta)$, respectivamente, onde θ é um parâmetro. Seja t_c um tempo pré-determinado, dito *tempo de censura*. Uma *amostra sob esquema de censura do tipo I* é uma amostra $X_1, \ldots, X_n$ dada por

$$X_i = \min(T_i, t_c) = \begin{cases} T_i, \text{se } T_i \leq t_c \\ t_c, \text{se } T_i > t_c \end{cases}.$$

Seja Δ_i a variável aleatória definida por:

$$\Delta_i = \begin{cases} 1, & T_i \leq t_c \\ 0, & T_i > t_c \end{cases} \quad i = 1, \ldots, n.$$

Nessa abordagem, podemos dizer que Δ_i é a variável que indica se o i-ésimo tempo de falha T_i é censurado ou não. Dessa forma, temos que a função densidade de probabilidade conjunta de X_i e Δ_i é dada por

$$f_{X_i,\Delta_i}(t,\delta;\theta) = [f(t;\theta)]^{\delta}[1 - F(t_c;\theta)]^{1-\delta}, \ i = 1,\ldots,n,$$

para $t > 0$ e $\delta = \begin{cases} 1, & t \leq t_c \\ 0, & t > t_c \end{cases}$.

Na presença de censura, a função de verossimilhança é modificada. Em amostras com dados completos, ou seja, sem censura, a função de verossimilhança é o produto das densidades completas, como pode ser visto na Seção 1.2. Na presença de censura tipo I, o tempo de falha T_i é observado apenas de $T_i \leq t_c$, $i = 1,\ldots,n$; e portando, se $T_i > t_c$, não é conhecido exatamente quando a falha ocorre.

Quando uma censura ocorre, o que se sabe é que o intervalo do tempo de falha é (t_c, ∞) e essa informação é uma importante contribuição na função de verossimilhança, que pode ser resumida por $P(T_i > t_c)$.

Agora, se $t_1,\ldots,t_n$ é uma amostra observada de $X \stackrel{D}{=} X_i$, $i = 1,\ldots,n$, temos que a função de verossimilhança de θ, para o caso de censura do tipo I, é dada por:

$$L(\theta) = \prod_{i=1}^{n} [f(t_i;\theta)]^{\delta_i}.[1 - F(t_c;\theta)]^{1-\delta_i}, \tag{1.9}$$

onde $\delta_i = \begin{cases} 1 & \text{se } t_i \leq t_c \\ 0 & \text{se } t_i > t_c \end{cases}$, $i = 1,\ldots,n$, onde notação $\stackrel{D}{=}$ indica "mesma distribuição que".

Maiores detalhes podem ser encontrados em Lawless [24] e Klein [23].

Exemplo 1.14. *Estimação do EMV para uma amostra aleatória de tamanho n com distribuição Exponencial sob censura do tipo I [24].*

O modelo Exponencial é um dos mais utilizados em análises de dados de confiabiliade, pois se adequa a várias situações práticas, como por exemplo, modelagem de tempo de vida de produtos elétricos e óleos isolantes. A distribuição exponencial é a única distribuição absolutamente contínua que possui função de taxa de falha (ou taxa de risco) constante no tempo, propriedade dita como *falta de memória*. Seja T uma variável com distribuição exponencial de parâmetro θ, com funções densidade e de distribuição dadas, respectivamente, por

$$f(t) = \theta e^{-\theta t}, \qquad t \geq 0,$$

e

$$F(t) = 1 - e^{-\theta t}, \qquad t \geq 0.$$

Suponhamos, sem perda de generalidade, que os dados observados $t_1, \ldots, t_n$ são apresentados por

$$t_1 < \ldots < t_m : \text{ dados não-censurados}$$

e

$$t_{m+1} = t_c, \ldots, t_n = t_c : \text{ dados censurados.}$$

Por (1.9), a função de verossimilhança, neste caso, é dada por

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} [\theta e^{-\theta t_i}]^{\delta_i} [e^{-\theta t_i}]^{1-\delta_i} \\ &= \prod_{i=1}^{m} [\theta e^{-\theta t_i}].[e^{-\theta t_c}]^{n-m} \\ &= \theta^m e^{-\theta \sum_{i=1}^{m} t_i} .e^{-\theta t_c(n-m)}, \end{aligned}$$

e daí, a função log-verossimilhança é dada por

$$l(\theta) = m \log \theta - \theta \sum_{i=1}^{m} t_i - \theta t_c(n-m). \tag{1.10}$$

Derivando (1.10) em relação a θ, obtemos a equação de verossimilhança

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{m}{\theta} - \left(\sum_{i=1}^{m} t_i + t_c(n-m) \right) = 0,$$

e resolvendo-a, obtemos o EMV para θ:

$$\hat{\theta} = \frac{m}{\sum_{i=1}^{m} t_i + t_c(n-m)}.$$

Censura do tipo II

Um outro tipo de censura à direita é a *censura do tipo II*, na qual é observado o tempo de falha de um número pré-determinado de itens em teste, isto é, se n unidades forem colocadas em teste, este será finalizado quando ocorrer a m-ésima falha, sendo $m < n$ um número previamente fixado. Censura tipo II é comum em experimentos de teste de vida útil de equipamentos quando se tem pouca ou nenhuma informação sobre sua durabilidade. Neste processo, todos os itens são colocados em teste ao mesmo tempo e o teste é terminado quando é observado um certo número pré-estabelecido de falhas. Tal experiência pode economizar tempo e dinheiro pois, geralmente, são testes de alto custo para realização e que pode-se demorar muito tempo para que os itens falhem.

Definição 1.15. Sejam $T_1, \ldots, T_n$ variáveis aleatórias i.i.d. que caracterizam tempos de falhas, com função densidade de probabilidade e função de distribuição comum dadas por $f(\cdot;\theta)$ e $F(\cdot;\theta)$, respectivamente, onde θ é um parâmetro. Seja $m < n$ o número pré-fixado de falhas observadas. Uma *amostra sob esquema de censura do tipo II* é uma amostra $X_{(1)}, \ldots, X_{(n)}$ tal que $X_{(1)}, \ldots, X_{(n)}$ são estatísticas de ordem definidas por

$$X_{(i)} = \begin{cases} T_{(i)}, & \text{se } T_{(i)} \leq T_{(m)} \\ T_{(m)}, & \text{se } T_{(i)} > T_{(m)}, \end{cases}$$

onde $T_{(m)}$ é o tempo de vida aleatório da m-ésima falha.

Obtemos agora a função de verossimilhança para o parâmetro θ. Considerando $x_{(1)}, \ldots, x_{(n)}$ os valores observados de $X_{(1)}, \ldots, X_{(n)}$, a função de verossimilhança para este modelo com m falhas observadas é dada por

$$L(\theta) = \frac{n!}{(n-m)!} \left[\prod_{i=1}^{m} f(x_{(i)};\theta) \right] .[1 - F(x_{(m)};\theta)]^{n-m}, \tag{1.11}$$

onde $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(m)}$ e $x_{(m+1)} = \cdots = x_{(n)} = x_{(m)}$.

Maiores detalhes podem ser encontrados novamente em Lawless [24] e [23].

Exemplo 1.16. *Estimação dos EMV para uma amostra aleatória de tamanho n com distribuição Gumbel sob censura do tipo II [24].*

A distribuição Gumbel (ou distribuição de Valor Extremo), assim como a Exponencial, é muito utilizada em análise de dados de confiabilidade, por também se adequar a várias situações práticas. Essa distribuição está diretamente relacionada com a distribuição Weibull, pois é obtida considerando-se o logaritmo natural de uma variável aleatória com distribuição Weibull. Mais detalhes dessa relação serão tratados na Seção 2.2.

As funções de densidade de probabilidade e de distribuição acumulada de uma variável X com distribuição Gumbel com parâmetros μ e b (parâmetros de locação e de escala, respectivamente) são dadas por

$$f(x) = \frac{1}{b} \exp\left[\frac{x-\mu}{b} - \exp\left(\frac{x-\mu}{b} \right) \right], \quad -\infty < x < \infty$$

e

$$F(x) = 1 - \exp\left[- \exp\left(\frac{x-\mu}{b} \right) \right],$$

respectivamente, onde $b > 0$ e $-\infty < \mu < \infty$.

Suponhamos que $x_1, \ldots, x_n$ sejam observações ordenadas de uma amostra aleatória $X_1, \ldots, X_n$ com distribuição Gumbel de parâmetro b e μ sob esquema de censura do tipo II com m falhas observadas.

Por (1.11), a função de verossimilhança é dada por

$$\begin{aligned} L(\mu, b) &= \frac{n!}{(n-m)!} \prod_{i=1}^{m} \left[\frac{1}{b} e^{\frac{x_i-\mu}{b}} . \exp\left(-e^{\frac{x_i-\mu}{b}}\right) \right] \left[\exp\left(-e^{\frac{x_m-\mu}{b}}\right) \right]^{n-m} \\ &= \frac{n!}{(n-m)!} \frac{1}{b^m} \exp\left(\sum_{i=1}^{m} \frac{x_i-\mu}{b} \right) . \exp\left(-\sum_{i=1}^{m} e^{\frac{x_i-\mu}{b}} \right) \exp\left(-(n-m) e^{\frac{x_m-\mu}{b}} \right) \\ &= \frac{n!}{(n-m)!} \frac{1}{b^m} \exp\left[\sum_{i=1}^{m} \frac{x_i-\mu}{b} - (n-m) e^{\frac{x_m-\mu}{b}} - \sum_{i=1}^{m} e^{\frac{x_i-\mu}{b}} \right], \end{aligned}$$

e daí, a função log-verossimilhança, por:

$$l(\mu, b) = \log \frac{n!}{(n-m)!} - m \log b + \sum_{i=1}^{m} \frac{x_i-\mu}{b} - (n-m) e^{\frac{x_m-\mu}{b}} - \sum_{i=1}^{m} e^{\frac{x_i-\mu}{b}}. \quad (1.12)$$

Assim, derivando (1.12) em relação a μ e b obtemos as equações de verossimilhança

$$\begin{cases} \frac{\partial l(\mu,b)}{\partial \mu} = \frac{1}{b} \left[-m + (n-m) e^{\frac{x_m-\mu}{b}} + \sum_{i=1}^{m} e^{\frac{x_i-\mu}{b}} \right] = 0 \\ \\ \frac{\partial l(\mu,b)}{\partial b} = \frac{1}{b^2} \left[-mb - \sum_{i=1}^{m} (x_i + \mu) + (n-m)(x_m - \mu) e^{\frac{x_m-\mu}{b}} + \sum_{i=1}^{m} (x_i - \mu) e^{\frac{x_i-\mu}{b}} \right] = 0, \end{cases}$$

e resolvendo-as, obtemos

$$\begin{cases} e^{\mu} = \frac{1}{m^b} \left[(n-m) e^{\frac{x_m}{b}} + \sum_{i=1}^{m} e^{\frac{x_i}{b}} \right], \\ \\ \frac{1}{m} \sum_{i=1}^{m} x_i = \dfrac{(n-m) x_m e^{\frac{x_m}{b}} + \sum_{i=1}^{m} x_i e^{\frac{x_i}{b}}}{(n-m) e^{\frac{x_m}{b}} + \sum_{i=1}^{m} e^{\frac{x_i}{b}}} - b, \end{cases}$$

que nos possibilita encontrar os valores dos EMV $\hat{\mu}$ e $\hat{b}$ por meio de métodos numéricos.

1.3.2 Censura do Tipo II progressiva

Uma generalização da censura do tipo II é a *censura do tipo II progressiva*. Aqui, determina-se um número fixo de falhas a serem observadas no teste de vida e a cada uma delas, retira-se aleatoriamente outras unidades que ainda estão em funcionamento. Esses experimentos também são bastante utilizados em testes de vida útil de equipamentos, de modo a economizar tempo e dinheiro. Essa diferença em relação a censura do tipo II é que aqui, pressupõe-se que ao ocorrer uma falha em um equipamento por algum motivo, outros equipamentos estariam para falhar pelo mesmo motivo e por isso a retirada de itens do teste após cada falha. Isto economiza tempo e custos.

Nesse caso, n unidades são colocadas em teste e deseja-se observar o tempo em que ocorrem as m primeiras falhas, e a cada i-ésima falha são retirados aleatoriamente R_i unidades que ainda estão em funcionamento, onde $R_1, \ldots, R_m$ são valores pré-estabelecidos. Ou seja, quando ocorre a primeira falha, são retiradas R_1 unidades das $(n-1)$ que não falharam, restando então $n - R_1 - 1$ unidades. Da mesma forma, quando ocorre a segunda falha, são retiradas R_2 unidades, restando agora $n - R_1 - R_2 - 2$ unidades, e assim segue o experimento até ocorrer a m-ésima falha. Nesse momento restam $R_m = n - R_1 - R_2 - \cdots - R_{m-1} - m$ unidades em funcionamento e todas, portanto, são censuradas, já que foi atingido o número m de falhas, previamente estabelecido.

Observe que na censura do tipo II progressiva se $R_1 = R_2 = \cdots = R_{m-1} = R_m = 0$, então $n = m$ o que corresponde ao caso sem censura. Agora, se $R_1 = R_2 = \cdots = R_{m-1} = 0$, então $R_m = n - m$, o que corresponde à censura do tipo II convencional.

Sejam T_{ij}, $i = 1, \ldots, m$ e $j = 0, 1, \ldots, R_i$, variáveis aleatórias i.i.d. que caracterizam tempos de falha, com função densidade de probabilidade e função de distribuição comum dadas por $f(\cdot;\theta)$ e $F(\cdot;\theta)$ respectivamente, onde θ é um parâmetro e $m + \sum_{i=i}^{m} R_i = n$. Seja $m < n$ o número pré-fixado de falhas observadas e $R_1, \ldots, R_m$ o esquema de censura pré-determinado. Uma *amostra sob esquema de censura do tipo II* é uma amostra $X_1, \ldots, X_n$ tal que os m tempos de falha observados são dados por

$$X_i = T_{i0}, \quad i = 1, \ldots, m,$$

e os tempos censurados são dados por

$$X_{ij} = T_{i0}, \quad j = 1, \ldots, R_i,$$

onde X_{ij} representa o tempo observado das R_i unidades retiradas após a i-ésima falha.

Sejam então $X_1, \ldots, X_m$ as variáveis aleatórias relativas aos tempos de falha ordenados das m unidades que falharam, denominados *estatísticas de ordem censuradas progressiva-*

mente, e $x_1, \ldots, x_m$ suas respectivas observações. Para obter a densidade conjunta dessas estatísticas de ordem, como apresentam Balakrishnan e Aggarwala em [3], observe que para a primeira observação de falha, a probabilidade de $X_1 = x_1$ é dada por

$$P(X_1 = x_1) = nf(x_1;\theta)[1 - F(x_1;\theta)]^{n-1},$$

já que uma das n unidades falhou e outras $n-1$ não. Após essa falha, serão retiradas R_1 unidades da amostra.

Para a segunda observação de falha, uma das $n - R_1 - 1$ unidades, que ainda estão em funcionamento, irá falhar. Então, temos que a probabilidade de $X_2 = x_2$, dado que $X_1 = x_1$ e que R_1 unidades foram retiradas, é

$$P(X_2 = x_2 | x_1, R_1) = (n - R_1 - 1)f(x_2;\theta)\frac{[1 - F(x_2;\theta)]^{n-R_1-2}}{[1 - F(x_1;\theta)]^{n-R_1-1}}.$$

Seguimos esse raciocínio até a m-ésima observação de falha, obtendo

$$\begin{aligned}&P(X_m = x_m | x_1, R_1; \ldots; x_{m-1}, R_{m-1}) = \\ &(n - R_1 - \cdots - R_{m-1} - m + 1)f(x_m;\theta)\frac{[1 - F(x_m;\theta)]^{n-R_1-\cdots-R_{m-1}-m}}{[1 - F(x_{m-1};\theta)]^{n-R_1-\cdots-R_{m-1}-m+1}}.\end{aligned}$$

Como $X_1, \ldots, X_m$ são i.i.d., sua função de densidade conjunta fica

$$\begin{aligned}&f_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = \\ &P(X_1 = x_1).P(X_2 = x_2|x_1, R_1)\cdots P(X_m = x_m|x_1, R_1; \ldots; x_{m-1}, R_{m-1}),\end{aligned}$$

e então, pelas probabilidades já definidas, segue que

$$f_{X_1,\ldots,X_m}(x_1, \ldots, x_m;\theta) = C\prod_{i=1}^{m} f(x_i;\theta)[1 - F(x_i;\theta)]^{R_i},$$

onde $x_1 < \ldots < x_m$ e $C = n(n - R_1 - 1)(n - R_1 - R_2 - 2)\cdots(n - R_1 - \cdots - R_{m-1} - m + 1)$, ou seja, C é a constante normalizadora e representa o número de maneiras que m estatísticas de ordem censuradas do tipo II progressivamente podem ocorrer.

Segue então que a função de verossimilhança para o parâmetro θ baseada na amostra sob censura do tipo II progressiva é dada por

$$L(\theta) = C\prod_{i=1}^{m} f(x_i;\theta)[1 - F(x_i;\theta)]^{R_i}. \tag{1.13}$$

Maiores detalhes podem ser encontrados em Balakrishnan [2] e [3].

Observação 1.17. No Capítulo 3 estudaremos propriedades assintóticas do estimador de máxima verossimilhança de θ, baseado em esquemas de censura do tipo II progressiva. Nesse caso, quando o tamanho da amostra é suficientemente grande, o número de falhas observadas deve ser proporcional ao número de unidades da amostra, isto é, $\lim\limits_{n\to\infty} \dfrac{m}{n} = \tau$.

Além disso, assumimos que os números de unidades censuradas R_i, apesar de fixados, também tenham proporção fixa quando n tende a infinito, isto é, $\lim\limits_{n\to\infty} \dfrac{R_i}{n} = \tau_i$.

Dessa forma, quando $n \to \infty$, essas proporções são tais que $\sum\limits_{i=1}^{n} \tau_i \overset{n\to\infty}{\longrightarrow} 1 - \tau$, onde τ é a proporção das unidades observadas e $1 - \tau$ é a proporção das unidades censuradas.

A seguir, enunciamos um importante princípio no desenvolvimento de estudos de análise de sobrevivência com dados censurados.

Observação 1.18. Princípio da Informação Perdida - (Tanner [32] e Louis [27])

A informação sobre dados incompletos pode ser decomposta como:

$$\text{Informação Observada} = \text{Informação Completa} - \text{Informação Perdida}$$

É importante ressaltar que no caso de amostras com censura, os dados censurados não são necessariamente perdidos, pois a informação de sua sobrevivência até o tempo da censura deve ser levada em conta na análise estatística. Em todo caso, esse princípio será fundamental para obtermos as propriedades assintóticas do EMV do parâmetro em amostras sob censura do tipo II progressiva, abordadas nas Seções 3.2 e 3.3.

