ESTIMAÇÃO POR PSEUDO MÁXIMA VEROSSIMILHANÇA

José Julio Flores Delgado

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

Área de Concentração: **Estatística**

Orientador: **Prof. Dr. Heleno Bolfarine**

Durante a elaboração deste trabalho o autor recebeu apoio financeiro do CNPq.

- São Paulo, julho de 1995 -

Estimação por Pseudo Máxima Verossimilhança

Este exemplar corresponde à redação final
da dissertação devidamente corrigida e
defendida por José Julio Flores Delgado
e aprovada pela comissão julgadora.

São Paulo, agosto 1995.

Banca examinadora:

- Prof. Dr. Heleno Bolfarine (Orientador) - IME - USP
- Prof. Dr. José Galvão Leite - IME - USP
- Prof$^{\underline{a}}$. Dr$^{\underline{a}}$. Beatriz Mendes - IM- UFRJ.

Agradecimentos

São muitas as pessoas às quais eu gostaria de agradecer, professores, colegas e amigos pela sua colaboração e incentivo, que nos tornou possível alcançar este objetivo.

Agradeço aos professores deste Instituto, assim, como aos da *Pontificia Universidad Católica del Perú* (PUCP), pela oportunidade, confiança e estímulo de poder melhorar minha formação acadêmica.

Sou grato em especial ao meu orientador, o Prof. Heleno Bolfarine, pela excelente orientação durante o curso de Mestrado e desta Dissertação.

Sou grato também ao Prof. Galvão pela atenção dispensada na elaboração deste trabalho e pelas suas críticas e sugestões.

Aos amigos Ana e Christiam por todo o apoio e amizade durante o período do mestrado.

Aos colegas e amigos do Instituto, em particular, aos amigos Gladys, Viviana, Péricles, Luiz, Héliton, Regina e Reinaldo. Pelo companherismo, obrigado.

Ao Walter pela excelente digitação deste trabalho.

Pela amizade, aos amigos Alfredo e Iván.

Finalmente, pelo apoio incondicional, agradeço aos meus irmãos e, em especial, aos meus pais.

Resumo

As dificuldades na obtenção das estimativas de máxima verossimilhança em problemas com parâmetros "nuisance" tem levado ao estudo de procedimentos de estimação alternativos, os quais têm o espírito do método de máxima verossimilhança. Um destes é o método de pseudo máxima verossimilhança.

O objetivo da dissertação é apresentar o método de pseudo máxima verossimilhança e discutir algumas das suas propriedades. Estudamos condições sob as quais os estimadores obtidos por este método são fortemente consistentes e assintoticamente normais.

Abstract

The difficulties in obtaining maximum likelihood estimates in problems with nuisance parameters, has led to the investigation of alternative estimation procedures which have the spirit of likelihood procedures. One of these, is the method of pseudo maximum likelihood estimation.

The purpose of this work is to present the method of pseudo maximum likelihood estimation and discuss some of its properties. We give conditions under which the estimators derived by this method are strongly consistent and asymptotically normal.

Conteúdo

Introdução

Esta monografia está baseada no artigo "Pseudo maximum likelihood estimation: theory and applications", de Gail Gong e Francisco J. Samaniego, publicado na revista *The Annals of Statistics*, vol. 9 (1981), 861–869.

Em problemas de estimação na presença de parâmetros "nuisance", a aplicação do método de máxima verossimilhança pode ser inviável. Um dos procedimentos alternativos para este tipo de problema e que conserva o espírito do método anterior, é o de pseudo máxima verossimilhança.

Basicamente, o método desenvolvido neste trabalho consiste na substituição dos parâmetros "nuisance" por estimadores consistentes e na solução do sistema de equações de verossimilhança correspondentes aos parâmetros de interesse.

Para o caso de um parâmetro de interesse, θ, e um outro "nuisance", π, Gong e Samaniego mostram, usando um estimador $\widehat{\pi}_n$, fracamente consistente de π que, sob certas condições de regularidade, o estimador de pseudo máxima verossimilhança é fracamente consistente e assintoticamente normal.

No presente texto, generalizamos os resultados acima mencionados, considerando porém estimadores fortemente consistentes dos parâmetros "nuisance". Como resultado, temos que o estimador de pseudo máxima verossimilhança é fortemente consistente. Temos também simplificações na maioria das demonstrações que apresentamos.

Incorporamos, ainda, o resultado apresentado por Parke na revista mencionada acima, vol. 14 (1985), 355–357, o qual estabelece que, sob certas condições de regularidade, se o estimador do parâmetro "nuisance" $\hat{\pi}_n$, e a função $\frac{1}{n}\mathcal{L}_n(\theta, \hat{\pi}_n)$ ($\mathcal{L}_n(\theta, \hat{\pi}_n)$ é o logaritmo da função de verossimilhança com π_n substituído por $\hat{\pi}_n$) possuem distribuição assintótica normal bivariada, então a covariância assintótica entre ambos é nula. Implicações deste resultado são consideradas no texto.

No Capítulo 1, fazemos uma descrição do problema e apresentamos um exemplo simples para ilustrar o método.

O Capítulo 2 trata o caso uniparamétrico, isto é, quando temos apenas um parâmetro de interesse e um outro "nuisance".

No Capítulo 3, generalizamos a teoria desenvolvida no Capítulo 2, para o caso multiparamétrico, isto é, podemos ter mais de um parâmetro de interesse e também mais de um parâmetro "nuisance".

No Capítulo 4, apresentamos aplicações do método a um modelo de convolução discreto.

Capítulo 1

Descrição do problema

1.1 O problema

O método de máxima verossimilhança é um dos mais conhecidos e utilizados na estimação dos parâmetros de um modelo estatístico. Cramér (1946) define o caso "regular", no qual as equações de verossimilhança tem uma raiz fortemente consistente e assintoticamente eficiente e normalmente distribuída. No entanto, este método tem mostrado algumas dificuldades que tornam inviável sua aplicação em algumas situações particulares. Por exemplo, quando os estimadores de máxima verossimilhança não têm uma fórmula explícita e tem-se que usar algum método computacional para obter estimativas, pode ocorrer o problema de uma convergência muito lenta. Outra dificuldade tem ocorrido na estimação com presença de parâmetros "nuisance". Isto tem originado o estudo de procedimentos de estimação alternativos, que conservam o espírito do método de máxima verossimilhança. Dentre outros, temos, por exemplo, o método da verossimilhança parcial introduzido e estudado por Cox (1975).

Uma revisão compreenssível da literatura existente sobre estimação na presença de parâmetros "nuisance" é feita por Basu (1977). As abordagens para estimação com parâmetros "nuisance" tem-se centralizado na eliminação de parâmetros "nuisance" usando condicionamento ou redução dos dados. Estas abordagens nem sempre podem ser aplicadas em problemas práticos de importância. Por exemplo, em alguns modelos de convolução, como os que serão tratados no Capítulo 4, as abordagens mencionadas anteriormente falham. Esta é, precisamente, a característica destes modelos, que levaram os autores à abordagem desenvolvida aqui. No entanto, existem, na literatura, outras aplicações desta abordagem como, por exemplo, Chuang e Cox (1985), que utilizam este método para estimar os parâmetros da distribuição de Dirichlet.

Em geral, a estimação por pseudo máxima verossimilhança consiste em substituir todos os parâmetros "nuisance" de um modelo por estimadores consistentes e resolver o sistema reduzido de equações de verossimilhança correspondentes aos parâmetros de interesse.

1.2 Um exemplo simples

Para ilustrar o método de pseudo-máxima verossimilhança de uma maneira simples, consideremos o problema de estimação dos parâmetros α e β de uma distribuição gama.

Seja $(X_1, \ldots, X_n)$ uma amostra aleatória de uma distribuição gama com parâmetros α e β, isto é, uma distribuição cuja função de densidade é dada por

$$f(x|\beta,\alpha) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}}\ , \quad x > 0;\ \alpha, \beta > 0.$$

A função de verossimilhança correspondente é

$$\mathcal{L}_n(\beta,\alpha) = \prod_{j=1}^{n} X_j^{\alpha-1} e^{-\sum_{j=1}^{n} X_j/\beta}[\Gamma(\alpha)]^{-n}\beta^{-n\alpha}$$

e, assim:

$$\log \mathcal{L}_n(\beta,\alpha) = (\alpha-1)\sum_{j=1}^{n} \ln X_j - \beta^{-1}\sum_{j=1}^{n} X_j - n\ln\Gamma(\alpha) - n\alpha\ln\beta\ .$$

Logo, se $\widehat{\alpha}_n$ e $\widehat{\beta}_n$ são os estimadores de máxima verossimilhança, então satisfazem as seguintes equações:

$$\psi(\widehat{\alpha}_n) - \ln(\widehat{\alpha}_n) - \ln\left[\left(\prod_{j=1}^{n} X_j\right)^{1/n} \Big/ \ \overline{X}_n\right] = 0 \tag{1.1}$$

$$\widehat{\beta}_n = \overline{X}_n/\widehat{\alpha}_n\ , \tag{1.2}$$

onde ψ é a função digama, isto é,

$$\psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha) \ . \tag{1.3}$$

Assim, para encontrar estimativas destes parâmetros, devemos resolver numericamente a equação (1.1) e depois calcular $\widehat{\beta}_n$ a partir da equação (1.2).

Podemos verificar que a família de distribuições gama pertence ao caso "regular". Assim, $\widehat{\alpha}_n$ e $\widehat{\beta}_n$ são fortemente consistentes e assintoticamente eficientes e normalmente distribuídos com

$$\sqrt{n}(\widehat{\beta}_n - \beta, \widehat{\alpha}_n - \alpha) \overset{\mathcal{D}}{\mapsto} \mathbf{W} = (W_1, W_2) \sim N_2(\mathbf{0}, I_{\beta,\alpha}^{-1}) \ , \tag{1.4}$$

onde $I_{\beta\alpha}$ é a matriz de informação de Fisher, isto é,

$$I_{\beta,\alpha} = \begin{pmatrix} \alpha/\beta^2 & 1/\beta \\ 1/\beta & \psi'(\alpha) \end{pmatrix} \tag{1.5}$$

e

$$\psi'(\alpha) = \frac{\partial}{\partial\alpha}\psi(\alpha) \ .$$

Propriedades sobre a função ψ podem ser encontradas em Bowman e Shenton (1988). Em particular, para α positivo, $\alpha\psi(\alpha) > 1$. Portanto, a matriz de informação de Fisher, $I_{\beta,\alpha}$, dada em (1.5), é definida positiva e

$$I_{\beta,\alpha}^{-1} = D_{\beta,\alpha}^{-1} \begin{pmatrix} \psi'(\alpha) & -1/\beta \\ -1/\beta & \alpha/\beta^2 \end{pmatrix} \ , \tag{1.6}$$

onde $D_{\beta,\alpha} = [\alpha\psi'(\alpha) - 1]/\beta^2$.

Agora, usemos o método de pseudo máxima verossimilhança para estimar o parâmetro de interesse β, com α sendo "nuisance". Para isto, consideramos $\widetilde{\alpha}_n$ um estimador fortemente consistente do parâmetro α, por exemplo, usando o método dos momentos, podemos considerar:

$$\widetilde{\alpha}_n = \overline{X}_n^2/S_n^2 \ , \tag{1.7}$$

onde $S_n^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2$.

Então, $\widetilde{\beta}_n$, o estimador de pseudo máxima verossimilhança de β, associado a $\widetilde{\alpha}_n$, deve maximizar a função de verossimilhança, quando o parâmetro α é substituído por seu estimador $\widetilde{\alpha}_n$, isto é,

$$\widetilde{\beta}_n = \max_{\beta>0} \mathcal{L}_n(\beta, \widetilde{\alpha}_n) \ . \tag{1.8}$$

Assim, para obter $\widetilde{\beta}_n$, as equações de verossimilhança (1.1) e (1.2) se reduzem a

$$\widetilde{\beta}_n = \overline{X}_n / \widetilde{\alpha}_n \tag{1.9}$$

e, daí, segue de (1.7) que

$$\widetilde{\beta}_n = S_n^2 / \overline{X}_n \ . \tag{1.10}$$

Neste caso, o estimador de pseudo máxima verossimilhança $\widetilde{\beta}_n$, coincide com o obtido pelo método dos momentos e, a partir de (1.10), podemos deduzir as seguintes propriedades:

(i) $\widetilde{\beta}_n$ é fortemente consistente, isto é, $\widetilde{\beta}_n \xrightarrow[n\uparrow\infty]{} \beta$, com probabilidade 1;

(ii) $\widetilde{\beta}_n$ é assintoticamente normal e

$$\sqrt{n}(\widetilde{\beta}_n - \beta) \overset{\mathcal{D}}{\mapsto} U \sim N(0, \sigma_{PMV}^2) \ , \tag{1.11}$$

com $\sigma_{PMV}^2 = 3\beta^2(\alpha + 1)/\alpha$.

No entanto, $\widetilde{\beta}_n$ não é assintoticamente eficiente como é $\widehat{\beta}_n$ o estimador de máxima verossimilhança. Segue de (1.5) e (1.11), que a eficiência assintótica relativa de $\widetilde{\beta}_n$ com respeito a $\widehat{\beta}_n$, $EAR(\widetilde{\beta}_n, \widehat{\beta}_n)$, é dada por

$$EAR(\widetilde{\beta}_n, \widehat{\beta}_n) = [\alpha\psi'(\alpha)]/\{3[\alpha\psi'(\alpha) - 1](\alpha + 1)\} \ . \tag{1.12}$$

Daqui, usando propriedades da função ψ (veja Bowman e Shenton, 1988), podemos verificar que para todo α positivo, $EAR(\widetilde{\beta}_n, \widehat{\beta}_n) < \frac{2}{3}$.

Nos capítulos 2 e 3 generalizamos os resultados obtidos neste exemplo, ou seja, mostramos que para famílias de distribuições "regulares", e estimando apropriadamente os parâmetros "nuisance", os estimadores de pseudo máxima verossimilhança são fortemente consistentes e assintoticamente normalmente distribuídos. No entanto, como será estabelecido, também a eficiência assintótica deles dependerá da eficiência correspondente dos estimadores dos parâmetros "nuisance", em particular, se eles forem assintoticamente eficientes, os estimadores de pseudo máxima verossimilhança também serão.

Capítulo 2

Caso uniparamétrico

Neste capítulo, definimos o estimador de pseudo máxima verossimilhança considerando dois parâmetros reais, um deles de interesse e o outro de incômodo ("nuisance"); e sob condições de "regularidade", provaremos suas propriedades de consistência forte e distribuição assintótica normal.

Primeiramente apresentamos algumas notações que serão usadas ao longo deste capítulo.

Seja $\mathcal{F} = \{F_{\theta,\pi}\}_{(\theta,\pi)\in\Theta\times\Pi}$ uma família de distribuições biparamétricas definida sobre o conjunto dos números reais, $\mathbb{R}$, sendo Θ e Π dois subconjuntos de $\mathbb{R}$. Para cada par $(\theta,\pi) \in \Theta\times\Pi$, $f(\cdot|\theta,\pi)$ denotará a função de densidade (ou de probabilidade) associada a $F_{\theta,\pi}$.

Consideraremos θ como parâmetro estrutural (ou de interesse) e π como "nuisance" (ou de incômodo).

Se $(X_1,\ldots,X_n)$ é uma amostra aleatória da distribuição $F_{\theta,\pi}$, com $(\theta,\pi)\in\Theta\times\Pi$:

$$\begin{aligned}
\mathcal{L}_n(\theta,\pi) &= \mathcal{L}_n(X_1,\ldots,X_n;\theta,\pi) = \ln\prod_{j=1}^{n} f(X_j|\theta,\pi) \\
&= \sum_{j=1}^{n}\ln f(X_j|\theta,\pi)
\end{aligned}$$

e

$$\overline{\mathcal{L}}_n(\theta,\pi) = \frac{1}{n}\mathcal{L}_n(\theta,\pi).$$

Definição 2.0.1 *Seja* $(X_1, \ldots, X_n)$ *uma amostra aleatória da distribuição* $F_{\theta,\pi}$ *e* $\widehat{\pi}_n = \widehat{\pi}_n(X_1, \ldots, X_n)$ *um estimador do parâmetro* π. *Se:*

$$\widehat{\theta}_n = \widehat{\theta}_n(\widehat{\pi}_n) = \widehat{\theta}_n(X_1, \ldots, X_n) = \max_{\theta \in \Theta} \overline{\mathcal{L}}_n(X_1, \ldots, X_n; \theta, \widehat{\pi}_n) \ ;$$

dizemos que $\widehat{\theta}_n$ *é um estimador de pseudo máxima verossimilhança do parâmetro* θ.

Segue da definição anterior que, se $\widehat{\pi}_n$ *é o estimador de máxima verossimilhança para* π, *então* $\widehat{\theta}_n$ *é também o estimador de máxima verossimilhança para* θ.

As condições de regularidade que consideramos para a família $\mathcal{F}$ são as seguintes:

(C1) Para cada par $(\theta, \pi) \in \Theta \times \Pi$, $f(\cdot|\theta, \pi)$ tem como suporte o espaço amostral $\mathcal{X} \subseteq \mathbb{R}$.