No Capítulo 2 serão apresentados detalhadamente exemplos de amostras sob censura tipo II progressiva, onde são obtidos os EMV's a partir de sua função de verossimilhança. Em um dos casos, podemos encontrar esses estimadores utilizando o algoritmo EM, que será apresentado na próxima seção. Esse algoritmo é uma importante ferramenta para a estimação de parâmetros em amostras com dados incompletos, como por exemplo, as amostras sob censura que apresentamos nesta seção.

1.4 - O Algoritmo EM

O algoritmo EM (Expectation-Maximization) proposto por Dempster et al. [15] é um processo iterativo que tem como objetivo encontrar o EMV para amostras aleatórias quando, por algum motivo, os dados observados forem considerados incompletos. Isto é, seja Y uma variável aleatória com função de densidade $f_Y(y;\theta)$, onde $\theta \in \Theta \subset \mathbb{R}^r$ é um parâmetro desconhecido. Queremos obter o EMV para θ quando $Y = y$ é observado, porém, a maximização da verossimilhança do dado observado é complexa. Buscamos outra forma de resolver esse problema, ou diminuir sua dificuldade introduzindo um dado latente z, de modo que $c = (y, z)$ seja tratado como um dado completo e a maximização de $f_{Y,Z}(y, z; \theta) = f_C(c; \theta)$ seja mais simples.

Porém, a verossimilhança completa $f_{Y,Z}(y, z; \theta)$ nos fornece um estimador que depende do valor introduzido z, o que não faz sentido. Sendo assim, a proposta do algoritmo EM consiste em calcular a esperança de $\log f_{Y,Z}(y, z; \theta)$ com respeito à densidade preditiva da variável latente Z dado $Y = y$, para um valor ajustado de θ, $f_{Z|Y}(z|y; \theta)$ e em seguida, atualizar um novo valor para o parâmetro θ, maximizando tal esperança com respeito a θ. Estas são basicamente as duas etapas do algoritmo EM, denominadas passo E (expectation) e passo M (maximization), respectivamente. Formalizemos, então, o algoritmo EM.

Seja Y um vetor aleatório com valores em $E_Y \subseteq \mathbb{R}^l$ e densidade $f_Y(y; \theta)$ relativa a medida μ_Y, onde $\theta \in \Theta \subset \mathbb{R}^r$ é um parâmetro desconhecido. O objetivo é encontrar o EMV de θ quando $Y = y$ é observado, ou seja, obter $\theta_{\max} = argmax_\theta f_Y(y; \theta)$. Nas situações em que este cálculo é complexo, uma alternativa é considerar o dado observado y como um dado incompleto e completá-lo introduzindo um dado não observado z, de modo que o cálculo de $\theta_{\max} = argmax_\theta f_{Y,Z}(y, z; \theta)$ seja mais simples. Formalmente, seja Z um vetor aleatório com valores em $E \in \mathbb{R}^p$, μ_y uma medida σ-finita em $(E, \mathscr{E})$, onde $\mathscr{E}$ é uma σ-álgebra de subconjuntos de E tais que

$$f_Y(y; \theta) = \sum_E f_{Y,Z}(y, z; \theta) \mu_y(dz).$$

Neste caso, dado $Y = y$ a variável Z possui uma densidade condicional (preditiva) relativa a medida μ_y dada por

$$f_{Z|Y}(z|y; \theta) = \frac{f_{Y,Z}(y, z; \theta)}{f_Y(y; \theta)}.$$

A estimativa de $\theta_{\max}$ não deve depender do valor acrescido z e a proposta do algoritmo EM é a substituição do cálculo do $argmax_\theta f_{Y,Z}(y, z; \theta)$ pela maximização do valor esperado de $\log f_{Y,Z}(y, z; \theta)$ relativa a distribuição preditiva,

$$\begin{aligned} Q(\theta, \theta') &= \int_E [\log f_{Y,Z}(y, z; \theta)] f_{Z|Y}(z|y; \theta') \mu_y(dz) \\ &= E[\log f_{Y,Z}(Y, Z; \theta) | \theta', y]. \end{aligned} \tag{1.14}$$

Cada iteração do algoritmo EM consiste em dois passos: o cálculo de $Q(\theta, \theta')$ (passo E) e a determinação de $argmax_\theta Q(\theta, \theta')$ (passo M).

Algoritmo EM. Observa-se o dado $Y = y$ e seleciona-se um valor inicial de $\theta(0) \in \Theta$. Na iteração $k + 1$, *passo E*: calcula-se a esperança

$$\begin{aligned} Q(\theta, \theta(k)) &= \int_E [\log f_{Y,Z}(y, z; \theta)] f_{Z|Y}(z|y; \theta(k)) \mu_y(dz) \\ &= E[\log f_{Y,Z}(Y, Z; \theta) | \theta(k), y] \end{aligned}$$

passo M: determina-se

$$\theta(k+1) = argmax_\theta Q(\theta, \theta(k)),$$

isto é, $\theta(k+1)$ tal que $Q(\theta(k+1), \theta(k)) \geq Q(\theta, \theta(k))$.

Para entender a aplicabilidade do algoritmo, segue um exemplo clássico, exibido por Rao [29].

Exemplo 1.19. *(Modelo do Elo Genético)* Assuma que temos 4 categorias de animais e que Y_i representa o número de animais da categoria i numa amostra de tamanho r. Suponha que $Y = (Y_1, Y_2, Y_3, Y_4)$ tem distribuição multinomial com probabilidades $\left(\frac{1}{2} + \frac{\theta}{4}, \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{\theta}{4}\right)$ e $\theta \in (0, 1)$. Para o dado observado $y = (y_1, y_2, y_3, y_4)$, temos a densidade

$$f_Y(y; \theta) = \binom{r}{y_1\ y_2\ y_3\ y_4} \left(\frac{1}{4}\right)^r (2+\theta)^{y_1} (1-\theta)^{y_2+y_3} \theta^{y_4},$$

e portanto, a menos de uma constante de proporcionalidade independente de θ, temos que

$$f_Y(y; \theta) \propto (2+\theta)^{y_1} (1-\theta)^{y_2+y_3} \theta^{y_4}, \tag{1.15}$$

e segue então que

$$\log f_Y(y; \theta) \propto y_1 \log(2+\theta) + (y_2 + y_3) \log(1-\theta) + y_4 \log \theta. \tag{1.16}$$

O cálculo de $\theta_{\max}$ em (1.15) é complexo, pois ao derivar (1.16) em relação a θ e igualar a 0, recaímos na equação do segundo grau

$$\frac{\partial}{\partial \theta} \log f_Y(y; \theta) = r\theta^2 - [y_1 - 2(y_2 + y_3) - y_4]\theta - 2y_4 = 0.$$

Este cálculo pode ser simplificado introduzindo uma variável latente $Z = z$ de modo que a primeira categoria seja subdividida em duas categorias com probabilidades $\frac{1}{2}$ e $\frac{\theta}{4}$. Neste caso, o dado completo é $(y_1 - z, z, y_2, y_3, y_4)$ e a densidade completa dada por

$$f_{Y,Z}(y, z; \theta) = \binom{r}{(y_1 - z)\ z\ y_2\ y_3\ y_4} \left(\frac{1}{4}\right)^r 2^{y_1 - z} \theta^{z + y_4} (1-\theta)^{y_2+y_3},$$

isto é, a menos de uma constante de proporcionalidade,

$$f_{Y,Z}(y,z;\theta) \propto \theta^{z+y_4}(1-\theta)^{y_2+y_3}, \tag{1.17}$$

e então segue que

$$\log f_{Y,Z}(y,z;\theta) \propto (z+y_4)\log\theta + (y_2+y_3)\log(1-\theta). \tag{1.18}$$

Ao derivar (1.18) em relação a θ e igualar a 0, a maximizição de (1.17) resulta na solução da equação do primeiro grau

$$\theta(r-y_1+z)-(z+y_4)=0.$$

Para melhor ilustrar a notação utilizada no algoritmo EM, temos para este exemplo: $E=\{0,1,\ldots,r\}$, μ_y é a medida contadora em $\{0,1,\ldots,y_1\}$ e

$$f_{Z|Y}(z|y;\theta) = \binom{y_1}{z}\left(\frac{\theta}{2+\theta}\right)^{z}\left(\frac{2}{2+\theta}\right)^{y_1-z}.$$

Pela definição dada em (1.14) temos que

$$\begin{aligned}
Q(\theta,\theta(k)) &= \int_E [\log f_{Y,Z}(y,z;\theta)] f_{Z|Y}(z|y;\theta(k))\mu_y(dz) \\
&= \int_E [(z+y_4)\log\theta + (y_2+y_3)\log(1-\theta)] f_{Z|Y}(z|y;\theta(k))\mu_y(dz) \\
&= y_4\log\theta + (y_2+y_3)\log(1-\theta) + \log\theta\left[\sum_{z=0}^{y_1} z f_{Z|Y}(z|y;\theta(k))\right] + cte \\
&= \left(y_4 + y_1\frac{\theta(k)}{2+\theta(k)}\right)\log\theta + (y_2+y_3)\log(1-\theta), \qquad (1.19)
\end{aligned}$$

e $\theta(k+1) = argmax_\theta Q(\theta,\theta(k))$. Então, derivando (1.19) em relação a θ e igualando a 0 obtemos

$$\theta(k+1) = \frac{(y_1+y_4)\theta(k)+2y_4}{r\theta(k)+2(r-y_1)}.$$

A grande popularidade do algoritmo EM se dá pelo fato de que, sob determinadas condições, garante que $\theta(k) \overset{k\to\infty}{\longrightarrow} \theta_{max}$. Este resultado foi dado por Wu [ref] e enunciamos aqui no Teorema 1.21. Primeiramente, note que:

Lema 1.20. *Se* $Q(\theta,\theta') \geq Q(\theta',\theta')$ *então* $f_Y(y;\theta) \geq f_Y(y;\theta')$.

Demonstração. Temos que

$$\log \frac{f_Y(y;\theta)}{f_Y(y;\theta')} = \log\left(\frac{f_{Y,Z}(y,z;\theta)}{f_{Z|Y}(z|y;\theta)}.\frac{f_{Z|Y}(z|y;\theta')}{f_{Y,Z}(y,z;\theta')}\right),$$

e como $\int_E f_{Z|Y}(z|y;\theta')dz = 1$, segue que

$$\begin{aligned}\log \frac{f_Y(y;\theta)}{f_Y(y;\theta')} &= \int_E \left[\log \frac{f_Y(y;\theta)}{f_Y(y;\theta')}\right] f_{Z|Y}(z|y;\theta')\mu_y(dz)\\ &= Q(\theta,\theta') - Q(\theta',\theta') - \int_E \left[\log \frac{f_{Z|Y}(z|y;\theta)}{f_{Z|Y}(z|y;\theta')}\right] f_{Z|Y}(z|y;\theta')\mu_y(dz).\end{aligned}$$

Pela desigualdade de Jensen,

$$\int_E \left[\log \frac{f_{Z|Y}(z|y;\theta)}{f_{Z|Y}(z|y;\theta')}\right] f_{Z|Y}(z|y;\theta')\mu_y(dz) \leq \log\left[\int_E f_{Z|Y}(z|y;\theta)\mu_y(dz)\right] = 0.$$

Dessa forma, se

$$Q(\theta(k+1),\theta(k)) \geq Q(\theta(k),\theta(k-1)) \geq \ldots \geq Q(\theta(1),\theta(0)),$$

temos que

$$\log \frac{f_Y(y;\theta(k+1))}{f_Y(y;\theta(k))} \geq 0,$$

e portanto

$$f_Y(y;\theta(k+1)) \geq f_Y(y;\theta(k)) \geq \ldots \geq f_Y(y;\theta(0)).$$

■

Teorema 1.21. *(Wu [ref]). Seja $f_Y(y;\theta)$ contínua em Θ e diferenciável em $int(\Theta)$. Seja $\{\theta(k)\}_{k\geq 0}$ a sequência gerada pelo algoritmo EM com $\theta(0)$ tal que $f_Y(y;\theta(0)) > 0$ e tal que o conjunto $\{\theta : \theta \in \Theta, f_Y(y;\theta) > f_Y(y;\theta(0))\}$ seja compacto. Assuma que para todo k tal que $\theta(k) \in \mathscr{L}^{\mathscr{C}}$, onde $\mathscr{L} = \left\{\theta : \theta \in \Theta \frac{\partial}{\partial\theta'} f_Y(y;\theta')\Big|_{\theta'=\theta} = 0, \right\}$, $f_Y(y;\theta(k+1)) > f_Y(y;\theta(k))$ e que se $\theta(k) \stackrel{k\to\infty}{\longrightarrow} \theta_*$ com $\theta_* \in \mathscr{L}^{\mathscr{C}}$, então $\theta_* \in M(\theta_*)$, onde $M(\theta') = \{\theta : \theta = argmax_\rho Q(\rho,\theta')\}$. Neste caso, toda subsequência convergente de $\{\theta(k)\}_{k\geq 0}$ converge para um limite em $\mathscr{L}$ e $\{f_Y(y;\theta(k))\}_{k\geq 0}$ converge para $f_Y(y;\theta_*)$ com $\theta_* \in \mathscr{L}$.*

O Teorema 1.21 exibe as condições suficientes para a convergência da sequência $\{\theta(k)\}$ obtida pelo algoritmo EM para um algum ponto crítico ou máximo local de $f_Y(y;\theta)$. Maiores detalhes podem ser encontrados em Dempster et al. [ref] e Wu[ref].

Veremos um exemplo dessa aplicação na Seção 2.4, onde obtemos, pelo uso do algoritmo EM, os EMV's dos parâmetros de um modelo lognormal com 3 parâmetros censurado tipo II progressivamente.

Capítulo 2

Modelos de Censura Tipo II Progressiva

2.1 - Introdução

A primeira discussão acerca de estimação de parâmetros com amostras progressivamente censuradas foi feita por Herd em [20], onde fez referência a essas amostras como "multi-censuradas". Alguns anos depois, Cohen, em [12], discutiu a importância da censura progressiva em testes de confiabilidade de tempo de falha. Posteriormente, em [11], Cohen sugere que a censura progressiva é a metodologia ideal para modelar exemplos práticos em que unidades são perdidas da amostras por motivos não esperados ou alheios ao teste. A partir daí, surgem na literatura trabalhos que tratam de modelos de distribuições específicas sob censura progressiva, em particular, sob censura tipo II progressiva. Neste capítulo apresentamos três desses modelos.

As distribuições de Valor Extremo e Weibull compõem um importante papel na modelagem de dados de diversos experimentos. Em especial, a distribuição Weibull é amplamente usada para descrever tempos de vida de produtos industriais, como por exemplo componentes eletrônicos, cerâmicas, capacitores, etc. Essa aplicabilidade se dá devido sua grande variedade de formas e sua propriedade de que a taxa de falha é monótona. A distribuição de Valor Extremo Gumbel está estritamente ligada à distribuição Weibull, pois é obtida aplicando-se o logaritmo natural em uma variável aleatória com distribuição Weibull, isto é, se X é uma variável com distribuição Weibull então $\log X$ possui distribuição de Valor Extremo Gumbel. Essa relação se torna uma importante ferramenta para a estimação de parâmetros em modelos paramétricos definidos por ambas distribuições, pois alguns pontos de difícil trato encontrados em um modelo podem ser simplificados no outro. Com essa ideia, Ding e Yu, em [16], obtém os estimadores de máxima verossimilhança para os parâmetros, considerando uma amostra aleatória relativa a tempos de vida com distribuição de Valor de Extremo Gumbel censurada tipo II progressivamente, convertendo o modelo de Valor Extremo para um modelo Weibull, o que simplifica o tratamento dos dados e, consequentemente, a estimação dos parâmetros para tal modelo. Na Seção 2.2 exibimos os estudos desenvolvidos por esses autores.

Na Seção 2.3, tratamos de uma amostra aleatória relativa a tempos de vida com distribuição Exponencial Generalizada sob testes parcialmente acelerados. Distribuições exponenciais são largamente usadas para modelar tempos de vida, principalmente em testes de vida de componentes eletrônicos. Entretanto, esse tipo de teste pode ter longa duração, acarretando a necessidade de um alto investimento financeiro. Para tentar solucionar esse problema, podem ser aplicados fatores de aceleração que diminuem o tempo de falha das unidades em teste, sem que se comprometa a qualidade dos dados obtidos. Motivados por esses casos práticos, exibimos o trabalho de Ismail em [22], onde o autor associa os estudos de censura tipo II progressiva e de testes parcialmente acelerados, construindo um modelo com distribuição Exponencial Generalizada com essas duas características. Após a obtenção do modelo, são obtidos os estimadores de máxima verossimilhança para seus parâmetros.

Na Seção 2.4 tratamos, de uma amostra aleatória relativa a tempos de vida com distribuição Lognormal com 3 parâmetros. Assim como a distribuição Weibull, distribuições lognormais são usadas para caracterizar tempo de vida de produtos eletrônicos e outros materiais, como semicondutores, por exemplo. Basak e Balakrishnan apresentam em [6], um modelo da distribuição Lognormal com 3 parâmetros censurado tipo II progressivamente e obtém os estimadores para os parâmetros utilizando o Princípio de Máxima Verossimilhança. Além disso, esses autores exibem um método numérico para a obtenção dos estimadores do modelo, construído com o uso do algoritmo EM. Os detalhes do desenvolvimento do trabalho desses autores serão apresentados e discutidos.

2.2 - Inferência para a Distribuição de Valor Extremo Gumbel

Nesta seção abordaremos a inferência estatística considerando uma amostra aleatória com distribuição de Valor Extremo sob censura do tipo II progressiva. Tomaremos como base o trabalho de Ding e Yu [16]. Estudos a respeito desse tipo de amostra já haviam sido desenvolvidos por outros autores, como por exemplo Balakrishnan et al. em [5], onde foram deduzidos alguns valores aproximados para o EMV usando a expansão de Taylor e aplicando processos numéricos de iteração a partir de valores iniciais. Ding e Yu conduziram sua análise sob outro ponto de vista, convertendo o modelo de Valor Extremo para um modelo Weibull, o que simplificou significativamente a complexidade dos algoritmos usados por Balakrishnan em [5].

O modelo Weibull oferece vantagens ao ser empregado, por facilitar a obtenção dos estimadores, sobretudo quando pretende-se calcular o estimador de apenas um parâmetro do modelo.

O desenvolvimento dos estudos apresentados em Ding e Yu são exibidos a seguir.