(C2) Para cada $x \in \mathcal{X}$, temos que para todo par $(\theta, \pi) \in \Theta \times \Pi$ estão definidas as seguintes derivadas parciais:

$$\frac{\partial^3}{\partial \theta^3} \ln f(x|\theta, \pi) \ , \quad \frac{\partial}{\partial \pi} \ln f(x|\theta, \pi) \ ,$$
$$\frac{\partial^3}{\partial \pi \partial \theta^2} \ln f(x|\theta, \pi) \quad \text{e} \quad \frac{\partial^3}{\partial \pi^2 \partial \theta} \ln f(x|\theta, \pi) \ .$$

(C3) Para cada par $(\theta_1, \pi_1) \in \Theta \times \Pi$ existem intervalos abertos $A_{\theta_1} \subseteq \Theta$ e $B_{\pi_1} \subseteq \Pi$ contendo θ_1 e π_1, respectivamente, e funções g_1, g_2 e g_3, definidas sobre $\mathcal{X}$, tais que $\forall i \in \{1,2,3\}$, $0 < \int_{\mathcal{X}} g_i(x) \mathrm{d}x < \infty$ e para todo par $(\theta, \pi) \in \Theta \times \Pi$ e todo $x \in \mathcal{X}$:

$$\left| \frac{\partial}{\partial \theta} f(x|\theta, \pi) \right| \leq g_1(x) \ ;$$
$$\left| \frac{\partial^2}{\partial \theta^2} f(x|\theta, \pi) \right| \leq g_2(x) \ ;$$
$$\left| \frac{\partial^2}{\partial \pi \partial \theta} f(x|\theta, \pi) \right| \leq g_3(x) \ .$$

(C4) Para todo par $(\theta, \pi) \in \Theta \times \Pi$:

$$0 < I_\theta = E_{\theta,\pi} \left[\frac{\partial}{\partial \theta} \ln f(X|\theta, \pi) \right]^2 < \infty$$

e

$$-\infty < I_{\theta,\pi} = E_{\theta,\pi}\left[\frac{\partial}{\partial\theta}\ln f(X|\theta,\pi)\frac{\partial}{\partial\pi}\ln f(X|\theta,\pi)\right] < \infty \ ;$$

onde $X \sim f(\cdot|\theta,\pi)$.

(C5) Para cada par $(\theta_1,\pi_1) \in \Theta \times \Pi$, existem funções M_1, M_2 e M_3, definidas sobre $\mathcal{X}$, tais que $\forall i \in \{1,2,3\}$, $E_{\theta_1,\pi_1}[M_i(X)] \in \mathbb{R}_+$ e para todo $x \in \mathcal{X}$:

(i) $\forall(\theta,\pi) \in A_{\theta_1} \times B_{\pi_1}$, $\left|\frac{\partial^3}{\partial\theta^3}\ln f(x|\theta,\pi)\right| \leq M_1(x)$;

(ii) $\forall\pi \in B_{\pi_1}$, $\left|\frac{\partial^3}{\partial\pi\partial\theta^2}\ln f(x|\theta_1,\pi)\right| \leq M_2(x)$;

(iii) $\forall\pi \in B_{\pi_1}$, $\left|\frac{\partial^3}{\partial\pi^2\partial\theta}\ln f(x|\theta_1,\pi)\right| \leq M_3(x)$.

onde A_{θ_1} e B_{π_1} são como em (C3) e $X \sim f(\cdot|\theta_1,\pi_1)$.

Observação 2.0.2 Estas condições são parte daquelas usadas para provar as propriedades assintóticas do estimador de máxima verossimilhança usual, quando o parâmetro a ser estimado é um vetor de duas componentes reais.

2.1 Consistência forte

Nesta seção provamos a consistência forte do estimador de pseudo máxima verossimilhança $\hat{\theta}_n(\hat{\pi}_n)$.

Resultados preliminares são apresentados nos lemas seguintes. Sejam θ_0 e π_0 os verdadeiros valores dos parâmetros.

Lema 2.1.1 *Seja $(X_1,\ldots,X_n)$ uma amostra aleatória da distribuição $F_{\theta_0,\pi_0} \in \mathcal{F}$. Então, sob as condições de regularidade (C1) - (C4), temos:*

(i) $\overline{\mathcal{L}}_\theta(\theta_0,\pi_0) = \frac{1}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta}\ln f(X_j|\theta_0,\pi_0) \xrightarrow{\text{q.c.}} 0$;

(ii) $\overline{\mathcal{L}}_{\theta\theta}(\theta_0,\pi_0) = \frac{1}{n}\sum_{j=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln f(X_j|\theta_0,\pi_0) \xrightarrow{\text{q.c.}} -I_{\theta_0}$;

(iii) $\overline{\mathcal{L}}_{\theta,\pi}(\theta_0,\pi_0)=\dfrac{1}{n}\sum_{j=1}^{n}\dfrac{\partial^2}{\partial\pi\partial\theta}\ln f(X_j|\theta_0,\pi_0)\xrightarrow{\text{q.c.}}-I_{\theta_0,\pi_0}$.

Prova: Das condições de regularidade (C3) e (C4) decorrem:

$$E_{\theta_0,\pi_0}\left[\frac{\partial}{\partial\theta}\ln f(X_1|\theta_0,\pi_0)\right]=0\ , \tag{2.1}$$

$$\begin{aligned} &E_{\theta_0,\pi_0}\left[\frac{\partial^2}{\partial\theta^2}\ln f(X_1|\theta_0,\pi_0)\right]=-E_{\theta_0,\pi_0}\left[\frac{\partial}{\partial\theta}\ln f(X_1|\theta_0,\pi_0)\right]^2 \\ &=-I_{\theta_0}\in\mathbb{R}_-\ , \end{aligned} \tag{2.2}$$

$$\begin{aligned} &E_{\theta_0,\pi_0}\left[\frac{\partial^2}{\partial\pi\partial\theta}\ln f(X_1|\theta_0,\pi_0)\right] \\ &=-E_{\theta_0,\pi_0}\left[\frac{\partial}{\partial\theta}\ln f(X_1|\theta_0,\pi_0)\frac{\partial}{\partial\theta}\ln f(X_1|\theta_0,\pi_0)\right] \\ &=-I_{\theta_0\pi_0}\in\mathbb{R}\ . \end{aligned} \tag{2.3}$$

Logo, como $X_1,\ldots,X_n$ são variáveis aleatórias independentes e identicamente distribuídas, então pela lei forte dos grandes números de Kolmogorov (James, 1981) e as expressões acima, provam-se (i), (ii) e (iii). □

Lema 2.1.2 *Sejam $(X_1,\ldots,X_n)$ uma amostra aleatória da distribuição $F_{\theta_0,\pi_0}\in\mathcal{F}$, $\widetilde{\theta}_n=\widetilde{\theta}_n(X_1,\ldots,X_n)$ e $\widehat{\pi}_n=\widehat{\pi}_n(X_1,\ldots,X_n)$ dois estimadores fortemente consistentes para θ_0 e π_0, respectivamente. Então, para todo $\delta>0$ e sob as condições de regularidade (C1), (C2) e (C5), temos com probabilidade 1 que, para n suficientemente grande:*

(i) $|\overline{\mathcal{L}}_{\theta\theta\theta}(\widetilde{\theta}_n,\widehat{\pi}_n)|<EM_1(X_1)+\delta$;

(ii) $|\overline{\mathcal{L}}_{\theta\theta\pi}(\theta_0,\widehat{\pi}_n)|<EM_2(X_1)+\delta$;

(iii) $|\overline{\mathcal{L}}_{\theta\pi\pi}(\theta_0,\widehat{\pi}_n)|<EM_3(X_1)+\delta$;

onde M_1, M_2 e M_3 são como na condição de regularidade (C5).

Prova: Prova de (i): Sejam A_{θ_0} e B_{π_0} como na condição de regularidade (C5). Então, como $(\theta_0,\pi_0)\in A_{\theta_0}\times B_{\pi_0}$ e $(\widetilde{\theta}_n,\widehat{\pi}_n)\xrightarrow{\text{q.c.}}(\theta_0,\pi_0)$, segue que, com probabilidade 1, $(\widetilde{\theta}_n,\widehat{\pi}_n)\in$

$A_{\theta_0} \times B_{\pi_0}$ para n suficientemente grande; assim, usando a condição de regularidade (C5) - (i), temos com probabilidade 1 que, para n suficientemente grande:

$$
\begin{aligned}
|\overline{\mathcal{L}}_{\theta\theta\theta}(\widetilde{\theta}_n, \widehat{\pi}_n)| &= \left| \frac{1}{n} \sum_{j=1}^{n} \frac{\partial^3}{\partial \theta^3} \ln f(X_j | \widetilde{\theta}_n, \widehat{\pi}_n) \right| \\
&\leq \frac{1}{n} \sum_{j=1}^{n} \left| \frac{\partial^3}{\partial \theta^3} \ln f(X_j | \widetilde{\theta}_n, \widehat{\pi}_n) \right| \\
&\leq \frac{1}{n} \sum_{j=1}^{n} M_1(X_j) \ . \qquad (2.4)
\end{aligned}
$$

Por outro lado, como as variáveis $X_1, \ldots, X_n$ são independentes e identicamente distribuídas e, pela condição de regularidade (C5), $EM_1(X_1) \in \mathbb{R}_+$, segue da lei forte dos grandes números de Kolmogorov que $\frac{1}{n}\sum_{j=1}^{n} M_1(X_j) \xrightarrow{\text{q.c.}} EM_1(X_1)$; assim, para $\delta > 0$, temos com probabilidade 1 que, para n suficientemente grande:

$$
\frac{1}{n} \sum_{j=1}^{n} M_1(X_j) < EM_1(X_1) + \delta \ . \qquad (2.5)
$$

A prova de (i) segue de (2.4) e (2.5).

De maneira análoga podemos provar (ii) e (iii). □

Vejamos agora que o Lema 2.1.1 ainda continua valendo se em (i), (ii) ou (iii) substituirmos π_0 por $\widehat{\pi}_n$, qualquer estimador fortemente consistente de π_0. Mais precisamente:

Lema 2.1.3 *Sejam $(X_1, \ldots, X_n)$ uma amostra aleatória da distribuição $F_{\theta_0,\pi_0} \in \mathcal{F}$ e $\widehat{\pi}_n = \widehat{\pi}_n(X_1, \ldots, X_n)$ um estimador fortemente consistente de π_0. Então, sob as condições de regularidade (C1) - (C5), temos:*

(i) $\overline{\mathcal{L}}_{\theta}(\theta_0, \widehat{\pi}_n) = \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} \frac{\partial}{\partial \theta} \ln f(X_j | \theta_0, \widehat{\pi}_n) \xrightarrow{\text{q.c.}} 0$;

(ii) $\overline{\mathcal{L}}_{\theta\theta}(\theta_0, \widehat{\pi}_n) = \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} \frac{\partial^2}{\partial \theta^2} \ln f(X_j | \theta_0, \widehat{\pi}_n) \xrightarrow{\text{q.c.}} -I_{\theta_0}$;

(iii) $\overline{\mathcal{L}}_{\theta\pi}(\theta_0, \widehat{\pi}_n) = \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} \frac{\partial^2}{\partial \pi \partial \theta} \ln f(X_j | \theta_0, \widehat{\pi}_n) \xrightarrow{\text{q.c.}} -I_{\theta_0,\pi_0}$.

Prova: Prova de (i): Pela condição de regularidade (C2) e o teorema de Taylor com resto de Lagrange (Serfling, 1980), temos para cada $n \in \mathbb{N}_+$:

$$\overline{\mathcal{L}}_\theta(\theta_0, \widehat{\pi}_n) = \overline{\mathcal{L}}_\theta(\theta_0, \pi_0) + (\widehat{\pi}_n - \pi_0)\overline{\mathcal{L}}_{\theta\pi}(\theta_0, \pi_0) + \frac{1}{2}(\widehat{\pi}_n - \pi_0)^2\overline{\mathcal{L}}_{\theta\pi\pi}(\theta_0, \pi_n^*) \ , \tag{2.6}$$

onde π_n^* está entre π_0 e $\widehat{\pi}_n$.

Desta última igualdade decorre (i), pois cada um dos somandos do lado direito de (2.6) convergem a zero quase certamente. Com efeito:

$\overline{\mathcal{L}}_\theta(\theta_0, \pi_0) \xrightarrow{\text{q.c.}} 0$, pelo Lema 2.1.1 (i);

$(\widehat{\pi}_n - \pi_0)\overline{\mathcal{L}}_{\theta\pi}(\theta_0, \pi_0) \xrightarrow{\text{q.c.}} 0$, pois $\widehat{\pi}_n \xrightarrow{\text{q.c.}} \pi_0$, pela hipótese, e $\overline{\mathcal{L}}_{\theta\pi}(\theta_0, \pi_0) \xrightarrow{\text{q.c.}} -I_{\theta_0,\pi_0} \in \mathbb{R}$, pelo Lema 2.1.1 (iii);

$\frac{1}{2}(\widehat{\pi}_n - \pi_0)^2\overline{\mathcal{L}}_{\theta\pi\pi}(\theta_0, \pi_n^*) \xrightarrow{\text{q.c.}} 0$, pois $\widehat{\pi}_n \xrightarrow{\text{q.c.}} \pi_0$ e, com probabilidade 1, para n suficientemente grande, $|\overline{\mathcal{L}}_{\theta\pi\pi}(\theta_0, \pi_n^*)| \leq EM_3(X_1) + 1$, que segue do Lema 2.1.2 (iii), observando que $\pi_n^* \xrightarrow{\text{q.c.}} \pi_0$, já que π_n^* está entre π_0 e $\widehat{\pi}_n$ e $\widehat{\pi}_n \xrightarrow{\text{q.c.}} \pi_0$.

Isso prova (i).

Analogamente, podemos provar (ii) e (iii). □

Lema 2.1.4 *Sejam $(X_1, \ldots, X_n)$ uma amostra aleatória de $F_{\theta_0,\pi_0} \in \mathcal{F}$ e $\widehat{\pi}_n = \widehat{\pi}_n(X_1, \ldots, X_n)$ um estimador fortemente consistente de π_0. Então, sob as condições de regularidade (C1) - (C5), temos com probabilidade 1 que, para cada $\varepsilon > 0$ suficientemente pequeno,*

$$\forall\theta \in \{\theta_0 - \varepsilon,\ \theta_0 + \varepsilon\} \ , \quad \overline{\mathcal{L}}_n(\theta, \widehat{\pi}_n) < \overline{\mathcal{L}}_n(\theta_0, \widehat{\pi}_n) \ ,$$

para n suficientemente grande.

Prova: Pela condição de regularidade (C2) e o teorema de Taylor com resto de Lagrange, temos, para $\varepsilon > 0$, $\theta \in \{\theta_0 - \varepsilon, \theta_0 + \varepsilon\}$, e para cada $n \in \mathbb{N}_+$, que:

$$\begin{aligned} &\overline{\mathcal{L}}_n(\theta, \widehat{\pi}_n) - \overline{\mathcal{L}}_n(\theta_0, \widehat{\pi}_n) = (\theta - \theta_0)\overline{\mathcal{L}}_\theta(\theta_0, \widehat{\pi}_n) \\ &+\frac{1}{2}(\theta - \theta_0)^2\overline{\mathcal{L}}_{\theta\theta}(\theta_0, \widehat{\pi}_n) + \frac{1}{6}(\theta - \theta_0)^3\overline{\mathcal{L}}_{\theta\theta\theta}(\theta_n^*, \widehat{\pi}_n) \\ &= S_{1n} + S_{2n} + S_{3n} \ , \end{aligned} \tag{2.7}$$

onde:

$$\begin{aligned}
S_{1n} &= (\theta-\theta_0)\overline{\mathcal{L}}_\theta(\theta_0,\widehat{\pi}_n) \ , \\
S_{2n} &= \frac{1}{2}(\theta-\theta_0)^2\overline{\mathcal{L}}_{\theta\theta}(\theta_0,\widehat{\pi}_n) = \frac{1}{2}\varepsilon^2\overline{\mathcal{L}}_{\theta\theta}(\theta_0,\widehat{\pi}_n) \ , \\
S_{3n} &= \frac{1}{6}(\theta-\theta_0)^3\overline{\mathcal{L}}_{\theta\theta\theta}(\theta_n^*,\widehat{\pi}_n)
\end{aligned}$$

e $\widehat{\theta}_n^*$ está entre θ_0 e θ.

Provaremos abaixo que com probabilidade 1 para $\varepsilon > 0$ suficientemente pequeno tem-se, para $\theta \in \{\theta_0-\varepsilon, \theta_0+\varepsilon\}$ e n suficientemente grande:

$$|S_{1n}| < \varepsilon^3 \ , \tag{2.8}$$

$$S_{2n} < -\frac{1}{4}\varepsilon^2 I_{\theta_0} \ , \tag{2.9}$$

$$|S_{3n}| < \varepsilon^3 EM_1(X_1) \ , \tag{2.10}$$

onde M_1 é como na condição de regularidade (C5). Como conseqüência, as desigualdades anteriores, junto com (2.7), implicam que se $0 < \varepsilon < \frac{1}{4}[1+EM_1(X_1)]^{-1}I_{\theta_0}$, então, com probabilidade 1, para $\theta \in \{\theta_0-\varepsilon, \theta_0+\varepsilon\}$ e n suficientemente grande:

$$\overline{\mathcal{L}}_n(\theta,\widehat{\pi}_n) - \overline{\mathcal{L}}_n(\theta_0,\widehat{\pi}_n) = S_{1n}+S_{2n}+S_{3n} < 0 \ ;$$

o que prova o lema.