Consideremos uma amostra aleatória de tamanho n relativa a um teste de vida sob esquema de censura do tipo II progressiva com m falhas observadas. Suponhamos que os tempos de falha $X_1, \ldots, X_n$ possuem distribuição comum de Valor Extremo Gumbel, com função densidade de probabilidade dada por

$$\begin{aligned} f_X(x;\sigma,\mu) &= \frac{1}{\sigma}\exp\left(\frac{x-\mu}{\sigma}\right)\exp\left\{-\exp\left(\frac{x-\mu}{\sigma}\right)\right\}, \\ &= \frac{1}{\sigma}.(e^x)^{\frac{1}{\sigma}}.e^{-\frac{\mu}{\sigma}}.\exp\left\{-(e^x)^{\frac{1}{\sigma}}.e^{-\frac{\mu}{\sigma}}\right\}, \qquad -\infty < x < +\infty. \end{aligned}$$

Se uma variável aleatória Y tem distribuição Weibull, então $X = \log Y$ possui distribuição de Valor Extremo, [24]. Dessa forma, definindo $Y = e^X$ segue que Y possui distribuição Weibull com função densidade de probabilidade dada por

$$g(y;\sigma,\mu) = \frac{1}{\sigma}.(y)^{\frac{1}{\sigma}}.e^{-\frac{\mu}{\sigma}}.\exp\left\{-(y)^{\frac{1}{\sigma}}.e^{-\frac{\mu}{\sigma}}\right\}. \tag{2.1}$$

Fazendo $\alpha = e^{\mu}$ e $\delta = 1/\sigma$ reescrevermos (2.1) como

$$g(y,\alpha,\delta) = \frac{\delta}{\alpha}\left(\frac{y}{\alpha}\right)^{\delta-1}\exp\left\{-\left(\frac{y}{\alpha}\right)^{\delta}\right\}, \quad y > 0, \tag{2.2}$$

ou seja, Y tem distribuição Weibull com parâmetros δ e α. Além disso, sua função de sobrevivência é dada por

$$1 - G(y;\alpha,\delta) = \exp\left\{-\left(\frac{y}{\alpha}\right)^{\delta}\right\},$$

onde $G(y;\alpha,\delta)$ é a função de distribuição acumulada de Y.

Sejam $x_1, \ldots, x_m$ os tempos de falha observados da amostra e sejam $R_1, \ldots, R_m$ as censuras correspondentes ao esquema tipo II progressivo. Definamos $y_i = e^{x_i}$, então $Y_1, \ldots, Y_m$ é uma amostra censurada tipo II progressivamente com distribuição Weibull e função densidade de probabilidade dada por (2.2), onde $y_i, \ldots, y_m$ são suas respectivas observações ordenadas.

Por (1.13), temos que a função de verossimilhança para α e δ é dada por

$$\begin{aligned} L(\alpha,\delta) &= C.\prod_{i=1}^{m} g(y_i).[1-G(y_i)]^{R_i} \\ &= C.\prod_{i=1}^{m} \frac{\delta}{\alpha}\left(\frac{y_i}{\alpha}\right)^{\delta-1} \exp\left\{-\left(\frac{y_i}{\alpha}\right)^{\delta}\right\}\left(\exp\left\{-\left(\frac{y_i}{\alpha}\right)^{\delta}\right\}\right)^{R_i} \\ &= C.\frac{\delta^m}{\alpha^{m\delta}}.\prod_{i=1}^{m} y_i^{\delta-1}\exp\left\{-(1+R_i)\left(\frac{y_i}{\alpha}\right)^{\delta}\right\}, \end{aligned}$$

onde $C = n(n-R_1-1)\cdots(n-R_1-\cdots R_{m-1}-m+1)$ é constante normalizadora, e daí, obtemos facilmente a função log-verossimilhança

$$l(\alpha,\delta) = \log L(\alpha,\delta) = \log C + m\log\delta - m\delta\log\alpha + \sum_{i=1}^{m}\left\{(\delta-1)\log y_i - (R_i+1).\left(\frac{y_1}{\alpha}\right)^{\delta}\right\}.$$

Para encontrar os EMV's $\hat{\alpha}$ e $\hat{\delta}$ para α e δ, respectivamente, devemos encontrar as soluções das equações de verossimilhança, dadas por

$$\begin{cases} \frac{\partial l(\alpha,\delta)}{\partial\alpha} = -\frac{m\delta}{\alpha} + \sum_{i=1}^{m} -(R_i-1)y_i^{\delta}.(-\delta)\alpha^{-\delta-1} = \frac{\delta}{\alpha}\left[\sum_{i=1}^{m}(R_i-1)\left(\frac{y_i}{\alpha}\right)^{\delta} - m\right] = 0 \\ \frac{\partial l(\alpha,\delta)}{\partial\delta} = \frac{m}{\delta} - m\log\alpha + \sum_{i=1}^{m}\left\{\log y_i - (R_i+1).\left(\frac{y_i}{\alpha}\right)^{\delta}.\log\frac{y_i}{\alpha}\right\} = 0, \end{cases}$$

e utilizar a Hessiana de $l(\alpha,\delta)$ para verificar que tais soluções são máximos locais da função log-verossimilhança, e portanto, da função de verossimilhança.

Primeiramente, como $\delta \neq 0$,

$$\begin{aligned} \frac{\partial l(\alpha,\delta)}{\partial\alpha} = 0 \ &\Leftrightarrow \ \sum_{i=1}^{m}(R_i-1)\left(\frac{y_i}{\alpha}\right)^{\delta} - m = 0 \\ &\Leftrightarrow \ \alpha^{\delta} = \frac{1}{m}\sum_{i=1}^{m}(R_i-1){y_i}^{\delta} \\ &\Leftrightarrow \ \alpha = \left[\frac{1}{m}\sum_{i=1}^{m}(R_i-1){y_i}^{\delta}\right]^{\frac{1}{\delta}}. \end{aligned} \tag{2.3}$$

E assim,

$$\frac{\partial l(\alpha,\delta)}{\partial \delta} = 0 \quad \Leftrightarrow \quad \frac{1}{\delta} - \log\alpha + \frac{1}{m}\sum_{i=1}^{m}\log y_i - \frac{1}{m\alpha^{\delta}}\sum_{i=1}^{m}(R_i+1).{y_i}^{\delta}.[\log y_i - \log\alpha] = 0$$

$$\Leftrightarrow \quad \frac{1}{\delta} - \log\alpha + \frac{1}{m}\sum_{i=1}^{m}\log y_i - \frac{\sum_{i=1}^{m}(R_i+1).{y_i}^{\delta}.\log y_i}{\sum_{i=1}^{m}(R_i+1){y_i}^{\delta}} + \frac{\sum_{i=1}^{m}(R_i+1).{y_i}^{\delta}\log\alpha}{\sum_{i=1}^{m}(R_i+1){y_i}^{\delta}} = 0$$

$$\Leftrightarrow \quad \frac{\sum_{i=1}^{m}(R_i+1).y_i^{\delta}.\log y_i}{\sum_{i=1}^{m}(R_i+1).y_i^{\delta}} - \frac{1}{\delta} = \frac{1}{m}\sum_{i=1}^{m}\log y_i \tag{2.4}$$

Os EMV's, $\hat{\alpha}$ e $\hat{\delta}$, são os valores de α e δ respectivamente, que resolvem (2.3) e (2.4) conjuntamente. Eles podem ser obtidos por meio do uso de métodos numéricos, como Newton-Raphson, por exemplo.

Como $\alpha = e^{\mu}$ e $\delta = 1/\sigma$, então $\mu = \log\alpha$ e $\sigma = 1/\delta$. Dessa forma, os EMV's para esses dois últimos parâmetros são dados por

$$\hat{\mu} = \log\hat{\alpha} \text{ e } \hat{\sigma} = 1/\hat{\delta}.$$

Além disso, obtemos a Matriz de Informação de Fisher de (α, δ), dada por

$$I(\alpha,\delta) = \begin{pmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{pmatrix},$$

onde

$$I_{11} = E\left(-\frac{\partial^2 \log L}{\partial \alpha^2}\right) = \frac{\delta}{\alpha^2}\left[(\delta+1).E\left(\sum_{i=1}^{m}(R_i+1)\left(\frac{Y_i}{\alpha}\right)^{\delta}\right) - m\right],$$

$$I_{22} = E\left(-\frac{\partial^2 \log L}{\partial \delta^2}\right) = \frac{m}{\delta^2} + E\left[\sum_{i=1}^{m}(R_i+1)\left(\frac{Y_i}{\alpha}\right)^{\delta}\log^2\left(\frac{Y_i}{\alpha}\right)\right],$$

$$I_{12} = E\left(-\frac{\partial^2 \log L}{\partial \alpha\partial \delta}\right) = \frac{1}{\alpha}\left\{m - E\left[\sum_{i=1}^{m}(R_i+1)\left(\frac{Y_i}{\alpha}\right)^{\delta}\right] - \delta.E\left[\sum_{i=1}^{m}(R_i+1)\left(\frac{Y_i}{\alpha}\right)^{\delta}\log^2\left(\frac{Y_i}{\alpha}\right)\right]\right\},$$

e a matriz de variância-covariância assintótica de $(\hat{\alpha}, \hat{\delta})$ é dada por

$$G(\hat{\alpha},\hat{\delta}) = I^{-1}(\alpha,\delta).$$

A matriz de variância-covariância dos EMV é utilizada para obtenção dos intervalos de confiança para tais estimadores, como pode ser visto em [16].

2.3 - Inferência para a Distribuição Exponencial Generalizada sob Testes Parcialmente Acelerados

Nesta seção trataremos da inferência estatística considerando uma amostra aleatória com distribuição exponencial generalizada em testes parcialmente acelerados sob censura do tipo II progressiva, tendo como objetivo construir um modelo estatístico para esse caso. A construção desse modelo é apresentada por Ismail em [22] e a exibiremos nessa Seção.

A distribuição exponencial generalizada foi introduzida por Gupta e Kundu em [19], com a ideia de contrapor algumas desvantagens encontradas nas distribuições Gama e Weibull e desde então, tem sido amplamente estudada. Sua estrutura simples permite que seja aplicada efetivamente na modelagem de diversos testes de vida, dentre os quais, testes que envolvem censuras.

A família da Exponencial Generalizada com dois parâmetros tem função densidade de probabilidade e função de distribuição, respectivamente, dadas por

$$f(y;\alpha,\theta) \;=\; \alpha\theta(1-e^{-\alpha y})^{\theta-1}e^{-\alpha y}, \quad y>0, \tag{2.5}$$

e

$$F(y;\alpha,\theta) \;=\; (1-e^{-\alpha y})^{\theta}, \qquad y>0, \tag{2.6}$$

onde $\theta > 0$ e $\alpha > 0$ são os parâmetros de forma e escala, respectivamente.

Essa distribuição possui também várias interpretações físicas. Por exemplo, consideremos um sistema paralelo formado por n componentes, isto é, um sistema que funciona somente quando pelo menos uma das n componentes funciona. Se a distribuição dos tempos de vida das componentes são variáveis aleatórias exponenciais i.i.d., então a distribuição do tempo de vida do sistema é dada por

$$F(y;\alpha,n) = (1-e^{-\alpha y})^{n}, \qquad y>0,$$

que representa claramente a distribuição de uma exponencial generalizada com $\theta = n$.

Com a constante melhora do processo de fabricação de produtos, muitas vezes lidamos com produtos altamente confiáveis e com uma substancial vida útil. Nessas situações, os testes padrão de confiabilidade podem demorar muito tempo e, possivelmente, requererem altos custos de investimento para que sejam realizados, até se obtenha os dados de tempo de falha desejados para realizar a inferência. Afim de assegurar a ocorrência de falhas rápidas e confiáveis nestes testes, é possível encurtar o período de suas ocorrências em todas ou em pelo menos algumas unidades do teste, submetendo-as a condições de estresse mais graves que as normais. Esses tipos de teste são denominados *Testes Acelerados* (quando todas as unidades são testadas em condições de aceleramento) e *Testes Parcialmente Acelerados* (quando existem unidades testadas em condições normais e outras testadas em condições de aceleramento).

Testes acelerados ocorrem na prática elevando-se o nível das tensões, como por exemplo temperatura, pressão, carga elétrica, umidade, etc, ou uma combinação desses fatores. A tensão sobre as unidades em teste de vida podem ser aplicadas de diferentes maneiras, sendo que os métodos mais comumente utilizados são os chamados *stress por passo* e os de *stress constante.*

Em um *teste parcialmente acelerado com stress constante*, cada item é colocado em condições de nível de tensão constante, isto é, apenas sob condições normais ou apenas sob condições de aceleramento, até que o teste termine. Já no *teste parcialmente acelerado com stress por passo*, cada item é executado em condições normais até determinado tempo e, se ele não falhar durante esse tempo, passa a ser executado sob condições de aceleramento até que ocorra falha ou que seja censurado. O objetivo de um teste parcialmente acelerado é coletar mais dados de falha em um tempo menor, sem necessariamente usar altas tensões para todas as unidades.

Nosso objetivo, a partir daqui, é combinar os estudos de censura progressiva com testes parcialmente acelerados com stress por passo, construindo um modelo de tempo de vida para uma amostra com distribuição exponencial generalizada sob essas condições.

Suponhamos que n unidades são colocadas em teste de vida sob censura do tipo II progressiva com m falhas observadas. Considere que os tempos de falha $X_1, \ldots, X_n$ das unidades dessa amostra possuem distribuição Exponencial Generalizada com função de densidade dada por (2.5) e função de distribuição dada por (2.6).

Diferentemente do exemplo apresentado na Seção 2.2, não iremos agora pressupor que os números de censura $R_1, \ldots, R_m$ sejam previamente definidos. Nesse caso, suponhamos que cada unidade é censurada do teste independentemente das outras, mas todas com a mesma probabilidade p de serem removidas. Então o número de unidades censuradas após cada falha seguirá uma distribuição binomial, isto é,

$$R_1 \sim \text{binomial}(n-m, p),$$

e

$$R_i \sim \text{binomial}\left(n-m-\sum_{j=1}^{i-1} R_j, p\right), \quad \text{para } i = 2, 3, \ldots, m-1,$$

com $R_m = n - m - R_1 - R_2 - \cdots - R_{m-1}$.

Cada uma das n unidades será colocada em teste inicialmente sob condições normais de tensão. Se a unidade não falhar ou não for censurada até um tempo pré-determinado τ, será colocada sob condições de aceleração (stress). A partir disso, podemos definir a amostra $Y_1, \ldots, Y_n$ sob teste parcialmente acelerado com stress por passo, tal que

$$Y_i = \begin{cases} X_i, & \text{se } X_i \leq \tau \\ \tau + (X_i - \tau)/\beta, & \text{se } X_i > \tau, \end{cases} \tag{2.7}$$

onde $X_1, \ldots, X_n$ são os tempos de falha da amostra sob condições normais e $\beta > 1$ é o fator de aceleração do teste.

Dessa forma, a função de densidade de Y_i pode ser definida por

$$f(y) = \begin{cases} 0, & \text{se } y \leq 0 \\ f_1(y) \equiv f(y; \alpha, \theta), & \text{se } 0 < y \leq \tau \\ f_2(y), & \text{se } y > \tau, \end{cases}$$

onde

$$f_2(y) \equiv f_2(y; \alpha, \theta, \beta) = \beta\alpha\theta\{1 - e^{-\alpha[\tau+\beta(y-\tau)]}\}^{\theta-1} e^{(\alpha[\tau+\beta(y-\tau)])},$$

obtida pela mudança de variável definida em (2.7).

Sejam $y_1, \ldots, y_m$ os tempos de falha observados da amostra $Y_1, \ldots, Y_n$. Definimos as indicadoras do processo de aceleração na amostra

$$\delta_{1i} = \begin{cases} 1, & \text{se } X_i \leq \tau \\ 0, & \text{se } X_1 > \tau \end{cases}$$

e

$$\delta_{2i} = \begin{cases} 0, & \text{se } X_i \leq \tau \\ 1, & \text{se } X_1 > \tau \end{cases}$$

e, então, obtemos a função de verossimilhança

$$L_1(y; \theta, \alpha, \beta) = \prod_{i=1}^{m} \Big\{ \big\{ f_1(y)[1 - F_1(y)]^{R_i} \big\}^{\delta_{1i}} \big\{ f_2(y)[1 - F_2(y)]^{R_i} \big\}^{\delta_{2i}} \Big\}, \tag{2.8}$$

onde $F_1(y) = F(y; \alpha, \theta)$ dada por (2.6) e $F_2(y) = \big(1 - e^{-\alpha[\tau+\beta(y-\tau)]}\big)^{\theta}$, obtida pela mudança de variável definida em (2.7).

Como o número de unidades censuradas r_i após a i-ésima falha segue distribuição binomial para todo $i = 1, \ldots, m$, e definimos que m falhas serão observadas, temos que

$$P(R_1 = r_1) = \binom{n-m}{r_1} p^{r_1} (1-p)^{n-m-r_1},$$

e

$$P(R_i = r_i | R_{i-1} = r_{i-1}; \dots, R_1 = r_1) = \binom{n - m - \sum_{j=1}^{i-i} r_j}{r_i} p^{r_i} (1-p)^{n-m-\sum_{j=1}^{i} r_j},$$

para $i = 2, 3, \dots, m-1$, e $0 \leq r_i \leq n - m - (r_1 + r_2 + \cdots + r_{i-1})$.

Definindo $\mathbf{R} = (R_1, \dots, R_m)$ o vetor do esquema de censura do modelo e $\mathbf{r} = (r_1, \dots, r_m)$ o vetor das quantidades de unidades retiradas do teste após cada falha, temos que

$$\begin{aligned} P(\mathbf{R} = \mathbf{r}) &= P(R_{m-1} = r_{m-1}, R_{m-2} = r_{m-2}, \dots, R_1 = r_1) \\ &= P(R_{m-1} = r_{m-1} | R_{m-2} = r_{m-2}, \dots, R_1 = r_1) \cdots P(R_2 = r_2 | R_1 = r_1) P(R_1 = r_1), \end{aligned}$$

isto é

$$P(\mathbf{R} = \mathbf{r}) = \frac{(n-m)!}{\left(n - m - \sum_{i=1}^{m-1} r_i\right)! \prod_{i=1}^{m-1} r_i!} p^{\sum_{i=1}^{m-1} r_i} (1-p)^{(m-1)(n-m) - \sum_{i=1}^{m-1} (m-i) r_i}.$$

Suponhamos que o número de unidades retiradas R_i sejam independentes das variáveis Y_i para todo i. Então a função de verossimilhança completa pode ser definida como

$$L(y; \alpha, \theta, \beta, p) = L_1(y; \alpha, \theta, \beta) P(\mathbf{R} = \mathbf{r}).$$

Discutiremos primeiramente o processo de obtenção dos EMV's para os parâmetros θ, α e β.