Prova de (2.8): $S_{1n} = (\theta-\theta_0)\overline{\mathcal{L}}_\theta(\theta_0,\widehat{\pi}_n)$. Pelo Lema 2.1.3 (i), $\overline{\mathcal{L}}_\theta(\theta_0,\widehat{\pi}_n) \xrightarrow{\text{q.c.}} 0$, portanto, $S_{1n} \xrightarrow{\text{q.c.}} 0$ e assim, segue (2.8). □

Prova de (2.9): $S_{2n} = \frac{1}{2}\varepsilon^2\overline{\mathcal{L}}_{\theta\theta}(\theta_0,\widehat{\pi}_n)$. Pelo Lema 2.1.3 (ii), $\overline{\mathcal{L}}_{\theta\theta}(\theta_0,\widehat{\pi}_n) \xrightarrow{\text{q.c.}} -I_{\theta_0}$, daí, como $I_{\theta_0} \in \mathbb{R}_+$, segundo a condição de regularidade (C4), segue que com probabilidade 1 para n suficientemente grande, $\overline{\mathcal{L}}_{\theta\theta}(\theta_0,\widehat{\pi}_n) < -I_{\theta_0} + \frac{1}{2}I_{\theta_0} = -\frac{1}{2}I_{\theta_0}$ e, portanto, $S_{2n} < -\frac{1}{4}\varepsilon^2 I_{\theta_0}$. □

Prova de (2.10): $S_{3n} = \frac{1}{6}(\theta-\theta_0)^3\overline{\mathcal{L}}_{\theta\theta\theta}(\theta_n^*,\widehat{\pi}_n)$. Sejam A_{θ_0} e B_{π_0} como na condição de regularidade (C5). Como $(\theta_0,\pi_0) \in A_{\theta_0}\times B_{\pi_0}$, θ_n^* está entre θ e θ_0 e $\widehat{\pi}_n \xrightarrow{\text{q.c.}} \pi_0$;

então para cada $\varepsilon > 0$ suficientemente pequeno, temos com probabilidade 1 que, para $\theta \in \{\theta_0 - \varepsilon, \theta_0 + \varepsilon\}$ e n suficientemente grande, $(\theta_n^*, \widehat{\pi}_n) \in A_{\theta_0} \times B_{\pi_0}$ e, assim, procedendo como na prova do Lema 2.1.2 (i), substituindo $\widehat{\theta}_n$ por θ_n^*,

$$|\overline{\mathcal{L}}_{\theta\theta\theta}(\theta_n^*, \widehat{\pi}_n)| < EM_1(X_1) + 5EM_1(X_1) = 6EM_1(X_1) \ .$$

Daqui decorre (2.10).

O resultado principal da seção é enunciado e provado a seguir.

Teorema 2.1.5 *Sejam* $(X_1, \ldots, X_n)$ *uma amostra aleatória de* $F_{\theta_0,\pi_0} \in \mathcal{F}$ *e* $\widehat{\pi}_0 = \widehat{\pi}_n(X_1, \ldots, X_n)$ *um estimador fortemente consistente de* π_0*. Então, sob as condições de regularidade (C1) - (C5), existe uma seqüência* $(\widetilde{\theta}_n)_{n\geq 1}$ *tal que, com probabilidade 1:*

(i) $\overline{\mathcal{L}}_\theta(\widetilde{\theta}_n, \widehat{\pi}_n) = 0$, *para n suficientemente grande;*

(ii) $\widetilde{\theta}_n \underset{n\uparrow\infty}{\longrightarrow} \theta_0$.

Prova: Da condição de regularidade (C2) segue que para cada $\varepsilon > 0$ e $n \in \mathbb{N}_+$, existe $\widetilde{\theta}_{n,\varepsilon} \in [\theta_0 - \varepsilon, \theta_0 + \varepsilon]$ tal que

$$\overline{\mathcal{L}}_n(\widetilde{\theta}_{n,\varepsilon}, \widehat{\pi}_n) = \max\left\{\overline{\mathcal{L}}_n(\theta, \widehat{\pi}_n) : \theta \in [\theta_0 - \varepsilon, \theta_0 + \varepsilon]\right\} \ .$$

Assim, pelo Lema 2.1.4 e a condição de regularidade (C2), temos, com probabilidade 1, que para cada $\varepsilon > 0$ suficientemente pequeno:

$$\widetilde{\theta}_{n,\varepsilon} \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon) \quad \text{e} \quad \overline{\mathcal{L}}_\theta(\widetilde{\theta}_{n,\varepsilon}, \widehat{\pi}_n) = 0 \ , \tag{2.11}$$

para n suficientemente grande.

Seja $\varepsilon_0 > 0$ satisfazendo (2.11) e definamos para cada n suficientemente grande $\widetilde{\theta}_n$, de modo que:

$$|\widetilde{\theta}_n - \theta_0| = \inf\left\{|\theta - \theta_0| : \theta \in [\theta_0 - \varepsilon_0, \theta_0 + \varepsilon_0] \ \text{ e } \ \overline{\mathcal{L}}_\theta(\theta, \widehat{\pi}_n) = 0\right\} \ . \tag{2.12}$$

Então, pela continuidade das funções valor absoluto e $\overline{\mathcal{L}}_\theta$, temos com probabilidade 1 que para n suficientemente grande:

$$\widetilde{\theta}_n \in [\theta_0 - \varepsilon_0, \theta_0 + \varepsilon_0] \quad \text{e} \quad \overline{\mathcal{L}}_\theta(\widetilde{\theta}_n, \widehat{\pi}_n) = 0 \ . \tag{2.13}$$

Por último, seja $\varepsilon \in (0, \varepsilon_0)$; então, como $\varepsilon < \varepsilon_0$, segue de (2.11) que, com probabilidade 1, para cada n suficientemente grande, $|\widetilde{\theta}_{n,\varepsilon} - \theta_0| < \varepsilon < \varepsilon_0$ e $\overline{\mathcal{L}}_\theta(\widetilde{\theta}_{n,\varepsilon}, \widehat{\pi}_n) = 0$, de modo que, por (2.12), $|\widetilde{\theta}_n - \theta_0| \leq |\widetilde{\theta}_{n,\varepsilon} - \theta_0| < \varepsilon$. Portanto, $\widehat{\theta}_n \xrightarrow[n \uparrow \infty]{} \theta_0$, com probabilidade 1. □

Observação 2.1.6 O teorema anterior não determina a existência de um estimador fortemente consistente, pois a seqüência obtida depende de θ_0. Uma exceção ocorre no caso em que a equação $\mathcal{L}_\theta(\theta, \widehat{\pi}_n) = 0$, vista como função de θ, tenha para cada $n \in \mathbb{N}_+$, uma única raiz.

Corolário 2.1.7 *Sejam $(X_1, \ldots, X_n)$ uma amostra aleatória de $F_{\theta_0,\pi_0} \in \mathcal{F}$ e $\widehat{\pi}_n = \widehat{\pi}_n(X_1, \ldots, X_n)$ um estimador fortemente consistente de π_0. Suponhamos que as condições de regularidade (C1) - (C5) são satisfeitas pela família $\mathcal{F}$ e que para cada $n \in \mathbb{N}_+$, a equação $\mathcal{L}_\theta(\theta, \widehat{\pi}_n) = 0$, vista como função de θ, tenha uma única raiz. Então, o estimador de pseudo máxima verossimilhança, $\widehat{\theta}_n$, é fortemente consistente para θ_0.*

Prova: Pela condição de regularidade (C2), para cada $n \in \mathbb{N}_+$, $\mathcal{L}_\theta(\widehat{\theta}_n, \widehat{\pi}_n) = 0$; assim $\widehat{\theta}_n$ é a única raiz da equação $\mathcal{L}_\theta(\theta, \widehat{\pi}_n) = 0$ e o corolário decorre do Teorema 2.1.5. □

2.2 Distribuição assintótica

Nesta seção estudaremos a distribuição assintótica do estimador de pseudo máxima verossimilhança. Como na seção anterior, resultados preliminares são apresentados nos lemas seguintes.

Lema 2.2.1 *Seja $(X_1, \ldots, X_n)$ uma amostra aleatória da distribuição $F_{\theta_0,\pi_0} \in \mathcal{F}$. Sejam $\widetilde{\theta}_{in} = \widetilde{\theta}_{in}(X_1, \ldots, X_n)$, $i = 1, 2$, dois estimadores fortemente consistentes de θ_0 e $\widehat{\pi}_{in} =$*

$\widehat{\pi}_{in}(X_1,\ldots,X_n)$, $i=1,2$, *dois estimadores fortemente consistentes de* π_0. *Então, sob as condições de regularidade (C1), (C2) e (C5), temos*

(i) $(\widetilde{\theta}_{1n}-\theta_0)\overline{\mathcal{L}}_{\theta\theta\theta}(\widetilde{\theta}_{2n},\widehat{\pi}_{1n}) \xrightarrow{q.c.} 0$;

(ii) $(\widehat{\pi}_{1n}-\pi_0)\overline{\mathcal{L}}_{\theta\pi\pi}(\theta_0,\widehat{\pi}_{2n}) \xrightarrow{q.c.} 0$.

Prova: (i) segue do Lema 2.1.2 (i); e (ii) do Lema 2.1.2 (iii). □

Lema 2.2.2 *Sejam* $(X_1,\ldots,X_n)$ *uma amostra aleatória da distribuição* $F_{\theta_0,\pi_0}\in\mathcal{F}$ *e* $\widehat{\pi}_n=\widehat{\pi}_n(X_1,\ldots,X_n)$ *tais que*

$$\sqrt{n}(\overline{\mathcal{L}}_\theta(\theta_0,\pi_0),\widehat{\pi}_n-\pi_0) \overset{\mathcal{D}}{\mapsto} \mathbf{W}\sim N_2(\mathbf{0},\Sigma) ,$$

onde $\Sigma=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}$.