Como $P(\mathbf{R} = \mathbf{r})$ não depende dos parâmetros θ, α e β, então os EMV's para esses parâmetros podem ser determinados diretamente pela maximização da função $L_1(y; \theta, \alpha, \beta)$ definida em (2.8). Para isso utilizaremos a função de log-verossimilhança, aplicando o logaritmo natural na função $L_1(y; \theta, \alpha, \beta)$, obtendo

$$\begin{aligned} l(\theta, \alpha, \beta) &= m \log \alpha + m \log \theta - \alpha \sum_{i=1}^{m_u} y_i + (\theta - 1) \sum_{i=1}^{m_u} \log\left(1 - e^{-\alpha y_i}\right) + \\ &+ \sum_{i=1}^{m_u} R_i \log\left[1 - (1 - e^{-\alpha y_i})^{\theta}\right] + m_a \log \beta - \alpha \sum_{i=1}^{m_a} [\tau + \beta(y_i - \tau)] + \\ &+ (\theta - 1) \sum_{i=1}^{m_a} \log\left(1 - e^{-\alpha[\tau + \beta(y_i - \tau)]}\right) + \sum_{i=1}^{m_a} R_i \log\left[1 - (1 - e^{-\alpha[\tau + \beta(y_i - \tau)]})^{\theta}\right]; \end{aligned}$$

onde

$$\begin{cases} m_u = \sum_{i=1}^{m} \delta_{1i} \\ m_a = \sum_{i=1}^{m} \delta_{21}, \end{cases}$$

tal que $m = m_u + m_a$.

Derivando essa função em relação aos parâmetros desejados, segue que

$$\begin{aligned} \frac{\partial l}{\partial \theta} &= \frac{m}{\theta} + \sum_{i=1}^{m_u} \log\left(1 - e^{-\alpha y_i}\right) - \sum_{i=1}^{m_u} R_i \frac{(1 - e^{-\alpha y_i})^{\theta} \log\left(1 - e^{-\alpha y_i}\right)}{1 - (1 - e^{-\alpha y_i})^{\theta}} + \\ &+ \sum_{i=1}^{m_a} \log\left(1 - e^{-\alpha[\tau + \beta(y_i - \tau)]}\right) - \sum_{i=1}^{m_a} R_i \frac{(1 - e^{-\alpha[\tau+\beta(y_i-\tau)]})^{\theta} \log\left(1 - e^{-\alpha[\tau+\beta(y_i-\tau)]}\right)}{1 - (1 - e^{-\alpha[\tau+\beta(y_i-\tau)]})^{\theta}}, \end{aligned}$$

$$\begin{aligned} \frac{\partial l}{\partial \alpha} &= \frac{m}{\alpha} + \sum_{i=1}^{m_u} y_i + (\theta - 1) \sum_{i=1}^{m_u} \frac{y_i e^{-\alpha y_i}}{1 - e^{-\alpha y_i}} - \theta \sum_{i=1}^{m_u} R_i \frac{(1 - e^{-\alpha y_i})^{\theta - 1} y_i e^{-\alpha y_i}}{1 - (1 - e^{-\alpha y_i})^{\theta}} + \\ &+ \sum_{i=1}^{m_a} [\tau + \beta(y_i - \tau)] + (\theta - 1) \sum_{i=1}^{m_a} \frac{[\tau + \beta(y_i - \tau)] e^{-\alpha}[\tau + \beta(y_i - \tau)]}{1 - e^{-\alpha[\tau+\beta(y_i-\tau)]}} - \\ &- \theta \sum_{i=1}^{m_a} R_i \frac{(1 - e^{-\alpha[\tau+\beta(y_i-\tau)]})^{\theta-1} [\tau + \beta(y_i - \tau)] e^{-\alpha[\tau+\beta(y_i-\tau)]}}{1 - (1 - e^{-\alpha[\tau+\beta(y_i-\tau)]})^{\theta}} \end{aligned}$$

e

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \frac{m_a}{\beta} - \alpha \sum_{i=1}^{m_a} (y_i - \tau) - (\theta - 1)\alpha \sum_{i=1}^{m_a} \frac{(y_i - \tau) e^{-\alpha[\tau+\beta(y_i-\tau)]}}{1 - e^{-\alpha[\tau+\beta(y_i-\tau)]}} - \\ &- \theta\alpha \sum_{i=1}^{m_a} R_i \frac{(1 - e^{-\alpha[\tau+\beta(y_i-\tau)]})^{\theta-1} (y_i - \tau) e^{-\alpha[\tau+\beta(y_i-\tau)]}}{1 - (1 - e^{-\alpha[\tau+\beta(y_i-\tau)]})^{\theta}}. \end{aligned}$$

Daí obtemos as equações de verossimilhança $\frac{\partial l}{\partial \theta} = 0$, $\frac{\partial l}{\partial \alpha} = 0$ e $\frac{\partial l}{\partial \beta} = 0$, cujas soluções são os EMV's desejados, desde que sejam um máximo local de $l(\theta, \alpha, \beta)$. Observe que pela complexidade das expressões obtidas se torna difícil obter uma forma fechada para os EMV's, $\hat{\theta}$, $\hat{\alpha}$ e $\hat{\beta}$, entretanto, podemos utilizar métodos numéricos iterativos para resolver esse sistema de equações e obtê-los.

Independentemente, podemos encontrar agora o EMV para o parâmetro binomial p do número de unidades censuradas. Para isso, derivamos a função de log-verossimilhança de

$L(y;\alpha,\theta,\beta,p)$ em relação ao parâmetro p e obtemos a equação de verossimilhança

$$\frac{\partial \log L}{\partial p} = \frac{\sum_{i=1}^{m-1} r_i}{p} - \frac{(m-1)(n-m) - \sum_{i=1}^{m-1}(m-i)r_i}{1-p} = 0,$$

obtendo assim o EMV para p:

$$\hat{p} = \frac{\sum_{i=1}^{m-1} r_i}{(m-1)(n-m) - \sum_{i=1}^{m-1}(m-i-1)r_i}.$$

Agora, encontramos a Matriz de Informação de Fisher

$$I(\theta,\alpha,\beta) = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{12} & I_{22} & I_{23} \\ I_{13} & I_{23} & I_{33} \end{pmatrix},$$

onde

$$\begin{aligned} I_{11} = E\left(-\frac{\partial^2 l}{\partial \theta^2}\right) &= \frac{m}{\theta^2} + E\left[\sum_{i=1}^{m_u} R_i \frac{\psi_{2i}^{\theta}(\log \psi_{2i})^2(1-\psi_{2i}^{\theta}) + (\psi_{2i}\log\psi_{2i})^2}{(1-\psi_{2i}^{\theta})^2}\right] + \\ &+E\left[\sum_{i=1}^{m_a} R_i \frac{\psi_{4i}^{\theta}(\log \psi_{4i})^2(1-\psi_{4i}^{\theta}) + (\psi_{4i}^{\theta}\log\psi_{4i})^2}{(1-\psi_{4i}^{\theta})^2}\right]; \end{aligned}$$

$$\begin{aligned} I_{22} = E\left(-\frac{\partial^2 l}{\partial \alpha^2}\right) &= -\frac{m}{\alpha^2} + (\theta-1)E\left[\sum_{i=1}^{m_u} \frac{y_i^2\psi_{3i}(\psi_{2i}+\psi_{3i})}{\psi_{2i}^2} + \sum_{i=1}^{m_a}\frac{\psi_{1i}^2\psi_{5i}(\psi_{4i}+\psi_{5i})}{\psi_{4i}^2}\right] + \\ &+\theta E\left[\sum_{i=1}^{m_u} R_i y_i^2 \frac{[(\theta-1)\psi_{2i}^{\theta-2}\psi_{3i}^2 - \psi_{2i}^{\theta-1}\psi_{3i}](1-\psi_{2i}^{\theta}) + \theta\psi_{2i}^{2(\theta-1)}\psi_{3i}^2}{(1-\psi_{2i}^{\theta})^2}\right]; \end{aligned}$$

$$\begin{aligned} I_{33} = E\left(-\frac{\partial^2 l}{\partial \beta^2}\right) &= \frac{m_a}{\beta^2} + \alpha(\theta-1)E\left[\sum_{i=1}^{m_a}(y_i-\tau)^2\frac{\psi_{4i}\psi_{5i}+\psi_{5i}^2}{\psi_{4i}^2}\right] + \\ &+\alpha^2\theta E\left[\sum_{i=1}^{m_a} R_i (y_i-\tau)^2 \frac{[(\theta-1)\psi_{4i}^{\theta-2}\psi_{5i}^2 - \psi_{4i}^{\theta-1}\psi_{5i}](1-\psi_{4i}^{\theta}) + \theta\psi_{4i}^{2(\theta-1)}\psi_{5i}^2}{(1-\psi_{4i}^{\theta})^2}\right]; \end{aligned}$$

$$
\begin{aligned}
I_{12} &= E\left(-\frac{\partial^2 l}{\partial\theta\partial\alpha}\right) = -E\left[\sum_{i=1}^{m_u} y_i\frac{\psi_{3i}}{\psi_{2i}}\right] - E\left[\sum_{i=1}^{m_a}\psi_{1i}\frac{\psi_{5i}}{\psi_{4i}}\right] + \\
&+E\left[\sum_{i=1}^{m_u} R_i y_i\psi_{3i}\frac{[\psi_{2i}^{\theta-1}+\theta\psi_{2i}^{\theta-1}\log\psi_{2i}](1-\psi_{2i}^{\theta})+\theta\psi_{2i}^{2\theta-1}\log\psi_{2i}}{(1-\psi_{2i}^{\theta})^2}\right] + \\
&+E\left[\sum_{i=1}^{m_a} R_i\psi_{1i}\psi_{5i}\frac{[\psi_{4i}^{\theta-1}+\theta\psi_{4i}^{\theta-1}\log\psi_{4i}](1-\psi_{4i}^{\theta})+\theta\psi_{4i}^{2\theta-1}\log\psi_{4i}}{(1-\psi_{4i}^{\theta})^2}\right];
\end{aligned}
$$

$$
\begin{aligned}
I_{13} &= E\left(-\frac{\partial^2 l}{\partial\theta\partial\beta}\right) = -\alpha E\left[\sum_{i=1}^{m_a}(y_i-\tau)\frac{\psi_{5i}}{\psi 4i}\right] + \\
&+\alpha E\left[\sum_{i=1}^{m_a} R_i(y_i-\tau)\frac{[(\theta\log\psi_{4i}+1)\psi_{4i}^{\theta-1}\psi_{5i}](1-\psi_{4i}^{\theta})+\theta\psi_{4i}^{2\theta-1}\psi_{5_i}\log\psi_{4i}}{(1-\psi_{4i}^{\theta})^2}\right];
\end{aligned}
$$

$$
\begin{aligned}
I_{23} &= -E\left(\frac{\partial^2 l}{\partial\alpha\partial\beta}\right) = E\left[\sum_{i=1}^{m_a}(y_i-\tau)\right] + \\
&+(\theta-1)E\left[\sum_{i=1}^{m_a}(y_i-\tau)\frac{[\psi_{4i}\psi_{5i}-\alpha\psi_{1i}\psi_{4i}\psi_{5i}-\alpha\psi_{1i}\psi_{5i}^2]}{\psi_{4i}^2}\right] + \\
&+\alpha\theta E\left[\sum_{i=1}^{m_a} R_i(y_i-\tau)\frac{[(\theta-1)\psi_{4i}^{\theta-2}\psi_{1i}\psi_{5i}^2+\psi_{4i}^{\theta-1}\psi_{5i}(1-\psi_{1i})](1-\psi_{4i}^{\theta})+\theta\psi_{1i}\psi_{4i}^{2(\theta-1)}\psi_{51}^2}{(1-\psi_{4i}^{\theta})^2}\right];
\end{aligned}
$$

onde, $\psi_{1i}=\tau+\beta(Y_i-\tau)$, $\psi_{2i}=1-e^{-\alpha Y_i}$, $\psi_{3i}=e^{-\alpha Y_i}$, $\psi_{4i}=1-e^{-\alpha\psi_{1i}}$ e $\psi_{5i}=e^{-\alpha\psi_{1i}}$.

A partir da matriz de informação de Fisher, obtemos a matriz de variância-covariância assintótica de $(\hat{\theta},\hat{\alpha},\hat{\beta})$ dada por

$$G(\hat{\theta},\hat{\alpha},\hat{\beta}) = I^{-1}(\theta,\alpha,\beta),$$

utilizada na obtenção dos intervalos de confiança dos parâmetros.

2.4 - Inferência para a Distribuição Lognormal com 3 Parâmetros

Nesta seção abordamos a inferência estatística, considerando uma amostra com distribuição lognormal com três parâmetros sob censura do tipo II progressiva. Apresentamos os estudos do trabalho de Basak et al. em [6].

A distribuição lognormal é bastante usada para modelagem de tempos de falha em estudos de confiabilidade e é particularmente útil para modelagem de dados de cauda longa. Muitos trabalhos acerca desse tipo de distribuição podem ser encontrados na literatura, devido à fácil aplicabilidade em problemas práticos.

Existe uma clara relação entre a distribuição normal e a lognormal. Se $X = \log(Y - \gamma)$ é normalmente distribuída com média μ e desvio padrão σ, então a distribuição de Y será lognormal com três parâmetros, sendo estes $\theta = (\gamma, \mu, \sigma)$, onde σ é o parâmetro de forma, μ o de escala e γ o de posição. Sendo assim, a função de densidade de probabilidade de tal variável com distribuição lognormal com três parâmetros é dada por

$$f(y;\gamma,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}(y-\gamma)} \exp\left\{-\frac{[\log(y-\gamma)-\mu]^2}{2\sigma^2}\right\}, \tag{2.9}$$

com $\gamma < y < \infty$, $\sigma > 0$ e $-\infty < \mu < \infty$, onde σ^2 e μ são a variância e a média da variável normal X.

Quando o parâmetro γ é conhecido, a estimação dos outros dois parâmetros pode ser feita usando-se os resultados conhecidos para distribuição normal, simplesmente fazendo a mudança de variável de Y para X. Entretanto, quando γ não é conhecido os métodos de estimação se tornam mais complexos. Temos por objetivo, assim como nos modelos das seções 2.2 e 2.3, obter os EMV's para os parâmetros do modelo.

Suponha que n unidades são colocadas em um experimento de teste de vida sob censura do tipo II progressiva com m falhas observadas e esquema de censura pré-determinado $R_1, \ldots, R_m$. Suponha ainda que os tempos de vida $X_1, \ldots, X_n$ são variáveis aleatórias com distribuição lognormal com três parâmetros. Sejam $Y_1, \ldots, Y_m$ os m tempos de falha observados da amostra aleatória e $y_1, \ldots, y_m$ suas respectivas observações ordenadas. Então, por (1.13) obtemos a função de verossimilhança

$$L(\gamma,\mu,\sigma) = \prod_{i=1}^{m} \frac{1}{\sigma\sqrt{2\pi}(y_i-\gamma)} e^{\left\{-\frac{[\log(y_i-\gamma)-\mu]^2}{2\sigma^2}\right\}} \left[1 - \Phi\left(\frac{\log(y_i-\gamma)-\mu}{\sigma}\right)\right]^{R_i} \tag{2.10}$$

onde

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du.$$

Dessa forma, aplicamos o logaritmo em (2.10) para encontrar a função log-verossimilhança

$$l(\gamma,\mu,\sigma) = -m\log\sigma\sqrt{2\pi} - \sum_{i=1}^{m}\log(y_i-\gamma) - \frac{1}{2}\sum_{i=1}^{m}\psi_i^2 + \sum_{i=1}^{m} R_i \log[1-\Phi(\psi_i)] \tag{2.11}$$

onde $\psi_i = \frac{\log(y_i-\gamma)-\mu}{\sigma}$.

Derivando a função (2.11) em relação aos parâmetros desejados obtemos

$$
\begin{aligned}
\frac{\partial l(\gamma,\mu,\sigma)}{\partial \gamma} &= \sum_{i=1}^{m} \frac{1}{(y_i-\gamma)} + \frac{1}{\sigma}\sum_{i=1}^{m}\frac{1}{(y_i-\gamma)}\left[\psi_i + R_i\frac{\phi(\psi_i)}{1-\Phi(\psi_i)}\right];\\
\frac{\partial l(\gamma,\mu,\sigma)}{\partial \mu} &= \frac{1}{\sigma}\left[\sum_{i=1}^{m}\psi_i + \sum_{i=1}^{m} R_i\frac{\phi(\psi_i)}{1-\Phi(\psi_i)}\right];\\
\frac{\partial l(\gamma,\mu,\sigma)}{\partial \sigma} &= \frac{1}{\sigma}\left[-m\sum_{i=1}^{m}(\psi_i)^2 + \sum_{i=1}^{m} R_i\psi_i\frac{\phi(\psi_i)}{1-\Phi(\psi_i)}\right];
\end{aligned}
$$

onde $\phi(x)$ é a função de densidade normal padrão, isto é, $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

Desse modo, uma forma de se obter os EMV's dos parâmetros γ, μ e σ é resolver o sistema de equações de verossimilhança $\frac{\partial l}{\partial \gamma} = 0$, $\frac{\partial l}{\partial \mu} = 0$ e $\frac{\partial l}{\partial \sigma} = 0$, e verificar que são um máximo local da função $l(\gamma,\mu,\sigma)$, o que pode ser feito utilizando métodos numéricos clássicos, visto a complexidade de se encontrar uma forma fechada para as soluções desse sistema.

Além disso, encontramos a Matriz de Informação de Fisher

$$
I(\gamma,\mu,\sigma) = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{12} & I_{22} & I_{23} \\ I_{13} & I_{23} & I_{33} \end{pmatrix},
$$

com

$$
\begin{aligned}
I_{11} = E\left(-\frac{\partial^2 l}{\partial \gamma^2}\right) &= -E\left[\sum_{i=1}^{m}\frac{1}{(Y_i-\gamma)^2}\right] - \frac{1}{\sigma}E\left[\sum_{i=1}^{m}(\psi_i + R_iL_i)\frac{1}{(Y_i-\gamma)^2}\right] + \\
&\quad + \frac{1}{\sigma^2}E\left[\sum_{i=1}^{m}(1 - R_i\psi_iL_i + R_iL_i^2)\frac{1}{(Y_i-\gamma)^2}\right];\\
I_{22} = E\left(-\frac{\partial^2 l}{\partial \mu^2}\right) &= \frac{m}{\sigma^2} + \frac{1}{\sigma^2}E\left[\sum_{i=1}^{m} R_iL_i(L_i-\psi_i)\right];\\
I_{33} = E\left(-\frac{\partial^2 l}{\partial \sigma^2}\right) &= -\frac{m}{\sigma^2} + \frac{1}{\sigma^2}E\left[\sum_{i=1}^{m}[3\psi_i^2 + R_i(2\psi_iL_i - \psi_i^3L_i + \psi_i^2L_i^2)]\right];\\
I_{12} = E\left(-\frac{\partial^2 l}{\partial \gamma\partial \mu}\right) &= \frac{m}{\sigma^2} - \frac{1}{\sigma^2}E\left[\sum_{i=1}^{m_u}[R_i\psi_iL_i - R_iL_i^2]\frac{1}{(Y_i-\gamma)}\right];\\
I_{13} = E\left(-\frac{\partial^2 l}{\partial \gamma\partial \sigma}\right) &= \frac{1}{\sigma^2}E\left[\sum_{i=1}^{m}[2\psi_i + R_i(L_i - \psi_i^2L_i + \psi_iL_i^2)]\frac{1}{(Y_i-\gamma)}\right];
\end{aligned}
$$

$$I_{23} = E\left(-\frac{\partial^2 l}{\partial\mu\partial\sigma}\right) = \frac{1}{\sigma^2}E\left[\sum_{i=1}^{m}[2\psi_i + R_i(L_i - \psi_i^2 L_i + \psi L_i^2)]\right];$$

onde $L_i = \frac{\phi(\psi_i)}{1-\Phi(\psi_i)}$.