Então, se $\mathcal{F}$ *satisfaz as condições de regularidade (C1) - (C5), temos*

(i) $\sigma_{11}=I_{\theta_0}$;

(ii) $\sigma_{12}=0$.

Prova: Como $\overline{\mathcal{L}}_\theta(\theta_0,\pi_0)=\frac{1}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta}\ln f(X_j|\theta_0,\pi_0)$, com $\frac{\partial}{\partial\theta}\ln f(X_j|\theta_0,\pi_0)$, $i=1,\ldots,n$, são variáveis aleatórias independentes e identicamente distribuídas tendo média zero, segundo (2.1), e variância I_{θ_0}, segundo a condição de regularidade (C4); então pelo teorema do limite central (James, 1981) segue que $\sqrt{n}\,\overline{\mathcal{L}}_\theta(\theta_0,\pi_0)\overset{\mathcal{D}}{\mapsto}N(0,I_{\theta_0})$, o que prova (i).

Para provar (ii), seja $(\widetilde{\theta}_n)_{n\geq 1}$ uma seqüência tal que com probabilidade 1, $\overline{\mathcal{L}}_\theta(\widetilde{\theta}_n,\pi_0)=0$ para n suficientemente grande, e $\widetilde{\theta}_n \xrightarrow[n\uparrow\infty]{} \theta_0$. A existência de tal seqüência está garantida pelo Teorema 2.1.5 e o fato de que $\pi_0 \xrightarrow[n\uparrow\infty]{q.c.} \pi_0$.

Assim, da condição de regularidade (C2) e da fórmula de Taylor com resto de Lagrange, segue com probabilidade 1 que, para n suficientemente grande:

$$0=\sqrt{n}\,\overline{\mathcal{L}}_\theta(\theta_0,\pi_0)+\sqrt{n}(\widetilde{\theta}_n-\theta_0)\overline{\mathcal{L}}_{\theta\theta}(\theta_0,\pi_0)+\frac{1}{2}\sqrt{n}(\widetilde{\theta}_n-\theta_0)^2\overline{\mathcal{L}}_{\theta\theta\theta}(\theta_n^*,\pi_0) ,$$

onde θ_n^* está entre θ_0 e $\widetilde{\theta}_n$. Portanto, com probabilidade 1:

$$\sqrt{n}\mathcal{L}_\theta(\theta_0, \pi_0) + X_n\sqrt{n}(\widetilde{\theta}_n - \theta_0) \underset{n\uparrow\infty}{\longrightarrow} 0 , \tag{2.14}$$

onde $X_n = \overline{\mathcal{L}}_{\theta\theta}(\theta_0, \pi_0) + \frac{1}{2}(\widetilde{\theta}_n - \theta_0)\overline{\mathcal{L}}_{\theta\theta\theta}(\theta_n^*, \pi_0)$. Então, por (2.14), pelo teorema de Slutsky (James, 1981) e pela hipótese, segue que

$$\sqrt{n}(-X_n(\widetilde{\theta}_n - \theta_0), \widehat{\pi}_n - \pi_0) \overset{\mathcal{D}}{\mapsto} \mathbf{W} \sim N_2(\mathbf{0}, \Sigma) . \tag{2.15}$$

Além disso, como θ_n^* está entre θ_0 e $\widetilde{\theta}_n$ e $\widetilde{\theta}_n \xrightarrow{\text{q.c.}} \theta_0$, temos que $\theta_n^* \xrightarrow{\text{q.c.}} \theta_0$ e daí, pelo Lema 2.2.1 (i), segue que $-X_n \xrightarrow{\text{q.c.}} I_{\theta_0}$, com $I_{\theta_0} \in \mathbb{R}_+$, segundo a condição de regularidade (C4). Portanto, pelo teorema de Slutsky e (2.15), segue que

$$\begin{aligned}\sqrt{n}(\widetilde{\theta}_n - \theta_0, \widehat{\pi}_n - \pi_0) &= \sqrt{n}(-X_n(\widetilde{\theta}_n - \theta_0), \widehat{\pi}_n - \pi_0)\begin{pmatrix} -X_n^{-1} & 0 \\ 0 & 1 \end{pmatrix} \\ &\overset{\mathcal{D}}{\mapsto} WC \sim N_2(\mathbf{0}, C\Sigma C) ,\end{aligned}$$

onde $C = \begin{pmatrix} I_{\theta_0}^{-1} & 0 \\ 0 & 1 \end{pmatrix}$,
ou seja,

$$\sqrt{n}(\widetilde{\theta}_n - \theta_0, \widehat{\pi}_n - \pi_0) \overset{\mathcal{D}}{\mapsto} \mathbf{W}C \sim N_2(\mathbf{0}, \Sigma') , \tag{2.16}$$

onde $\Sigma' = C\Sigma C = \begin{pmatrix} I_{\theta_0}^{-2}\sigma_{11} & I_{\theta_0}^{-1}\sigma_{12} \\ I_{\theta_0}^{-1}\sigma_{12} & \sigma_{22} \end{pmatrix}$,
ou como por (i) $\sigma_{11} = I_{\theta_0}$,

$$\Sigma' = \begin{pmatrix} I_{\theta_0}^{-1} & I_{\theta_0}^{-1}\sigma_{12} \\ I_{\theta_0}^{-1}\sigma_{12} & \sigma_{22} \end{pmatrix} .$$

De (2.16) e do teorema de Slutsky, segue que $\forall c \in \mathbb{R}$:

$$\sqrt{n}(\widetilde{\theta}_n + c(\widehat{\pi}_n - \pi_0) - \theta_0) \overset{\mathcal{D}}{\mapsto} V_c \sim N(0, v_c) , \tag{2.17}$$

onde $v_c = I_{\theta_0}^{-1} + c^2\sigma_{22} + 2cI_{\theta_0}^{-1}\sigma_{12}$; e, em particular:

$$\sqrt{n}(\widetilde{\theta}_n - \theta_0) \overset{\mathcal{D}}{\mapsto} V_0 \sim N(0, I_{\theta_0}^{-1}) .$$

Deste resultado e das condições de regularidade (C1) - (C5), decorre que $\widetilde{\theta}_n$ é assintóticamente eficiente para θ_0 supondo π_0 conhecido (veja, por exemplo, Lehmann, 1984 e Bahadur, 1964), isto é, se $(\delta_n)_{n\geq 1}$ satisfaz $\sqrt{n}(\delta_n - \theta_0) \overset{\mathcal{D}}{\mapsto} N(0, v)$, então,

$$v \geq I_{\theta_0}^{-1} \,. \tag{2.18}$$

De (2.17) e (2.18) decorre que $\sigma_{12} = 0$; pois caso contrário, tomando $\delta_n = \widetilde{\theta}_n + c(\widehat{\pi}_n - \pi_0)$, com c um valor qualquer tal que $c^2\sigma_{22} + 2cI_{\theta_0}^{-1}\sigma_{12} < 0$, teríamos por (2.17) que $\sqrt{n}(\delta_n - \theta_0) \overset{\mathcal{D}}{\mapsto} V_c \sim N(0, v_c)$, com $v_c < I_{\theta_0}^{-1}$, o que contradiz (2.18). Isto prova (ii) e o lema. □

Observação 2.2.3 Gong e Samaniego (1981) em sua demonstração original não percebem que $\sigma_{12} = 0$. Tal fato foi provado em geral por Parke (1985).

Lema 2.2.4 *Sejam $(X_1, \ldots, X_n)$ uma amostra aleatória da distribuição $F_{\theta_0,\pi_0} \in \mathcal{F}$ e $\widehat{\pi}_n = \widehat{\pi}_n(X_1, \ldots, X_n)$ um estimador fortemente consistente de π_0. Então, sob as condições de regularidade (C1) - (C5), existe uma seqüência $(Y_n)_{n\geq 1} = (Y_n(X_1, \ldots, X_n))_{n\geq 1}$, tal que $Y_n \xrightarrow{\text{q.c.}} -I_{\theta_0,\pi_0}$ e com probabilidade 1 tem-se que:*

$$\sqrt{n}\,\overline{\mathcal{L}}_\theta(\theta_0, \widehat{\pi}_n) = \sqrt{n}\,\overline{\mathcal{L}}_\theta(\theta_0, \pi_0) + \sqrt{n}(\widehat{\pi}_n - \pi_0)Y_n \ ,$$

para n suficientemente grande.

Prova: Pela condição de regularidade (C2) e a fórmula de Taylor com resto de Lagrange temos que

$$\sqrt{n}\,\overline{\mathcal{L}}_\theta(\theta_0, \widehat{\pi}_n) = \sqrt{n}\,\overline{\mathcal{L}}_\theta(\theta_0, \pi_0) + \sqrt{n}(\widehat{\pi}_n - \pi_0)Y_n \ ,$$

onde $Y_n = \overline{\mathcal{L}}_{\theta\pi}(\theta_0, \pi_0) + \frac{1}{2}(\widehat{\pi}_n - \pi_0)\overline{\mathcal{L}}_{\theta\pi\pi}(\theta_0, \pi_n^*)$ e π_n^* está entre π_0 e $\widehat{\pi}_n$.