A partir da matriz de informação de Fisher, obtemos a matriz de variância-covariância assintótica de $(\hat{\gamma}, \hat{\mu}, \hat{\sigma})$ dada por

$$G(\hat{\gamma}, \hat{\mu}, \hat{\sigma}) = I^{-1}(\gamma, \mu, \sigma).$$

Uma alternativa para obtenção numérica dos EMV's apresentada por Basak e Balakrishnan em [6], é o uso do algoritmo EM, discutido na Seção 1.4. Como nosso modelo está sob censura, os dados obtidos são incompletos. Usaremos a ideia do algorimo EM para obter os EMV's dos parâmetros, isto é, tomamos os valores da variável observada e introduzimos uma variável latente, maximizando o valor esperado do logaritmo da função de densidade conjunta dessas variáveis, conforme apresentado na Seção 1.4.

Denotemos os dados censurados como um vetor $\mathbf{Z} = (Z_1, \ldots, Z_m)$, onde os dados da i-ésima censura, isto é, os dados não observados, podem ser representados por um vetor $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, \ldots, Z_{iR_i})$, para $i = 1, \ldots, m$, e como $\mathbf{Y} = (Y_1, \ldots, Y_m)$ os dados observados. Os dados completos são obtidos combinando os dados observados $\mathbf{Y}$ e os dados não observados $\mathbf{Z}$. Dessa forma, encontramos a função log-verossimilhança baseada na amostra "pseudo-completa"

$$\begin{aligned}
\log L(\mathbf{Y}, \mathbf{Z}; \theta) &= -\sum_{i=1}^{m}\left[\log\sigma\sqrt{2\pi} + \log(y_i - \gamma) + \frac{1}{2}\left(\frac{\log(y_i-\gamma)-\mu}{\sigma}\right)^2\right] - \\
&\quad -\sum_{i=1}^{m}\sum_{j=i}^{R_i}\left[\log\sigma\sqrt{2\pi} + \log(z_{ij} - \gamma) + \frac{1}{2}\left(\frac{\log(z_{ij}-\gamma)-\mu}{\sigma}\right)^2\right] \\
&= -n\log\sigma\sqrt{2\pi} - \sum_{i=1}^{m}\log(y_i-\gamma) - \frac{1}{2}\sum_{i=1}^{m}\left(\frac{\log(y_i-\gamma)-\mu}{\sigma}\right)^2 - \\
&\quad -\sum_{i=1}^{m}\sum_{j=1}^{R_i}\log(Z_{ij}-\gamma) - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{R_i}\left(\frac{\log(Z_{ij}-\gamma)-\mu}{\sigma}\right)^2.
\end{aligned}$$

Executando o passo E do algoritmo, denotando $\theta(k) = \theta_{(k)}$ como o valor do parâmetro θ na k-ésima iteração, encontramos a esperança condicional de $\log L(\mathbf{Y}, \mathbf{Z}; \theta)$ com respeito a densidade condicional das variáveis do vetor $\mathbf{Z}$ dado $\mathbf{Y} = \mathbf{y}$, onde $\theta = (\gamma, \mu, \sigma)$ e $\mathbf{y} = (y_1, \ldots, y_m)$ são os dados observados da $\mathbf{Y}$

$$\begin{aligned}
Q(\theta,\theta_{(k)}) &= E[\log L(\mathbf{Y},\mathbf{Z},\theta)|\theta_{(k)},\mathbf{y}] = \int_E [\log L(\mathbf{Y},\mathbf{Z},\theta)] f_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y};\theta_{(k)})\mu_{\mathbf{y}} d\mathbf{z} = \\
&= -n\log\sqrt{2\pi} - n\log\sigma - \sum_{i=1}^{m}\log(y_i-\gamma) - \frac{1}{2}\sum_{i=1}^{m}\left(\frac{\log(y_i-\gamma)-\mu}{\sigma}\right)^2 - \\
&\quad - \int_E \sum_{i=1}^{m}\sum_{j=1}^{R_i}\left[\log(z_{ij}-\gamma) + \left(\frac{\log(z_{ij}-\gamma)-\mu}{2\sigma}\right)^2\right] f_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y};\theta_{(k)})\mu_{\mathbf{y}} d\mathbf{z},
\end{aligned}$$

e então segue que

$$\begin{aligned}
Q(\theta,\theta_{(k)}) &= -n\log\sqrt{2\pi} - n\log\sigma - \sum_{i=1}^{m}\log(y_i-\gamma) - \frac{1}{2}\sum_{i=1}^{m}\left(\frac{\log(y_i-\gamma)-\mu}{\sigma}\right)^2 - \\
&\quad - \sum_{i=1}^{m}\sum_{j=1}^{R_i} E\left[\log(Z_{ij}-\gamma) + \left(\frac{\log(Z_{ij}-\gamma)-\mu}{2\sigma}\right)^2 |Z_{ij} > y_i\right].
\end{aligned} \tag{2.12}$$

Essas esperanças condicionais são obtidas usando-se o fato de que dado $Y_i = y_i$, as variáveis $\mathbf{Z}_i$'s têm distribuição truncada em y_i, isto é, a função densidade de probabilidade dessas $\mathbf{Z}_i$'s dado $Y_i = y_i$ são dadas por

$$f_{Z_{ij}|Y_i}(z_{ij}|y_i;\theta) = \frac{f(z_{ij};\theta)}{1-F(y_i,\theta)},$$

onde $f(z_{ij};\theta)$ é densidade dada por (2.9), e $F(y_i,\theta)$ sua respectiva função de distribuição (ver [3] e [28]).

A partir daí, obtemos a função densidade de probabilidade condicional de $\mathbf{Z}$ dado $\mathbf{Y}$:

$$f_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y};\theta) = \prod_{i=1}^{m}\prod_{j=1}^{R_i} f_{Z_{ij}|Y_i}(z_{ij}|y_i;\theta).$$

Agora, executemos o passo M, que consiste em maximizar $Q(\theta,\theta_{(k)})$ obtida em (2.12), isto é, encontrar $\theta_{(k+1)} = argmax_\theta Q(\theta,\theta_{(k)})$. Para isso, iniciemos o processo de iteração com um valor $\theta_{(0)}$, em que usamos $0 < \gamma_{(0)} < y_1$, e $\mu_{(0)}$ e $\sigma_{(0)}$ obtidos com base nos dados denominados "pseudo-completos", isto é, dados que envolvem os valores observados de $\mathbf{Y}$ e assumindo que todas as observações censuradas no i-ésimo passo $\mathbf{Z}_i$ tomem o valor y_i.

Então $\mu_{(0)} = \mu(\gamma_{(0)})$ e $\sigma_{(0)} = \sigma(\gamma_{(0)})$ são dados por

$$\begin{aligned}
\mu_{(0)} &= \frac{1}{n}\sum_{i=1}^{m}(R_i+1)\log(y_i-\gamma_{(0)}), \\
\sigma_{(0)} &= \left[\frac{1}{n}\sum_{i=1}^{m}(R_i+1)\log^2(y_i-\gamma_{(0)}) - \mu_{(0)}^2\right]^{1/2}.
\end{aligned}$$

Iniciando o processo de iteração com $\mu_{(0)}$ e $\sigma_{(0)}$, a $(k+1)$-ésima iteração é obtida dado o valor da k-ésima iteração $\theta_{(k)}$, dada por

$$\begin{aligned}
\mu_{(k+1)} &= \frac{1}{n}\left\{\sum_{i=1}^{m}\log(y_i-\gamma_{(k)})+\sum_{i=1}^{m}\sum_{j=1}^{R_i}E[\log(Z_{ij}-\gamma_{(k)})|Z_{ij}>y_i;\theta_{(k)}]\right\}\\
\sigma_{(k+1)} &= \left[\frac{1}{n}\left\{\sum_{i=1}^{m}\log^2(y_i-\gamma_{(k)})+\right.\right.\\
&\left.\left.+\sum_{i=1}^{m}\sum_{j=1}^{R_i}E[\log^2(Z_{ij}-\gamma_{(k)})|Z_{ij}>y_i;\mu_{(k+1)},\sigma_{(k)},\gamma_{(k)}]\right\}-\mu^2_{(k+1)}\right]^{1/2},
\end{aligned}$$

em que as esperanças condicionais são obtidas por

$$E[\log(Z_{ij}-\gamma_{(k)})|Z_{ij}>y_i;\theta_{(k)}]=\sigma_{(k)}L_{i(k)}+\mu_{(k)},$$
$$E[\log^2(Z_{ij}-\gamma_{(k)})|Z_{ij}>y_i;\mu_{(k+1)},\sigma_{(k)},\gamma_{(k)}]=\sigma^2_{(k)}[1+\psi_i^*L_i^*]+2\sigma_{(k)}\mu_{(k+1)}L_i^*+\mu^2_{(k+1)},$$

onde

$$\begin{aligned}
\psi_{i(k)} &= \frac{\log(y_i-\gamma_{(k)})-\mu_{(k)}}{\sigma_{(k)}},\\
L_{i(k)} &= \frac{\phi(\psi_{i(k)})}{1-\Phi(\psi_{i(k)})},\\
\psi_i^* &= \frac{\log(y_i-\gamma_{(k)})-\mu_{(k+1)}}{\sigma_{(k)}},\\
L_i^* &= \frac{\phi(\psi_i^*)}{1-\Phi(\psi_i^*)},
\end{aligned}$$

obtendo assim o valor da $(k+1)$-ésima iteração para os parâmetros μ e σ.

Resta agora obter $\gamma_{(k+1)}$. Para isso, derivamos (2.12) em relação a γ e igualamos a 0, obtendo a seguinte equação:

$$\begin{aligned}
&[\mu_{(k+1)}+\sigma^2_{(k+1)}]\sum_{i=1}^{m}\frac{1}{y_i-\gamma}-\sum_{i=1}^{m}\frac{\log(y_i-\gamma)}{y_i-\gamma}+\\
&+[\mu_{(k+1)}+\sigma^2_{(k+1)}]\sum_{i=1}^{m}\sum_{j=1}^{R_i}E\left[\frac{1}{Z_{ij}-\gamma}\middle|Z_{ij}>y_i;\gamma,\theta^*_{(k+1)}\right]-\\
&-\sum_{i=1}^{m}\sum_{j=1}^{R_i}E\left[\frac{\log(Z_{ij}-\gamma)}{Z_{ij}-\gamma}\middle|Z_{ij}>y_i;\gamma,\theta^*_{(k+1)}\right]=0,
\end{aligned}$$

onde $\theta^*_{(k+1)}=(\mu_{(k+1)},\sigma_{(k+1)})$.

As esperanças condicionais dessa expressão são obtidas como segue:

$$E\left[\frac{1}{Z_{ij}-\gamma}\middle|Z_{ij} > y_i; \gamma, \theta^*_{(k+1)}\right] = e^{\frac{\sigma^2_{(k+1)}}{2}-\mu_{(k+1)}} P_{i(k+1)}(\gamma)$$
$$E\left[\frac{\log(Z_{ij}-\gamma)}{Z_{ij}-\gamma}\middle|Z_{ij} > y_i; \gamma, \theta^*_{(k+1)}\right] = e^{\frac{\sigma^2_{(k+1)}}{2}-\mu_{(k+1)}}[\sigma_{(k+1)} P_{i(k+1)}(\gamma) + (\mu_{(k+1)} - \sigma^2_{(k+1)}) P_{i(k+1)}(\gamma)],$$

onde

$$\begin{aligned} \psi_{i(k+1)}(\gamma) &= \frac{\log(y_i - \gamma) - \mu_{(k+1)}}{\sigma_{(k+1)}}, \\ P_{i(k+1)}(\gamma) &= \frac{1 - \Phi(\psi_{i(k+1)}(\gamma) + \hat{\sigma}_{(k+1)})}{1 - \Phi(\psi_{i(k+1)}(\gamma))}. \end{aligned}$$

Pelo Teorema 1.21, a convergência para os EMV's dos parâmetros do modelo é garantida. Dessa forma, repetindo as iterações até que sequência $\theta_{(k)}$ convirja, obtemos o EMV $\hat{\theta}$ de θ desejado.

Capítulo 3

Propriedades Assintóticas dos Estimadores de Máxima Verossimilhança baseados na Censura do Tipo II Progressiva

3.1 - Introdução

No Capítulo 1 vimos que sob determinadas condições de regularidade sobre a função de densidade $f(x;\theta)$ as propriedades assintóticas de consistência e normalidade são garantidas ao EMV do parâmetro unidimensional θ de uma amostra aleatória da distribuição $f(x;\theta)$. Vimos também que nos casos de censura tipo II progressiva quando o tamanho n da amostra é suficientemente grande, o esquema de censura deve ser proporcionalmente grande, isto é, se $n \to \infty$ então $\frac{m}{n} \to \tau$. Isso garante que o esquema de censura tipo II progressiva não se aproxime assintoticamente de um esquema de censura do tipo II.

Ao estimar parâmetros para amostras censuradas tipo II progressivamente, naturalmente surge o questionamento se as propriedades assintóticas obtidas para amostras sem censura podem ser asseguradas também para esse caso. Neste capítulo veremos que é possível, mas que é necessário acrescentar condições de regularidade sobre a densidade $f(x;\theta)$ da amostra sob censura em relação às condições impostas no caso sem censura.

Lin e Balakrishnan exibem em [4] a garntia das propriedades de consistência e normalidade assintótica para amostras sob censura tipo II progressiva. Os autores utilizam o Princípio da Informação Perdida apresentado por Louis em [27] e Tanner em [32], para caracterizar a amostra observável como uma combinação de uma amostra completa e uma amostra censurada. Dessa forma, as propriedades desejadas são garantidas para a amostra considerada completa, como apresentamos na Seção 1.2.

A novidade do trabalho de Lin e Balakrishnan [4], está na garantia dessas propriedades para os dados censurados. Para demonstrá-las nesse caso, os autores acrescentam as condições existência da 1ª, 2ª e 3ª derivadas de $\log f(x;\theta)$ em relação a θ para quase todo x (condição A1), além de sua limitação por funções integráveis (condição A3), assim como a limitação da 1ª, 2ª e 3ª derivadas de $f(x;\theta)$ e da função $\frac{1}{1-F(x;\theta)}$ por funções também integráveis(condições A2 e A4).

A partir daí, é usada uma versão da Lei Fraca dos Grandes Números para variáveis aleatórias independentes mas não identicamente distribuídas, apresentada por Hoadley em [21], e então consegue-se garantir as duas propriedades desejadas para a amostra censurada.

Apresentamos detalhadamente todos esses passos e obtemos a consistência e a normalidade assintótica para amostras censuradas tipo II progressivamente nas Seções 3.2 e 3.3, respectivamente.

No Apêndice, demonstramos os dois teoremas apresentados por Hoadley [21] necessários na demonstração das propriedades estudadas neste capítulo.

3.2 - Consistência

A obtenção da consistência para o EMV para amostras sob esquemas de censura do tipo II progressiva segue os mesmos passos e ideias feitos na Seção 1.2 quando verificamos essa propriedade para amostras completas. Como era de se esperar, para dados censurados serão necessárias algumas hipóteses adicionais em relação à amostras completas. Assumimos novamente que θ é um parâmetro unidimensional com valores possíveis em $\Theta \subset \mathbb{R}$. Seja $\mathscr{F} = \{f(x;\theta); \theta \in \Theta\}$ uma família de densidades relativas a medida de Lebesgue μ. Considere um conjunto de dados $x_1, \ldots, x_n$ como realizações i.i.d. de uma variável aleatória X com densidade $f(x;\theta_0) \in \mathscr{F}$, onde $\theta_0 \in \Theta \subset \mathbb{R}$ é o parâmetro verdadeiro a ser estimado.

Seguem abaixo as condições de regularidade necessárias para os resultados que vêm a seguir.

(A1) Para quase todo x, as derivadas $\frac{\partial}{\partial\theta}\log f(x;\theta)$, $\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)$ e $\frac{\partial^3}{\partial\theta^3}\log f(x;\theta)$ existem para todo θ pertencendo a um intervalo não degenerado I.

(A2) Para todo $\theta \in I$, temos

$$\left|\frac{\partial}{\partial\theta}f(x;\theta)\right| \leq G_1, \quad \left|\frac{\partial^2}{\partial\theta^2}f(x;\theta)\right| \leq G_2, \quad \left|\frac{\partial^3}{\partial\theta^3}f(x;\theta)\right| \leq G_3,$$

onde

$$\int G_i(x)d\mu(x) < \infty, \quad i = 1, 2, 3$$

e μ é a medida de Lebesgue.

(A3) Para todo $\theta \in I$ e constantes positivas δ e K, temos

$$\left|\frac{\partial}{\partial\theta}\log f(x;\theta)\right| \leq G_1^*(x), \quad \left|\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right| \leq G_2^*(x), \quad \left|\frac{\partial^3}{\partial\theta^3}\log f(x;\theta)\right| \leq G_3^*(x),$$

e

$$\int |G_i^*(x)|^{1+\delta} f(x;\theta)d\mu(x) \leq K, \quad i = 1, 2, 3.$$

(A4) Para todo $\theta \in I$ e M constante positiva, $\frac{1}{1-F(x;\theta)}$ é limitada por $\eta(x)$, onde

$$\int \eta(x)f(x;\theta)d\mu(x) \leq M.$$

(A5) Para todo $\theta \in I$, a integral

$$\gamma^2 = \int \left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right]^2 f(x;\theta)d\mu(x)$$

é finita e positiva.

Primeiramente, demonstraremos o seguinte lema:

Lema 3.1. *Assumindo as condições (A2) − (A4) e que existe uma função mensurável e integrável T com $\int T(x_1, \ldots, x_m)d\mu(x) < \infty$, e ainda Q uma constante positiva independente de θ. Então temos que*

$$\frac{1}{n}\left|\frac{\partial^3}{\partial\theta^3}\log L(\theta; x_1, \ldots, x_m)\right| \leq T(x_1, \ldots, x_m)$$

e $E[T(x_1, \ldots, x_m)] \leq Q$.

Demonstração. Por (1.13) temos que a função de verossimilhança é dada por

$$L(\theta) = L(\theta; x_1, \ldots, x_m) = C.\prod_{i=1}^{m} f(x_i;\theta).[1 - F(x_i;\theta)]^{R_i},$$

onde $x_1 < \ldots < x_m$ e C é a constante normalizadora. Aplicando o logaritmo aí, temos

$$l(\theta) = \log L(\theta; x_1, \ldots, x_m) = \log C + \sum_{i=1}^{m}\log f(x_i;\theta) + \sum_{i=1}^{m} R_i \log[1 - F(x_i;\theta)].$$

Então

$$\frac{1}{n}\left|\frac{\partial^3}{\partial\theta^3}\log L(\theta)\right| \leq \frac{1}{n}\sum_{i=1}^{m}\left|\frac{\partial^3}{\partial\theta^3}\log f(x_i;\theta)\right| + \frac{1}{n}\sum_{i=1}^{m}R_i\left|\frac{\partial^3}{\partial\theta^3}\log\left[1-F(x_i;\theta)\right]\right| \tag{3.1}$$

Pela condição $(A3)$ temos que $\left|\frac{\partial^3}{\partial\theta^3}\log f(x;\theta)\right| \leq G_3^*(x_i)$; então basta provarmos que a segunda expressão do lado direito de (3.1) também é limitada por alguma função. Podemos escrever

$$\left|\frac{\partial^3}{\partial\theta^3}\log\left[1-F(x_i;\theta)\right]\right| =$$
$$\left|\frac{\frac{\partial^3}{\partial\theta^3}[1-F(x_i;\theta)]}{1-F(x_i;\theta)} - 3\frac{\frac{\partial}{\partial\theta}[1-F(x_i;\theta)]\frac{\partial^2}{\partial\theta^2}[1-F(x_i;\theta)]}{[1-F(x_i;\theta)]^2} + 2\frac{\left(\frac{\partial}{\partial\theta}[1-F(x_i;\theta)]\right)^3}{[1-F(x_i;\theta)]^3}\right|.$$

Pela condição $(A4)$, $\frac{1}{1-F(x_i;\theta)} \leq \eta(x_i)$ e então

$$\left|\frac{\partial^3}{\partial\theta^3}\log\left[1-F(x_i;\theta)\right]\right| \leq$$
$$\left|\frac{\partial^3}{\partial\theta^3}F(x_i;\theta)\right|\eta(x_i) + 3\left|\frac{\partial}{\partial\theta}F(x_i;\theta)\right|\left|\frac{\partial^2}{\partial\theta^2}F(x_i;\theta)\right|\eta^2(x_i) + 2\left|\frac{\partial}{\partial\theta}F(x_i;\theta)\right|^3\eta^3(x_i) =$$
$$\left|\frac{\partial^3}{\partial\theta^3}\int_{A_i}f(x;\theta)d\mu(x)\right|\eta(x_i) + 3\left|\frac{\partial}{\partial\theta}\int_{A_i}f(x;\theta)d\mu(x)\right|\left|\frac{\partial^2}{\partial\theta^2}\int_{A_i}f(x;\theta)d\mu(x)\right|\eta^2(x_i) +$$
$$+2\left|\frac{\partial}{\partial\theta}\int_{A_i}f(x;\theta)d\mu(x)\right|^3\eta^3(x_i);$$

onde $A_i = \{x : -\infty < x \leq x_i\}$ para $i = 1, \ldots, m$.

Note que pela condição $(A2)$ as funções $\frac{\partial^j}{\partial\theta^j}f(x;\theta)$ são limitadas por funções G_j integráveis, $j = 1, 2, 3$. Pelo Teorema da Convergência Dominada de Lebesgue segue que

$$\int\frac{\partial^j}{\partial\theta^j}f(x;\theta)d\mu(x) = \frac{\partial^j}{\partial\theta^j}\int f(x;\theta)d\mu(x), \quad j = 1,2,3.$$

Então

$$\left|\frac{\partial^3}{\partial\theta^3}\log\left[1-F(x_i;\theta)\right]\right| \leq$$
$$\left|\int_{A_i}\frac{\partial^3}{\partial\theta^3}f(x;\theta)d\mu(x)\right|\eta(x_i) + 3\left|\int_{A_i}\frac{\partial}{\partial\theta}f(x;\theta)d\mu(x)\right|\left|\int_{A_i}\frac{\partial^2}{\partial\theta^2}f(x;\theta)d\mu(x)\right|\eta^2(x_i) +$$
$$+2\left|\int_{A_i}\frac{\partial}{\partial\theta}f(x;\theta)d\mu(x)\right|^3\eta^3(x_i) \ \leq \ \left|\int_{A_i}G_3(x)d\mu(x)\right|\eta(x_i) +$$
$$+3\left|\int_{A_i}G_1(x)d\mu(x)\right|\left|\int_{A_i}G_2(x)d\mu(x)\right|\eta^2(x_i) + 2\left|\int_{A_i}G_1(x)d\mu(x)\right|^3\eta^3(x_i) \equiv \upsilon(x_i)$$

Pelas condições $(A2)$ e $(A4)$ a função obtida $\upsilon(x_i)$ é limitada, para todo $i = 1, \ldots, m$ e

$$\int \upsilon(x) f(x;\theta) d\mu(x) \leq M^*, \tag{3.2}$$

onde M^* é uma constante positiva independente de θ. Sendo assim, podemos definir

$$T(x_1, \ldots, x_m) = \frac{1}{n} \sum_{i=1}^{m} [G_3^*(x_i) + R_i \upsilon(x_i)],$$

de tal forma que, por (3.1) temos

$$\frac{1}{n} \left| \frac{\partial^3}{\partial \theta^3} \log L(\theta; x_1, \ldots, x_m) \right| \leq T(x_1, \ldots, x_m).$$

Observe que, pela condição $(A3)$ e por (3.2)

$$\begin{aligned} E[T(X_1, \ldots, X_m)] &= E\left[\frac{1}{n} \sum_{i=1}^{m} [G_3^*(X_i) + R_i \upsilon(X_i)]\right] \\ &= \frac{1}{n} \sum_{i=1}^{m} E[G_3^*(X_i)] + \frac{1}{n} \sum_{i=1}^{m} R_i E[\upsilon(X_i)] < Q, \end{aligned}$$

onde Q é uma constante positiva.

Logo, o lema está demonstrado. ■

Para mostrar a consistência do EMV em casos de censura do tipo II progressiva, usamos um resultado, estabelecido por Hoadley [21], que trata da Lei Fraca dos Grandes Números para variáveis aleatórias independentes mas não identicamente distribuídas. Este resultado é descrito abaixo e é demonstrado no Apêndice.

Teorema 3.2. *Sejam $\{Y_k : k = 1, 2, \ldots\}$ variáveis aleatórias independentes definidas em um espaço de probabilidade $(\Omega, \mathscr{F}, P_\theta)$ e tomando valores em um espaço de medida $(\Upsilon, \mathscr{A}, \mu)$. Seja $H_k : \Upsilon \times S \to \mathscr{R}^1$, onde $S \subset \mathscr{R}^p$ é compacto e seja $h_k(s) = E[H_k(Y_k, s)]$ Assumimos:*

(a) *Para cada $s \in S$, $H_k(\cdot, s)$ é $\mathscr{A}$-mensurável;*

(b) *$H_k(Y_k, \cdot)$ é contínua em S, uniformemente em k q.c.[P];*

(c) *Existe uma função mensurável $B_k : \Upsilon \to \mathscr{R}^1$ tal que $|H_k(\cdot, s)| < B_k(\cdot)$ para todo $s \in S$ e $E|B_k(Y_k)|^{1+\delta} \leq K$, onde K e δ são constantes positivas.*

Então:

(i) $h_k(\cdot)$ *é contínua em S, uniformemente em k;*

(ii) $\sup\left\{\left|\frac{1}{n}\sum_{k=1}^{n} H_k(Y_k, s) - \frac{1}{n}\sum_{k=1}^{n} h_k(s)\right| : s \in S\right\} \xrightarrow{P} 0.$

Para mostrar a propriedade de consistência do EMV, consideramos $1 \leq m < n$, já que quando $m = n$ temos uma amostra completa e esse caso já foi demonstrado no Teorema 1.10. Segue então o teorema:

Teorema 3.3. *Se as condições $(A1) - (A5)$ são satisfeitas, então a equação de verossimilhança*

$$\frac{\partial}{\partial\theta}\log L(\theta; x_1, \ldots, x_m) = \sum_{i=1}^{m}\left\{\frac{\partial}{\partial\theta}\log f(x_i;\theta) + R_i\frac{\partial}{\partial\theta}\log\left[1 - F(x_i;\theta)\right]\right\} = 0 \tag{3.3}$$

tem uma sequência de soluções $\hat{\theta}_n$, que converge em probabilidade para o verdadeiro valor de θ, digamos θ_0; ou seja, $\hat{\theta}_n$ é uma sequência de estimadores de máxima verossimilhança consistente de θ_0.

Demonstração. Da expansão de Taylor com resto de Lagrange da função $\frac{\partial}{\partial\theta}\log L(\theta; x_1, \ldots, x_m)$ em torno do verdadeiro valor θ_0 do parâmetro, temos que

$$\begin{aligned}\frac{1}{n}\frac{\partial}{\partial\theta}\log L(\theta) &= \frac{1}{n}\frac{\partial}{\partial\theta}\log L(\theta; x_1, \ldots, x_m)\Big|_{\theta=\theta_0} + \\ &+ \frac{(\theta-\theta_0)}{n}\frac{\partial^2}{\partial\theta^2}\log L(\theta; x_1, \ldots, x_m)\Big|_{\theta=\theta_0} + \frac{(\theta-\theta_0)^2}{2n}\frac{\partial^3}{\partial\theta^3}\log L(\theta; x_1, \ldots, x_m)\Big|_{\theta=\bar{\theta}},\end{aligned}$$

onde $\bar{\theta}$ é um valor entre θ e θ_0.

Observe que, pelo Lema 3.1, temos que $\frac{1}{n}\left|\frac{\partial^3}{\partial\theta^3}\log L(\theta; x_1, \ldots, x_m)\right| \leq T(x_1, \ldots, x_m)$ e então podemos definir

$$\Delta(x_1, \ldots, x_m) = \begin{cases} 0, & \text{se } T(x_1, \ldots, x_m) = 0 \\ \frac{1}{n}\frac{\partial^3}{\partial\theta^3}\log L(\theta; x_1, \ldots, x_m)\Big|_{\theta=\bar{\theta}}\frac{1}{T(x_1, \ldots, x_m)}, & \text{se } T(x_1, \ldots, x_m) > 0, \end{cases}$$

o que nos dá $0 \leq |\Delta(x_1, \ldots, x_m)| \leq 1$ e também

$$\begin{aligned}\frac{1}{n}\frac{\partial}{\partial\theta}\log L(\theta; x_1, \ldots, x_m) &= \frac{1}{n}\frac{\partial}{\partial\theta}\log L(\theta; x_1, \ldots, x_m)\Big|_{\theta=\theta_0} \\ &\quad + \frac{(\theta-\theta_0)}{n}\frac{\partial^2}{\partial\theta^2}\log L(\theta; x_1, \ldots, x_m)\Big|_{\theta=\theta_0} + \\ &\quad + \frac{1}{2}(\theta-\theta_0)^2\Delta(x_1, \ldots, x_m)T(x_1, \ldots, x_m) = \\ &= B_0 + (\theta-\theta_0)B_1 + \frac{1}{2}(\theta-\theta_0)^2\bar{B}_2,\end{aligned} \tag{3.4}$$

onde

$$
\begin{aligned}
B_0 &= \frac{1}{n}\frac{\partial}{\partial\theta}\log L(\theta; x_1,\ldots,x_m)\Big|_{\theta=\theta_0}\\
&= \frac{1}{n}\sum_{i=1}^{m}\left\{\frac{\partial}{\partial\theta}\log f(x_i;\theta) + R_i\frac{\partial}{\partial\theta}\log\left[1-F(x_i;\theta)\right]\right\}\Big|_{\theta=\theta_0};\\
B_1 &= \frac{1}{n}\frac{\partial^2}{\partial\theta^2}\log L(\theta; x_1,\ldots,x_m)\Big|_{\theta=\theta_0}\\
&= \frac{1}{n}\sum_{i=1}^{m}\left\{\frac{\partial^2}{\partial\theta^2}\log f(x_i;\theta) + R_i\frac{\partial^2}{\partial\theta^2}\log\left[1-F(x_i;\theta)\right]\right\}\Big|_{\theta=\theta_0};\\
\bar{B}_2 &= \Delta(x_1,\ldots,x_m)T(x_1,\ldots,x_m).
\end{aligned}
$$

Como $|\Delta(x_1,\ldots,x_m)| \leq 1$, temos que

$$
\begin{aligned}
|\bar{B}_2| &= |\Delta(x_1,\ldots,x_m)T(x_1,\ldots,x_m)|\\
&< |\Delta(x_1,\ldots,x_m)|\, T(x_1,\ldots,x_m)\\
&\leq T(x_1,\ldots,x_m) := B_2.
\end{aligned}
$$

Assim, $\bar{B}_2 = \Delta B_2$, onde $|\Delta| < 1$ e então podemos reescrever (3.4) da forma

$$
\frac{1}{n}\frac{\partial}{\partial\theta}\log L(\theta; x_1,\ldots,x_m) = B_0 + (\theta-\theta_0)B_1 + \frac{1}{2}(\theta-\theta_0)^2\Delta B_2. \tag{3.5}
$$

Como $x_1,\ldots,x_m$ são observações incompletas da amostra $X_1,\ldots,X_n$ censurada tipo II progressivamente, podemos reescrever suas informações utilizando o Princípio da Informação Perdida, apresentado na Observação 1.18.

Consideremos $W_i = X_i$, $i = 1,\ldots,n$, a amostra aleatória completa de tamanho n com função densidade de probabilidade $f(\cdot;\theta)$ e função de distribuição $F(\cdot;\theta)$ onde $w_1,\ldots,w_n$ são suas respectivas observações, e $Y_{ij} = X_{ij}$, $j = 1,\ldots,R_i$ para cada $i = 1,\ldots,m$, as variáveis censuradas com função densidade de probabilidade $\psi_i(y;\theta) = \frac{f(y;\theta)}{1-F(x_i;\theta)}$, com $y > x_i$, onde $y_{i1},\ldots,y_{iR_i}$ são suas respectivas observações, para $i = 1,\ldots,m$. Observe que as variáveis W_i e Y_{ij} são independentes.

Então podemos escrever

$$
\sum_{i=1}^{m}\log f(x_i;\theta) = \sum_{i=1}^{n}\log f(w_i;\theta) - \sum_{i=1}^{m}\sum_{j=1}^{R_i}\log f(y_{ij};\theta|X_i = x_i).
$$

No que segue, mostramos que a equação de verossimilhança (3.3) tem solução $\hat{\theta}_n$ que converge em probabilidade para θ_0, e fazemos isso seguindo os mesmos passos da demonstração do Lema 1.10 e Teorema 1.11. Verificamos que: *(i)* $B_0 \xrightarrow{P} 0$, *(ii)* $B_1 \xrightarrow{P} -\zeta_1^2$ e *(iii)* $B_2 \xrightarrow{P} \zeta_2$, onde ζ_1 e ζ_2 são constantes que serão definidas. Para simplificar a notação, a partir daqui denotaremos θ_0 simplesmente por θ. Vejamos então.

(i) Pelo Princípio da Informação Perdida, dado pela observação 1.18, podemos reescrever B_0 como

$$\begin{aligned}
B_0 &= \frac{1}{n}\left\{\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(w_i;\theta) - \sum_{i=1}^{m}\sum_{j=1}^{R_i}\frac{\partial}{\partial\theta}\log f(y_{ij};\theta|X_i=x_i)+\right.\\
&\left.+\sum_{i=1}^{m}R_i\frac{\partial}{\partial\theta}\log\left[1-F(x_i;\theta)\right]\right\} \equiv \frac{1}{n}(B_{01}-B_{02}),
\end{aligned}$$

onde

$$\begin{aligned}
B_{01} &= \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(w_i;\theta)\\
&\text{e}\\
B_{02} &= \sum_{i=1}^{m}\sum_{j=1}^{R_i}\frac{\partial}{\partial\theta}\log f(y_{ij};\theta|X_i=x_i) - \sum_{i=1}^{m}R_i\frac{\partial}{\partial\theta}\log\left[1-F(x_i;\theta)\right].
\end{aligned}$$

Pelo Lema 1.10, temos que

$$\frac{1}{n}B_{01} \xrightarrow{P} 0,$$

já que se trata da amostra completa, pelo Princípio da Informação Perdida. Basta, então, provarmos que $\frac{1}{n}B_{02} \xrightarrow{P} 0$.

Podemos reescrever B_{02} da seguinte forma

$$\begin{aligned}
B_{02} &= \sum_{i=1}^{m}\sum_{j=1}^{R_i}\frac{\partial}{\partial\theta}\log f(y_{ij};\theta|x_i) - \sum_{i=1}^{m}R_iE\left[\frac{\partial}{\partial\theta}\log f(Y_{i1};\theta|x_i)\right]+\\
&+\sum_{i=1}^{m}R_iE\left[\frac{\partial}{\partial\theta}\log f(Y_{i1};\theta|x_i)\right] - \sum_{i=1}^{m}R_i\frac{\partial}{\partial\theta}\log\left[1-F(x_i;\theta)\right].
\end{aligned}$$

Observe que pela densidade truncada das variáveis Y_{ij}, podemos escrever

$$\begin{aligned}
E\left[\frac{\partial}{\partial\theta}\log f(Y_{i1};\theta|x_i)\right] &= \int_{B_i}\frac{\partial}{\partial\theta}\log f(y;\theta).\psi_i(y;\theta)d\mu(y)\\
&= \int_{B_i}\frac{\partial}{\partial\theta}\log f(y;\theta).\frac{f(y;\theta)}{1-F(x_i;\theta)}d\mu(y),
\end{aligned}$$

e pelo Teorema da Convergência Dominada de Lebesgue

$$
\begin{aligned}
E\left[\frac{\partial}{\partial\theta}\log f(Y_{i1};\theta|x_i)\right] &= \frac{1}{1-F(x_i;\theta)}\int_{B_i}\frac{\partial}{\partial\theta}f(y;\theta)d\mu(y) \\
&= \frac{\frac{\partial}{\partial\theta}\int_{B_i}f(y;\theta)d\mu(y)}{1-F(x_i;\theta)} \\
&= \frac{\frac{\partial}{\partial\theta}[1-F(x_i;\theta)]}{1-F(x_i;\theta)} \\
&= \frac{\partial}{\partial\theta}\log\left[1-F(x_i;\theta)\right],
\end{aligned}
$$

onde $B_i = \{y : x_i < y < \infty\}$ para $i = 1, \ldots, m$.