Além disso, como $\widehat{\pi}_n \xrightarrow{\text{q.c.}} \pi_0$ segue que $\pi_n^* \xrightarrow{\text{q.c.}} \pi_0$ e assim, decorre dos lemas 2.2.1 (ii) e 2.1.1 (iii) que $Y_n \xrightarrow{\text{q.c.}} -I_{\theta_0,\pi_0}$. Isso prova o lema. □

O resultado principal desta seção é enunciado e provado a seguir.

Teorema 2.2.5 *Sejam* $(X_1, \ldots, X_n)$ *uma amostra aleatória da distribuição* $F_{\theta_0,\pi_0} \in \mathcal{F}$ *e* $\widehat{\pi}_n = \widehat{\pi}_n(X_1, \ldots, X_n)$ *um estimador fortemente consistente de* π_0*, tais que*

$$\sqrt{n}(\overline{\mathcal{L}}_\theta(\theta_0, \pi_0), \widehat{\pi}_n - \pi_0) \overset{\mathcal{D}}{\mapsto} \mathbf{W} \sim N_2(\mathbf{0}, \Sigma) \ , \tag{2.19}$$

onde $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$.

Seja $(\widetilde{\theta}_n)_{n\geq 1} = (\widehat{\theta}_n(X_1, \ldots, X_n))_{n\geq 1}$ *tal que* $\widetilde{\theta}_n \xrightarrow{\text{q.c.}} \theta_0$ *e com probabilidade 1,* $\overline{\mathcal{L}}_\theta(\widetilde{\theta}_n, \widehat{\pi}_n) = 0$*, para* n *suficientemente grande. Então se* $\mathcal{F}$ *satisfaz as condições de regularidade (C1) - (C5), temos que:*

$$\sqrt{n}(\widetilde{\theta}_n - \theta_0, \widehat{\pi}_n - \pi_0) \overset{\mathcal{D}}{\mapsto} \mathbf{V} \sim N_2(\mathbf{0}, \Sigma^*) \ ,$$

onde $\Sigma^* = \begin{pmatrix} I_{\theta_0}^{-1} + I_{\theta_0}^{-2} I_{\theta_0,\pi_0}^2 \sigma_{22} & -I_{\theta_0}^{-1} I_{\theta_0,\pi_0}\sigma_{22} \\ -I_{\theta_0}^{-1} I_{\theta_0,\pi_0}\sigma_{22} & \sigma_{22} \end{pmatrix}$.

Prova: Da condição de regularidade (C2), da fórmula de Taylor com resto de Lagrange e das propriedades de $\widetilde{\theta}_n$, temos com probabilidade 1 que, para n suficientemente grande:

$$\begin{aligned} 0 &= \sqrt{n}\,\overline{\mathcal{L}}_\theta(\theta_0, \widehat{\pi}_n) + \sqrt{n}(\widetilde{\theta}_n - \theta_0)\overline{\mathcal{L}}_{\theta\theta}(\theta_0, \widehat{\pi}_n) \\ &\quad + \frac{1}{2}\sqrt{n}(\widetilde{\theta}_n - \theta_0)^2 \overline{\mathcal{L}}_{\theta\theta\theta}(\theta_n^*, \widehat{\pi}_n) \ , \end{aligned}$$

e, assim,

$$\sqrt{n}(\widetilde{\theta}_n - \theta_0) X_n = -\sqrt{n}\,\overline{\mathcal{L}}_\theta(\theta_0, \widehat{\pi}_n) \ ,$$

onde θ_n^* está entre θ_0 e $\widetilde{\theta}_n$ e

$$X_n = \overline{\mathcal{L}}_{\theta\theta}(\theta_0, \widehat{\pi}_n) + \frac{1}{2}(\widetilde{\theta}_n - \theta_0)\overline{\mathcal{L}}_{\theta\theta\theta}(\theta_n^*, \widehat{\pi}_n) \ .$$

Daqui temos, pelos Lemas 2.1.3 (ii) e 2.2.1 (i), que

$$X_n \xrightarrow{\text{q.c.}} -I_{\theta_0} \ , \text{ com } I_{\theta_0} \in \mathbb{R}_+. \tag{2.20}$$

E, além disso, pelo Lema 2.2.4, segue que existe uma seqüência $(Y_n)_{n\geq 1}$ tal que:

$$Y_n \xrightarrow{\text{q.c.}} -I_{\theta_0,\pi_0} \ . \tag{2.21}$$

e com probabilidade 1, para n suficientemente grande

$$\sqrt{n}(\widetilde{\theta}_n - \theta_0)X_n = -[\sqrt{n}\,\overline{\mathcal{L}}_\theta(\theta_0, \pi_0) + \sqrt{n}(\widehat{\pi}_n - \pi_0)Y_n] ,$$

portanto, com probabilidade 1:

$$\sqrt{n}(\widetilde{\theta}_n - \theta_0) - \sqrt{n}Z_n \underset{n\uparrow\infty}{\longrightarrow} 0 , \tag{2.22}$$

onde $Z_n = -X_n^{-1}[\overline{\mathcal{L}}_\theta(\theta_0, \pi_0) + (\widehat{\pi}_n - \pi_0)Y_n]$.

De (2.22) e do teorema de Slutsky segue que $\sqrt{n}(\widetilde{\theta}_n - \theta_0, \widehat{\pi}_n - \pi_0)$ e $\sqrt{n}(Z_n, \widehat{\pi}_n - \pi_0)$ são assintoticamente equivalentes. Portanto, basta determinar a distribuição limite desta última seqüência. Para isto usamos o teorema de Cramér Wold (James, 1981): $\forall \mathbf{t} = (t_1, t_2) \in \mathbb{R}^2$, segue, usando as últimas expressões, o teorema de Slutsky e o Lema 2.2.3, que:

$$\begin{aligned}
&\sqrt{n}(Z_n, \widehat{\pi}_n - \pi_0)(t_1, t_2)^t \\
&= \sqrt{n}(\overline{\mathcal{L}}_\theta(\theta_0, \pi_0), \widehat{\pi}_n - \pi_0)(-t_1 X_n^{-1}, -t_1 X_n^{-1} Y_n + t_2)^t \\
&\overset{\mathcal{D}}{\longmapsto} \mathbf{W}(t_1 I_{\theta_0}^{-1}, -t_1 I_{\theta_0}^{-1} I_{\theta_0,\pi_0} + t_2)^t \\
&\sim N(0, \mathbf{t}\Sigma^* \mathbf{t}^t) .
\end{aligned}$$

Portanto, $\sqrt{n}(Z_n, \widehat{\pi}_n - \pi_0) \overset{\mathcal{D}}{\longmapsto} N(\mathbf{0}, \Sigma^*)$.

Assim, o teorema está provado. □

Corolário 2.2.6 *Consideremos as hipóteses do Teorema 2.2.5. Vamos assumir também que para cada $n \in N_+$, a equação $\overline{\mathcal{L}}_\theta(\theta, \widehat{\pi}_n) = 0$, vista como função de θ, tem uma única raiz. Então, se $\widehat{\theta}_n$ é o estimador de pseudo-máxima verossimilhança, temos que:*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0, \widehat{\pi}_n - \pi_0) \overset{\mathcal{D}}{\longmapsto} \mathbf{V} = (V_1, V_2) \sim N_2(\mathbf{0}, \Sigma^*) , \tag{2.23}$$

onde Σ^* é como no Teorema 2.2.5. Em particular,

$$\sqrt{n}(\widetilde{\theta}_n - \theta_0) \overset{\mathcal{D}}{\longmapsto} V_1 \sim N(0, \sigma^2) , \tag{2.24}$$

onde $\sigma^2 = I_{\theta_0}^{-1} + I_{\theta_0}^{-2} I_{\theta_0,\pi_0}^2 \sigma_{22}$.

Prova: Decorre do Corolário 2.1.7 e do Teorema 2.2.5. □

Observação 2.2.7 Consideremos o caso em que $\Pi = \{\pi_0\}$, com π_0 um valor conhecido, e $\widehat{\pi}_n = \pi_0$. Assim, o problema reduz-se à estimação do parâmetro θ. Neste caso, a estimação por pseudo máxima verossimilhança é equivalente à estimação por máxima verossimilhança usual, isto é, o estimador $\widehat{\theta}_n$ associado a $\widehat{\pi}_n$, obtido pelo primeiro método, é também o estimador de máxima verossimilhança usual do parâmetro θ. As condições de reuglaridade (C1) - (C5) são, neste caso, equivalentes às condições usuais dadas, por exemplo, em Serfling (1980) e as propriedades assintóticas do estimador de máxima verossimilhança ali demonstradas são conseqüências dos Teoremas 2.1.5 e 2.2.5.