Isso significa que

$$
\sum_{i=1}^{m} R_i E\left[\frac{\partial}{\partial\theta}\log f(Y_{i1};\theta|x_i)\right] = \sum_{i=1}^{m} R_i \frac{\partial}{\partial\theta}\log\left[1-F(x_i;\theta)\right]. \tag{3.6}
$$

Agora note que pelas condições $(A1)$ e $(A3)$ as variáveis $\sum_{j=1}^{R_i}\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta)$, $i = 1, \ldots, m$, satisfazem as condições do Teorema 3.2, pois $\sum_{j=1}^{R_i}\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta)$ é mensurável; para todo $\theta \in I$ é contínua em I, uniformemente em j q.c.[P]; e existe uma função G_1^* tal que $\left|\sum_{j=1}^{R_i}\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta)\right| \leq |R_i G_1^*|$ com $R_i^{1+\delta}E[G_1^*]^{1+\delta} \leq K$ para todo $i = 1, \ldots, m$, onde K e δ são constantes positivas.

Dessa forma, segue que

$$
\sup\left\{\left|\frac{B_{02}}{n-m}\right| : \theta \in I\right\} =
$$
$$
\sup\left\{\left|\frac{1}{n-m}\sum_{i=1}^{m}\sum_{j=1}^{R_i}\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i) - \frac{1}{n-m}\sum_{i=1}^{m}R_i E\left[\frac{\partial}{\partial\theta}\log f(Y_{i1};\theta|x_i)\right]\right| : \theta \in I\right\} \xrightarrow{P} 0.
$$

Daí, juntamente com (3.6), temos que $B_{02} \xrightarrow{P} 0$ e, como $B_{01} \xrightarrow{P} 0$, consequentemente temos que

$$
B_0 \xrightarrow{P} 0,
$$

como queríamos mostrar.

(ii) Usando novamente o Princípio da Informação Perdida, reescrevemos B_1:

$$\begin{aligned} B_1 &= \frac{1}{n}\left\{\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(w_i;\theta) - \sum_{i=1}^{m}\sum_{j=1}^{R_i}\frac{\partial^2}{\partial\theta^2}\log f(y_{ij};\theta|X_i=x_i) + \right. \\ &\left. + \sum_{i=1}^{m} R_i\frac{\partial^2}{\partial\theta^2}\log\left[1-F(x_i;\theta)\right]\right\} \equiv \frac{1}{n}(B_{11}-B_{12}), \end{aligned}$$

onde

$$\begin{aligned} B_{11} &= \sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(w_i;\theta) \\ \text{e} & \\ B_{12} &= \sum_{i=1}^{m}\sum_{j=1}^{R_i}\frac{\partial^2}{\partial\theta^2}\log f(y_{ij};\theta|X_i=x_i) - \sum_{i=1}^{m} R_i\frac{\partial^2}{\partial\theta^2}\log\left[1-F(x_i;\theta)\right]. \end{aligned}$$

Novamente, pelo Lema 1.10, temos que

$$\frac{1}{n}B_{11} \xrightarrow{P} E\left[\frac{\partial^2}{\partial\theta^2}\log f(W;\theta)\right] = -\gamma^2.$$

Agora, reescrevemos o termo $\frac{1}{n}B_{12}$ como

$$\begin{aligned} &\frac{n-m}{n}\left\{\frac{1}{n-m}\sum_{i=1}^{m}\sum_{j=1}^{R_i}\frac{\partial^2}{\partial\theta^2}\log f(y_{ij};\theta|X_i=x_i) - \frac{1}{n-m}\sum_{i=1}^{m} R_i E\left[\frac{\partial^2}{\partial\theta^2}\log f(Y_{i1};\theta|x_i)\right]\right\} \\ &-\frac{1}{n}\sum_{i=1}^{m} R_i\left\{\frac{\partial^2}{\partial\theta^2}\log\left[1-F(x_i;\theta)\right] - E\left[\frac{\partial^2}{\partial\theta^2}\log f(Y_{i1};\theta|x_i)\right]\right\}. \end{aligned}$$

Segue do Teorema 3.2 que o primeiro termo da expressão acima converge em probabilidade para zero quando $n\to\infty$. Resta então verificar o comportamento do segundo termo da expressão.

Note que

$$\begin{aligned} \frac{\partial^2}{\partial\theta^2}\log\left[1-F(x_i;\theta)\right] &= \frac{\frac{\partial^2}{\partial\theta^2}\left[1-F(x_i;\theta)\right]}{1-F(x_i;\theta)} - \left\{\frac{\partial}{\partial\theta}\log\left[1-F(x_i;\theta)\right]\right\}^2 \\ &= \frac{\frac{\partial^2}{\partial\theta^2}\left[1-F(x_i;\theta)\right]}{1-F(x_i;\theta)} - \left\{E\left[\frac{\partial}{\partial\theta}\log f(Y_{i1};\theta|x_i)\right]\right\}^2, \end{aligned}$$

e

$$E\left[\frac{\partial^2}{\partial\theta^2}\log f(Y_{i1};\theta|x_i)\right] = \int_{B_i}\left[\frac{\frac{\partial^2}{\partial\theta^2}f(y;\theta)}{f(y;\theta)} - \left(\frac{\partial}{\partial\theta}\log f(y;\theta)\right)^2\right]\frac{f(y;\theta)}{1-F(x_i;\theta)}d\mu(y) =$$
$$\int_{B_i}\frac{\frac{\partial^2}{\partial\theta^2}f(y;\theta)}{f(y;\theta)}\frac{f(y;\theta)}{1-F(x_i;\theta)}d\mu(y) - \int_{B_i}\left(\frac{\partial}{\partial\theta}\log f(y;\theta)\right)^2\frac{f(y;\theta)}{1-F(x_i;\theta)}d\mu(y) =$$
$$\frac{\frac{\partial^2}{\partial\theta^2}[1-F(x_i;\theta)]}{1-F(x_i;\theta)} - E\left[\frac{\partial}{\partial\theta}\log f(Y_{i1};\theta|x_i)\right]^2.$$

Então, temos que

$$\frac{\partial^2}{\partial\theta^2}\log\left[1-F(x_i;\theta)\right] - E\left[\frac{\partial^2}{\partial\theta^2}\log f(Y_{i1};\theta|x_i)\right] =$$
$$E\left[\frac{\partial}{\partial\theta}\log f(Y_{i1};\theta|x_i)\right]^2 - \left\{E\left[\frac{\partial}{\partial\theta}\log f(Y_{i1};\theta|x_i)\right]\right\}^2 = Var\left[\frac{\partial}{\partial\theta}\log f(Y_{i1};\theta|x_i)\right],$$

que é limitada e independente de θ, pelas condições $(A4)$ e $(A5)$.

Assim,

$$\frac{1}{n-m}\sum_{i=1}^{m}R_i\left\{\frac{\partial^2}{\partial\theta^2}\log\left[1-F(x_i;\theta)\right] - E\left[\frac{\partial^2}{\partial\theta^2}\log f(Y_{i1};\theta|x_i)\right]\right\} =$$
$$\frac{1}{n-m}\sum_{i=1}^{m}\sum_{j=1}^{R_i}Var\left[\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i)\right] \tag{3.7}$$

converge para um valor finito, digamos Λ.

Como, pela observação 1.17, $\lim_{n\to\infty}\frac{m}{n} = \tau$, segue que $\lim_{n\to\infty}\frac{n-m}{n} = 1-\tau$, e daí

$$\frac{1}{n}\sum_{i=1}^{m}R_i\left\{\frac{\partial^2}{\partial\theta^2}\log\left[1-F(x_i;\theta)\right] - E\left[\frac{\partial^2}{\partial\theta^2}\log f(Y_{i1};\theta|x_i)\right]\right\} \longrightarrow (1-\tau)\Lambda.$$

Portanto, combinando os resultados obtidos e definindo

$$\zeta_1^2 = \gamma^2 + (1-\tau)\Lambda > 0, \tag{3.8}$$

obtemos

$$B_1 = \frac{1}{n}(B_{01}+B_{02}) \xrightarrow{P} -\zeta_1^2,$$

como queríamos verificar.

*(iii)*Finalmente, seguindo as mesmas ideias da convergência de B_1, temos que B_2 converge em probabilidade para um valor finito, digamos ζ_2.

Seguindo os mesmos argumentos do Lema 1.10 e do Teorema 1.11, verificamos que a equação (3.3) possui uma sequência de soluções $\hat{\theta}_n$ que converge em probabilidade para o verdadeiro valor θ_0 do parâmetro, ou seja, $\hat{\theta}_n$ é uma sequência consistente de estimadores de máxima verossimilhança de θ_0.

■

3.3 - Normalidade Assintótica

Para estabelecer a normalidade assintótica do EMV do parâmetro θ, Lin e Balakrishnan [4] utilizam o resultados de Hoadley [21] e o Teorema de Slutsky, veja Serfling [31], para o caso multivariado. Hoadley exibe uma forma de Liapunov para o Teorema do Limite Central para o caso multivariado, que é uma importante ferramenta para mostrar que estimadores de máxima verossimilhança de θ têm normalidade assintótica em amostras de variáveis independentes mas não identicamente distribuídas. Segue o enunciado de um Teorema apresentado por Hoadley, demonstrado no Apêndice deste trabalho.

Teorema 3.4. *Sejam $\boldsymbol{X}_k, k = 1, 2, \ldots$ vetores aleatórios p-dimensionais independentes com $E\boldsymbol{X}_k = \boldsymbol{0}$, $Cov(\boldsymbol{X}_k) = \Gamma_k$. Assumimos que:*

(a) $\quad \bar{\Gamma}_n = \frac{1}{n}\sum_{k=1}^{n} \Gamma_k \longrightarrow \bar{\Gamma}$, *onde $\bar{\Gamma}$ é definida positiva;*

(b) $\quad$ *Para algum $\delta > 0$, $\frac{1}{n^{(2+\delta)/2}} \sum_k E|\lambda' \boldsymbol{X}_k|^{2+\delta} \longrightarrow 0$, para todo $\lambda \in \mathbb{R}^p$.*

Então $\frac{1}{\sqrt{n}} \sum_k \boldsymbol{X}_k \xrightarrow{D} N(\boldsymbol{0}, \bar{\Gamma})$.

O seguinte resultado mostra que a convergência de funções com distribuição univariada pode ser estendida para a convergência de funções com distribuição multivariada.

Teorema 3.5. *(Teorema de Slutsky) Sejam (X_n, Y_n), $n = 1, 2, \ldots$, e (X, Y) dois vetores aleatórios definidos em um mesmo espaço de probabilidade. Suponha que $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{D} Y$. Se X_n e Y_n são independentes para cada n, então*

$$(X_n, Y_n) \xrightarrow{D} (X^*, Y^*),$$

onde X^* e Y^* têm a mesma distribuição que X e Y, respectivamente, e são independentes no mesmo espaço.

No que segue, mostramos a normalidade assintótica do EMV de θ.

Teorema 3.6. *Sob as condições de regularidade $(A1)-(A5)$, a equação de verossimilhança (3.3) tem uma solução $\hat{\theta}$ que possui distribuição assintoticamente normal.*

Demonstração. Seja $\hat{\theta}$ uma solução consistente da equação (3.3), isto é,

$$\frac{\partial}{\partial\theta}\log L(\hat{\theta}) = 0.$$

Então, por (3.5) temos que

$$B_0 + (\hat{\theta}-\theta_0)B_1 + \frac{(\hat{\theta}-\theta_0)^2}{2}\Delta B_2 = 0$$

donde segue que

$$(\hat{\theta}-\theta_0) = \frac{B_0}{-B_1 - \frac{(\hat{\theta}-\theta_0)}{2}\Delta B_2}.$$

Logo temos que

$$\sqrt{n}(\hat{\theta}-\theta_0) = \frac{\frac{\sqrt{n}}{\zeta_1^2}B_0}{-\frac{B_1}{\zeta_1^2} - \frac{(\hat{\theta}-\theta_0)}{2\zeta_1^2}\Delta B_2}$$

e portanto

$$\sqrt{n}(\hat{\theta}-\theta_0) = \frac{\frac{1}{\zeta_1^2\sqrt{n}}\sum_{i=1}^{m}\left\{\frac{\partial}{\partial\theta}\log f(x_i;\theta_0) + R_i\frac{\partial}{\partial\theta}\log\left[1-F(x_i;\theta_0)\right]\right\}}{-\frac{B_1}{\zeta_1^2} - \frac{(\hat{\theta}-\theta_0)}{2\zeta_1^2}\Delta B_2}, \tag{3.9}$$

onde $\zeta_1^2 = \gamma^2 + (1-\tau)\Lambda$, dado em (3.8).

Vimos em *(ii)* e *(iii)* da demonstração do Teorema 3.3 que $-\frac{B_1}{\zeta_1^2} \xrightarrow{P} 1$ e B_2 é limitada por Q, logo $\frac{(\hat{\theta}-\theta_0)}{2\zeta_1^2}\Delta B_2 \xrightarrow{P} 0$.

Assim o denominador de (3.9) converge em probabilidade para 1. Resta verificar então a convergência do numerador da fração.

Pelo Princípio da Informação Perdida e usando o fato de que $\frac{\partial}{\partial\theta}\log\left[1-F(x_i;\theta)\right] = E\left[\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i)\right]$, podemos reescrever

$$\sum_{i=1}^{m}\left\{\frac{\partial}{\partial\theta}\log f(x_i;\theta_0) + R_i\frac{\partial}{\partial\theta}\log\left[1-F(x_i;\theta_0)\right]\right\} =$$

$$\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(w_i;\theta) - \sum_{i=1}^{m}\sum_{j=1}^{R_i}\frac{\partial}{\partial\theta}\log f(y_{ij};\theta|x_i) + \sum_{i=1}^{m}\sum_{j=1}^{R_i}E\left[\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i)\right] =$$

$$\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(w_i;\theta) - \sum_{i=1}^{m}\sum_{j=1}^{R_i}\left\{\frac{\partial}{\partial\theta}\log f(y_{ij};\theta|x_i) - E\left[\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i)\right]\right\}.$$

Segue do Teorema 1.11 que $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(w_i;\theta)$ é assintoticamente normal com média 0 e variância γ^2, pois w_i são consideradas as observações de uma amostra completa.

Note que pelas condições $(A3)$ e $(A4)$, $E\left|\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i) - E\left[\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i)\right]\right|^3$ é limitada por um valor independente de θ, digamos K^*. Então, segue que

$$\frac{1}{n}.\frac{1}{\sqrt{n}}\sum_{i=1}^{m}\sum_{j=1}^{R_i}E\left|\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i) - E\left[\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i)\right]\right|^3 = \frac{n-m}{n^{3/2}}K^* \overset{n\to\infty}{\longrightarrow} 0. \quad (3.10)$$

Além disso, por (3.7) temos que

$$\frac{1}{n-m}\sum_{i=1}^{m}\sum_{j=1}^{R_i}Var\left[\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i)\right] \xrightarrow{P} \Lambda. \quad (3.11)$$

Verifica-se então que as condições do Teorema 3.4 são satisfeitas para a sequência de vetores $X_j := \frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i) - E\left[\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i)\right]$, $i = 1,\ldots,m$ e $j = 1,\ldots,R_i$, já que facilmente verifica-se que $E|X_j| = 0$ e por (3.10) a média das $(n-m)$ covariâncias $Cov(X_j)$ converge para um valor Λ definido positivo, satisfazendo a condição (a) do Teorema. Além disso, por (3.10) a condição (b) também é satisfeita. Sendo assim, do Teorema 3.4, fazendo $\delta = 1$, segue que

$$\frac{1}{\sqrt{n-m}}\sum_{i=1}^{m}\sum_{j=1}^{R_i}\left\{\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i) - E\left[\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i)\right]\right\} \xrightarrow{D} N(0,\Lambda).$$

Pela observação 1.17 temos que $\frac{m}{n}\to\tau$ quando $n\to\infty$, e então segue que

$$\frac{\sqrt{n-m}}{\sqrt{n}}.\frac{1}{\sqrt{n-m}}\sum_{i=1}^{m}\sum_{j=1}^{R_i}\left\{\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i) - E\left[\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i)\right]\right\} \xrightarrow{D} N(0,(1-\tau)\Lambda),$$

e, portanto

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{m}\sum_{j=1}^{R_i}\left\{\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i) - E\left[\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i)\right]\right\} \xrightarrow{D} N(0,(1-\tau)\Lambda).$$

Agora, definindo

$$\begin{aligned} W_i^* &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(W_i;\theta) \\ &\text{e} \\ Y_{ij}^* &= \frac{1}{\sqrt{n}}\sum_{i=1}^{m}\sum_{j=1}^{R_i}\left\{\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i) - E\left[\frac{\partial}{\partial\theta}\log f(Y_{ij};\theta|x_i)\right]\right\}, \end{aligned}$$

com $i = 1, \ldots, m$ e $j = 1, \ldots, R_i$, como W_i^* e Y_{ij}^* são independentes para todo i e j, podemos aplicar o Teorema 3.5 e obter

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(W_i; \theta), \frac{1}{\sqrt{n}} \sum_{i=1}^{m} \sum_{j=1}^{R_i} \left\{ \frac{\partial}{\partial \theta} \log f(Y_{ij}; \theta | x_i) - E\left[\frac{\partial}{\partial \theta} \log f(Y_{ij}; \theta | x_i) \right] \right\} \right) \xrightarrow{D} (W^*, Y^*),$$

onde W^* e Y^* têm distribuição $N(0, \gamma^2)$ e $N(0, (1-\tau)\Lambda)$ respectivamente, e são independentes.

Daí segue do Teorema de Slutsky (ver em [1]) que

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(W_i; \theta) - \frac{1}{\sqrt{n}} \sum_{i=1}^{m} \sum_{j=1}^{R_i} \left\{ \frac{\partial}{\partial \theta} \log f(Y_{ij}; \theta | x_i) - E\left[\frac{\partial}{\partial \theta} \log f(Y_{ij}; \theta | x_i) \right] \right\} \xrightarrow{D} X^* - Y^*$$

Como $\gamma^2 + (1-\tau)\Lambda = \zeta_1^2$, temos que

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{m} \left\{ \frac{\partial}{\partial \theta} \log f(X_i; \theta) + R_i \frac{\partial}{\partial \theta} \log \left[1 - F(X_i; \theta) \right] \right\} \xrightarrow{D} N(0, \gamma^2 + (1-\tau)\Lambda) = N(0, \zeta_1^2),$$

donde segue que

$$\frac{1}{\zeta_1^2 \sqrt{n}} \sum_{i=1}^{m} \left\{ \frac{\partial}{\partial \theta} \log f(X_i; \theta) + R_i \frac{\partial}{\partial \theta} \log \left[1 - F(X_i; \theta) \right] \right\} \xrightarrow{D} N\left(0, \frac{1}{\zeta_1^2}\right),$$

e então concluímos que $\sqrt{n}(\hat{\theta} - \theta_0)$ tem distribuição assintoticamente normal $(0, [\zeta_1^2]^{-1})$. ∎

Apêndice

Neste apêndice demonstraremos os teoremas utilizados na obtenção das propriedades assintóticas do Capítulo 3 e enunciaremos resultados necessários para essas demonstrações.

Aqui tomamos $X_i, i \in I$, onde $I = 1, 2, \ldots$, variáveis aleatórias independentes definidas no espaço de probabilidade $(\Omega, \mathscr{F}, P)$ e tomando valores num espaço de medida $(\Upsilon, \mathscr{A}, \mu)$.

Primeiramente, introduzimos o conceito de integrabilidade uniforme, fundamental para garantir os resultados de consistência e normalidade assintótica dos EMV dos parâmetros em amostras censuradas tipo II progressivamente.

Definição A. 1. *Uma família de variáveis aleatórias $\{X_i, i \in I\}$ é dita uniformemente integrável (u.i.) se*

$$\lim_{M\to\infty} \sup_{i\in I} \int_{|X_i|>M} |X_i| dP = 0.$$

Lema A. 2. *Um condição suficiente para que $\{X_i, i \in I\}$ seja u.i. é que*

$$E|X_i|^{1+\delta} \leq \infty,$$

para algum $\delta > 0$.

Lema A. 3. *São equivalentes*
(i) $\{X_n; n = 1, 2, \ldots\}$ é u.i. e $X_n \xrightarrow{P} X$.
(ii) X é integrável e $E|X_n - X| \overset{n\to\infty}{\longrightarrow} 0$.

Lema A. 4. *Seja $U \subset \mathbb{R}^p$. Se $\{X_k(u); k = 1, 2, \ldots; u \in U\}$ é u.i e $\lim_{u\to u_0} X_k(u) = X_k$, q.c. [P], então*
(i) $\{X_k; k = 1, 2, \ldots\}$ é u.i.
Se além disso, $\lim_{u\to u_0} X_k(u) = X_k$ uniformemente em k, q.c.[P], então

(ii) $\lim\limits_{u\to u_0} E|X_k(u)-X_k| = 0$, *uniformemente em* k, *isto é,* $\lim\limits_{u\to u_0} EX_k(u) = EX_k$ *uniformemente em* k.

Demonstração.

(i) Como $\{X_k(u), k = 1,2,\ldots; u \in U\}$ é u.i., tomemos M suficientemente grande e ϵ suficientemente pequeno de forma que

$$\int_{|x_k(u)|>M} |X_k(u)|dP < \epsilon.$$

Para cada k fixo, definimos

$$\begin{aligned} A &= \{\lim_{u\to u_0} X_k(u) = X_k\}, \\ B(u) &= \{|X_k(u)| > M\}, \\ B &= \{|X_k| > M\}. \end{aligned}$$

Para algum $F \in F$, seja $I(F)$ a variável aleatória indicadora associada a F. Então segue que

$$\lim_{u\to u_0} \inf I(A\cap B(u))|X_k(u)| \geq I(A\cap B)|X_k|,$$

e então, pelo Lema de Fatou, segue que

$$\begin{aligned} \int_{|X_k|>M} |X_k|dP &= E[I(A\cap B)|X_k|] \\ &\leq E[\lim_{u\to u_0} \inf I(A\cap B(u))|X_k(u)|] \\ &\leq \lim_{u\to u_0} \inf E[I(A\cap B(u))|X_k(u)|] \\ &= \lim_{u\to u_0} \inf \int_{|X_k|>M} |X_k|dP < \epsilon. \end{aligned}$$

Segue então que $\int_{|X_k|>M} |X_k|dP < \epsilon$, e então

$$\lim_{M\to\infty} \sup_k \int_{|X_k|>M} |X_k|dP = 0,$$

o que significa que $\{X_k, k = 1,2,\ldots\}$ é u.i..

(ii) Suponhamos por contradição que $\lim\limits_{u\to u_0} E|X_k(u)-X_k| \neq 0$. Então existe $\epsilon > 0$ e sequências $k_m \stackrel{n\to\infty}{\longrightarrow} \infty$ e $s_n \stackrel{n\to\infty}{\longrightarrow} u_0$ tais que $E[Z_n] > \epsilon$, onde

$$Z_n = |X_{k_n}(s_n) - X_k|.$$

Como $\{X_{k_n}; n = 1,2,\ldots\}$ é u.i. por hipótese e $\{X_k, k = 1,2,\ldots\}$ também é u.i. pelo item *(i)*, segue que $\{Z_n; n = 1,2,\ldots\}$ também o é.

Além disso, como $X_k(u) \overset{u\to u_0}{\longrightarrow} X_k$ unif. em K, q.c.[P], segue que

$$\lim_{n\to\infty} |X_{k_n}(s_n) - X_k| = \lim_{n\to\infty} Z_n = 0, \text{ q.c.[P]}.$$

Assim, Z_n satisfaz as hipóteses do Teorema 2 e então segue que

$$E|Z_n - 0| \overset{n\to\infty}{\longrightarrow} 0 \Rightarrow \lim_{n\to\infty} E[Z_n] = 0,$$

o que é uma contradição.

Logo, $E|X_k(u) - X_k| \overset{n\to\infty}{\longrightarrow} 0.$

■

Teorema A. 5. *Sejam $\{Y_k : k = 1, 2, \ldots\}$ variáveis aleatórias independentes definidas em um espaço de probabilidade $(\Omega, \mathscr{F}, \mathscr{P}_\theta)$ e tomando valores em um espaço de medida $(\Upsilon, \mathscr{A}, \mu)$. Seja $H_k : \Upsilon \times S \to \mathscr{R}^1$, onde $S \subset \mathscr{R}^p$ é compacto e seja $h_k(s) = E[H_k(Y_k, s)]$ Assumimos:*

(a) *Para cada $s \in S$, $H_k(\cdot, s)$ é $\mathscr{A}$-mensurável;*

(b) *$H_k(Y_k, \cdot)$ é contínua em S, uniformemente em k q.c.[P];*

(c) *Existe uma medida $B_k : \Upsilon \to \mathscr{R}^1$ tal que $|H_k(\cdot, s)| < B_k(\cdot)$ para todo $s \in S$ e $E|B_k(Y_k)|^{1+\delta} \leq K$, onde K e δ são constantes positivas.*

Então:

(i) *$h_k(\cdot)$ é contínua em S, uniformemente em k;*

(ii) $\sup \left\{ \left| \frac{1}{n} \sum_{k=1}^{n} H_k(Y_k, s) - \frac{1}{n} \sum_{k=1}^{n} h_k(s) \right| : s \in S \right\} \overset{P}{\longrightarrow} 0$

Demonstração.

(i) Pela hipótese b, para cada $s_0 \in S$, temos que $\lim_{s\to s_0} H_k(Y_k, s) = H_k(Y_k, s_0)$ unif. em k, q.c.[P]. Pela hipótese c temos que a família $\{H_k(Y_k, s); k = 1, 2, \ldots; s \in S\}$ é u.i., já que é $H_k(Y_k, s)$ é limitada por uma função mensurável com esperança finita (ver [7]).

Então, pelo item *(ii)* do Teorema A.4 segue que $\lim_{s\to s_0} h_k(s) = h_k(s_0)$ unif. em k, isto é, $h_k(s)$ é contínua em S, unif. em k.

(ii) Pelo item *(i)*, podemos assumir sem perda de generalidade que $h_k(s) = EH_k(Y_k, s) = 0$.

Sejam

$$H_k^*(y, s, \rho) = \sup\{H_k(y, t); |t - s| \leq \rho\}$$
$$H_{k*}(y, s, \rho) = \inf\{H_k(y, t); |t - s| \leq \rho\},$$

onde $|.|$ é a norma euclidiana usual.

As funções $H_k^*(y,s,\rho)$ e $H_{k*}(y,s,\rho)$ são $\mathscr{A}$-mensuráveis, já que $H_k(Y+k,s)$ é contínua em S, e S é separável, já $S \subset \mathbb{R}^p$. Daí, pela hipótese b segue que

$$\lim_{\rho\to 0} H_k^*(Y_k,s,\rho) = H_k(Y_k,s)$$
$$\lim_{\rho\to 0} H_{k*}(Y_k,s,\rho) = H_k(Y_k,s),$$

unif. em k, q.c.[P].

Então, pela parte *(ii)* do Teorema 0.4,

$$\lim_{\rho\to 0} EH_k^*(Y_k,s,\rho) = EH_k(Y_k,s) = 0$$
$$\lim_{\rho\to 0} EH_{k*}(Y_k,s,\rho) = EH_k(Y_k,s) = 0,$$

unif. em k, q.c.[P]. Daí, para cada $s \in S$, dado $\epsilon > 0$, existe $\rho(s)$ suficientemente pequeno, tal que

$$-\epsilon < EH_{k*}(Y_k,s,\rho(s)) \leq EH_k^*(Y_k,s,\rho(s)) < \epsilon. \tag{12}$$

Como S é um conjunto compacto, toda cobertura aberta de S admite uma subcobertura finita. Observe que a coleção $\{S(s,\rho(s))\}$ forma uma cobertura aberta de S, então existem $s_1,\dots,s_m \in S$ tal que $S \subset \bigcup_{i=1}^{m} S(s_i,\rho(s_i))$, isto é, $\bigcup_{i=1}^{m} S(s_i,\rho(s_i))$ é uma sobcobertura finita de S.

Daí segue que para todo $s \in S$

$$\min_{1\leq i\leq m}\left\{\frac{1}{n}\sum_{k=1}^{n} H_{k*}(Y_k,s_i,\rho(s_i))\right\} \leq \frac{1}{n}\sum_{k=1}^{n} H_k(Y_k,s) \leq \max_{1\leq i\leq m}\left\{\frac{1}{n}\sum_{k=1}^{m} H_k^*(Y_k,s_i,\rho(s_i))\right\}. \tag{13}$$

Pela condição (c) temos que

$$E|H_k^*(Y_k,s,\rho)|^{1+\delta} \leq K$$
$$E|H_{k*}(Y_k,s,\rho)|^{1+\delta} \leq K,$$

o que implica que

$$\tfrac{1}{n^{1+\delta}}\sum_{k=1}^{n} E|H_k^*(Y_k,s,\rho)|^{1+\delta} \overset{n\to\infty}{\longrightarrow} 0$$
$$\tfrac{1}{n^{1+\delta}}\sum_{k=1}^{n} E|H_{k*}(Y_k,s,\rho)|^{1+\delta} \overset{n\to\infty}{\longrightarrow} 0,$$

e então, pela Lei Fraca dos Grandes Números de Markov (ver [30]) temos que

$$\begin{array}{l} \frac{1}{n}\sum_{k=1}^{n} H_k^*(Y_k,s,\rho) - \frac{1}{n}\sum_{k=1}^{n} E[H_k^*(Y_k,s,\rho)] \xrightarrow{P} 0 \\ \frac{1}{n}\sum_{k=1}^{n} H_{k*}(Y_k,s,\rho) - \frac{1}{n}\sum_{k=1}^{n} E[H_{k*}(Y_k,s,\rho)] \xrightarrow{P} 0. \end{array} \tag{14}$$

Combinando os resultados (12) e (14) temos que com probabilidade tendendo a $1-\epsilon_n$, com $\epsilon_n \to 0$

$$-2\epsilon < \frac{1}{n}\sum_{k=1}^{n} H_{k*}(Y_k,s,\rho(s)) \leq \frac{1}{n}\sum_{k=1}^{n} H_k^*(Y_k,s,\rho(s)) < 2\epsilon, \tag{15}$$

e, aplicando o resultado (15) na desigualdade (13), segue que com probabilidade tendendo a 1 quando $n \to \infty$, $\frac{1}{n}\sum_{k=1}^{n} H_k(Y_k,s)$ está entre -2ϵ e 2ϵ e daí segue o resultado

$$\sup_{s\in S}\left\{\left|\frac{1}{n}\sum_{k=1}^{n} H_k(Y_k,s) - \frac{1}{n}\sum_{k=1}^{n} EH_k(Y_k,s)\right|\right\} \xrightarrow{P} 0.$$

■

Teorema A. 6. *Sejam $\boldsymbol{X}_k, k=1,2,\ldots$ vetores aleatórios p-dimensionais independentes com $E\boldsymbol{X}_k = \boldsymbol{0}$, $Cov(\boldsymbol{X}_k) = \Gamma_k$. Assumimos que:*

(a) $\quad \bar{\Gamma}_n = \frac{1}{n}\sum_{k=1}^{n}\Gamma_k \longrightarrow \bar{\Gamma}$, *onde $\bar{\Gamma}$ é definida positiva;*

(b) $\quad$ *Para algum $\delta > 0$, $\frac{1}{n^{(2+\delta)/2}}\sum_k E|\lambda'\boldsymbol{X}_k|^{2+\delta} \longrightarrow 0$, para todo $\lambda \in \mathbb{R}^p$.*

Então $\frac{1}{\sqrt{n}}\sum_k \boldsymbol{X}_k \xrightarrow{D} N(\boldsymbol{0},\bar{\Gamma})$.

Demonstração. Pela hipótese (b), podemos aplicar o Teorema de Liapuonov para o caso multivariado em $\sum_k \lambda'\mathbf{X}_k$, para todo vetor $\lambda \neq \mathbf{0}$ (ver [26]), obtendo

$$\frac{1}{(n\lambda'\bar{\Gamma}_n\lambda)^{1/2}}\sum_k \lambda'\mathbf{X}_k \xrightarrow{D} N(0,1).$$

Pela hipótese (a), $\lambda'\bar{\Gamma}_n\lambda \overset{n\to\infty}{\longrightarrow} \lambda'\bar{\Gamma}\lambda \neq 0$.

Além disso,

$$\frac{\lambda'}{n^{1/2}}\sum_k \mathbf{X}_k \xrightarrow{D} N(0,\lambda'\bar{\Gamma}\lambda),$$

para todo $\lambda \neq \mathbf{0}$.

Segue então das propriedades da distribuição normal multivariada (ver [29]) que

$$\frac{1}{\sqrt{n}}\sum_{k}\mathbf{X}_k \xrightarrow{D} N(\mathbf{0},\bar{\Gamma}).$$

■

Referências Bibliográficas

[1] Ash, R. B., Doléans-Dade, C. A., *Probability and Measure Theory*, 2nd Ed., Academic Press, New York, 2000.

[2] Balakrishnan, N., *Progressive censoring methodology: an appraisal*, Test, v. 16, p. 211-296, 2007.

[3] Balakrishnan, N., Aggarwala, R., *Progressive Censoring: theory, methods and applications*, Birkäuser, Boston, 2000.

[4] Balakrishnan, N., Lin, C.T., *Asymptotic properties of maximum likelihood estimators based on progressive Type-II censoring*, Metrika, v. 74, p. 349-360, 2011.

[5] Balakrishnan, N., Kannan, N., Lin, C. T., Wu, S.J.S., *Inference for the Extreme Value Distribution Under Progressive Type-II Censoring*, Journal of Statistical Computation and Simulation, v. 74, p. 25-45, 2004.

[6] Basak, P., Basak, I., Balakrishnan, N., *Estimation for the Three-Parameter Lognormal Distribuiton Based on Progressively Censored Data*, Computational Statistics and Data Analysis, v. 53, p. 3580-3592, 2009.

[7] Billingsley, P., *Convergence of Probability Measures*, 3rd Ed., Wiley, New York, 1995.

[8] Casella, G., Berger, R. L., *Statistical Inference*, 2nd Ed., Duxburry, California, 2002.

[9] Casella, G., Lehmann, E. L., *Theory of Point Estimation*, 2nd Ed., Springer, New York, 1998.

[10] Cohen, A.C., *Estimating the Mean and the Variance of Normal Populations from Singly and Doubly Truncated Samples*, Annals of Mathematical Statistics, v. 21, p. 557-569, 1950.

[11] Cohen, A. C., *Life testing and early failure*, Technometrics, v. 8, p. 539-549, 1966.

[12] Cohen, A.C., *Progressively Censored Samples in Life Testing*,Technometerics, v. 5, p. 327-339, 1963.

[13] Cohen, A.C., *Tables for Maximum Likelihood Estimates; Singly Truncated and Singly Censored Samples*, Technometerics, v. 3, p. 535-541, 1961.

[14] Cramér, H., *Mathematical methods os statistics*, Princeton University Press, Princeton, 1946.

[15] Dempster, A. P., Laird, N. M., Rudin, D. B., *Maximum Likelihood From Incomplete Data Via The EM Algorithm*, JR Stat Soc Series B, v. 39, p. 1-38, 1977.

[16] Ding, C., Yu, D., *Statistical Inference on Progressive Type-II Censored Data from Extreme-Value Distribution*, CSO 2012, p. 62-66, 2012.

[17] Fisher, R. A., *On the Mathematical Foundations of Theoretical Statistics*, Philosophical Transactions of the Royal Society, London, v. 222, p. 309-368, 1922.

[18] Gupta, A. K., *Estimation of the Mean and Standard Deviation of a Normal Population from a Censored Sample*, Biometrika, v. 39, p. 269-273, 1952.

[19] Gupta, R. D., Kundu, D., *Generalized Exponential Distributions*, Australian and New Zealand Journal of Statistics, v. 41, p. 173-188, 1999.

[20] Herd, R. G., *Estimation of the parameters of a population from a multi-censored sample*, Ph.D. Thesis, Iowa State College, Ames, Iowa, 1956.

[21] Hoadley, B., *Asymptotic Properties of Maximum Likelihood Estimators for the Independent not Identically Distributed Case*, Annals of Mathematical Statistics, v. 42, p. 1977-1991, 1971.

[22] Ismail, A. A., *Inference in the Generalized Exponencial Distribution under Partially Accelerated Tests with Progressive Type-II Censoring*, Theoretical and Applied Fracture Mechanics, v. 59, p. 49-56, 2012.

[23] Klein, J. P., Moeschberger, M. L., *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd Ed., Springer, New York, 2003.

[24] Lawless. J. F., *Statistical Models and Methods for Lifetime Data*, 2nd Ed., Wiley, New Jersey, 2003.

[25] Lehmann, E. L., *Elements of Large-Sample Theory*, Springer, New York, 1983.

[26] Loève, M., *Probability Theory*, 2nd Ed., Van Nostrand, Princeton, 1960.

[27] Louis, T. A., *Finding the Observed Information Matrix When Using the EM algorithm*, Journal of the Royal Statistical Society: Series B, v. 44, p. 226-233, 1982.

[28] Ng, H. K. T., Chan. P. S., Balakrishnan, N., *Estimation of parameters from progressively censored data using EM algorithm*, Computational Statistics and Data Analysis, v. 39, p. 371-386, 1968.

[29] Rao, C. R., *Linear Statistical Inference and Its Applications*, Wiley, New York, 1973.

[30] Sen, P. K., Singer, J. M., Lima, A. C. P., *From Finite Sample to Asymptotic Methods in Statistics*, Cambridge University Press, New York, 2010.

[31] Serfling, R. J., *Approximation Theorems of Mathematical Statistics*, Wiley, 1980

[32] Tanner, M. A, *Tools fos Satatistical Inference*, 3th Ed., Springer, New York, 1996.

[33] Wu, C. F. J., *On the Convergence Properties of the EM Algorithm*, The Annals of Statistics, v. 11, p. 95-103, 1983.