Capítulo 3

Caso multiparamétrico

Neste capítulo generalizamos a teoria desenvolvida no Capítulo 2, considerando que temos um ou mais parâmetros de interesse e um ou mais de incômodo ("nuisance").

Primeiramente, apresentamos algumas notações que serão usadas ao longo deste capítulo.

Agora, $\mathcal{F} = \{F_{\boldsymbol{\theta},\boldsymbol{\pi}}\}_{(\boldsymbol{\theta},\boldsymbol{\pi})\in\Theta\times\Pi}$ denota uma família de distribuições $(p+q)$-paramétricas definidas sobre o conjunto dos números reais, $\mathbb{R}$, sendo Θ um subconjunto de $\mathbb{R}^p$ e Π um de $\mathbb{R}^q$, onde p e q estão em $\mathbb{N}_+$. Para cada par (θ, π), $f(\cdot|\boldsymbol{\theta}, \boldsymbol{\pi})$ denotará a função de densidade (ou de probabilidade) associada a $F_{\boldsymbol{\theta},\boldsymbol{\pi}}$. $\boldsymbol{\theta}$ será o vetor de parâmetros estruturais, ou de interesse, e $\boldsymbol{\pi}$ o de parâmetros "nuisance", ou de incômodo.

Dada $(X_1, \ldots, X_n)$, uma amostra aleatória da distribuição $F_{\boldsymbol{\theta},\boldsymbol{\pi}}$, com $(\boldsymbol{\theta}, \boldsymbol{\pi}) \in \Theta \times \Pi$, sejam

$$
\begin{aligned}
\mathcal{L}_n(\boldsymbol{\theta}, \boldsymbol{\pi}) &= \mathcal{L}_n(X_1, \ldots, X_n; \boldsymbol{\theta}, \boldsymbol{\pi}) = \ln \prod_{j=1}^{n} f(X_j|\boldsymbol{\theta}, \boldsymbol{\pi}) \\
&= \sum_{j=1}^{n} \ln f(X_j|\boldsymbol{\theta}, \boldsymbol{\pi}) \ ,
\end{aligned}
$$

a função log-verossimilhança correspondente a amostra observada e

$$
\overline{\mathcal{L}}_n(\theta, \pi) = \frac{1}{n}\mathcal{L}_n(\theta, \pi) \ .
$$

Definição 3.0.8 *Seja $(X_1, \ldots, X_n)$ uma amostra aleatória da distribuição $F_{\boldsymbol{\theta},\boldsymbol{\pi}}$ e $\widehat{\boldsymbol{\pi}}_n =$*

$\widehat{\boldsymbol{\pi}}_n(X_1, \ldots, X_n)$ *um estimador do parâmetro* $\boldsymbol{\pi}$. *Se:*

$$\widehat{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n(\widehat{\boldsymbol{\pi}}_n) = \widehat{\boldsymbol{\theta}}_n(X_1, \ldots, X_n) = \max_{\boldsymbol{\theta} \in \Theta} \overline{\mathcal{L}}_n(X_1, \ldots, X_n; \boldsymbol{\theta}, \widehat{\boldsymbol{\pi}}_n) \; ;$$

dizemos que $\widehat{\boldsymbol{\theta}}_n$ *é um estimador de pseudo máxima verossimilhança do parâmetro* $\boldsymbol{\theta}$.

Agora, as condições de regularidade para a família $\mathcal{F}$ tem a seguinte forma:

(C1) Para cada par $(\boldsymbol{\theta}, \boldsymbol{\pi}) \in \Theta \times \Pi$, $f(\cdot|\boldsymbol{\theta}, \boldsymbol{\pi})$ tem como suporte o espaço amostral $\mathcal{X} \subseteq \mathbb{R}$.

(C2) Para cada $x \in \mathcal{X}$, temos que para todo $(\boldsymbol{\theta}, \boldsymbol{\pi}) \in \Theta \times \Pi$, estão definidas as seguintes derivadas parciais:

$\dfrac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} \ln f(x|\boldsymbol{\theta}, \boldsymbol{\pi})$, para $i, j, k = 1, \ldots, p$;

$\dfrac{\partial}{\partial\pi_i} \ln f(x|\boldsymbol{\theta}, \boldsymbol{\pi})$, para $i = 1, \ldots, q$;

$\dfrac{\partial^3}{\partial\pi_i\partial\theta_j\partial\theta_k} \ln f(x|\boldsymbol{\theta}, \boldsymbol{\pi})$ e $\dfrac{\partial^3}{\partial\pi_i\partial\pi_l\partial\theta_k} \ln f(x|\boldsymbol{\theta}, \boldsymbol{\pi})$, para $i, l = 1, \ldots, q$, $j, k = 1, \ldots, p$.

(C3) Para cada par $(\boldsymbol{\theta}_1, \boldsymbol{\pi}_1) \in \Theta \times \Pi$, existem retângulos abertos $A_{\boldsymbol{\theta}_1} \subseteq \Theta$ e $B_{\boldsymbol{\pi}_1} \subseteq \Pi$ contendo $\boldsymbol{\theta}_1$ e $\boldsymbol{\pi}_1$, respectivamente, e funções g_1, g_2 e g_3, definidas sobre $\mathcal{X}$, tais que $\forall i \in \{1, 2, 3\}$, $0 < \int_{\mathcal{X}} g_i(x)\mathrm{d}x < \infty$ e para todo $(\boldsymbol{\theta}, \boldsymbol{\pi}) \in \Theta \times \Pi$ e todo $x \in \mathcal{X}$:

(i) $\left|\dfrac{\partial}{\partial\theta_i} f(x|\boldsymbol{\theta}, \boldsymbol{\pi})\right| \leq g_1(x)$, para $i = 1, \ldots, p$;

(ii) $\left|\dfrac{\partial^2}{\partial\theta_i\partial\theta_j} f(x|\boldsymbol{\theta}, \boldsymbol{\pi})\right| \leq g_2(x)$, para $i, j = 1, \ldots, p$;

(iii) $\left|\dfrac{\partial^2}{\partial\pi_i\partial\theta_j} f(x|\boldsymbol{\theta}, \boldsymbol{\pi})\right| \leq g_3(x)$, para $i = 1, \ldots, q$, $j = 1, \ldots, p$.

(C4) Para todo par $(\boldsymbol{\theta}, \boldsymbol{\pi}) \in \Theta \times \Pi$:

(*i*)
$$I_{\boldsymbol{\theta}} = E_{\boldsymbol{\theta},\boldsymbol{\pi}} \left\{ \left[\frac{\partial}{\partial\boldsymbol{\theta}} \ln f(X|\boldsymbol{\theta}, \boldsymbol{\pi})\right]^t \left[\frac{\partial}{\partial\boldsymbol{\theta}} \ln f(X|\boldsymbol{\theta}, \boldsymbol{\pi})\right] \right\}$$

é uma matriz definida positiva.

(*ii*)
$$I_{\boldsymbol{\theta},\boldsymbol{\pi}} = E_{\boldsymbol{\theta},\boldsymbol{\pi}} \left\{ \left[\frac{\partial}{\partial\boldsymbol{\theta}} \ln f(X|\boldsymbol{\theta}, \boldsymbol{\pi})\right]^t \left[\frac{\partial}{\partial\boldsymbol{\pi}} \ln f(X|\boldsymbol{\theta}, \boldsymbol{\pi})\right] \right\}$$

é uma matriz de componentes reais; onde $X \sim f(\cdot|\boldsymbol{\theta}, \boldsymbol{\pi})$,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(X|\boldsymbol{\theta}, \boldsymbol{\pi}) = \left[\frac{\partial}{\partial \theta_i} \ln f(X|\boldsymbol{\theta}, \boldsymbol{\pi}) \right]_{1\times p} \quad \text{e}$$

$$\frac{\partial}{\partial \boldsymbol{\pi}} \ln f(X|\boldsymbol{\theta}, \boldsymbol{\pi}) = \left[\frac{\partial}{\partial \pi_i} \ln f(X|\boldsymbol{\theta}, \boldsymbol{\pi}) \right]_{1\times q}.$$

(C5) Para cada $(\boldsymbol{\theta}_1, \boldsymbol{\pi}_1) \in \Theta \times \Pi$, existem funções M_1, M_2 e M_3, definidas sobre $\mathcal{X}$, tais que $\forall i \in \{1, 2, 3\}$, $0 < E_{\boldsymbol{\theta}_1, \boldsymbol{\pi}_1}[M_i(X)] < \infty$ e para todo $x \in \mathcal{X}$:

(i) $\forall(\boldsymbol{\theta}, \boldsymbol{\pi}) \in A_{\boldsymbol{\theta}_1} \times B_{\boldsymbol{\pi}_1}$, $\left| \frac{\partial^3}{\partial\theta_i \partial\theta_j \partial\theta_k} \ln f(x|\boldsymbol{\theta}, \boldsymbol{\pi}) \right| \leq M_1(x)$, para $i, j, k = 1, \ldots, p$;

(ii) $\forall \boldsymbol{\pi} \in B_{\boldsymbol{\pi}_1}$, $\left| \frac{\partial^3}{\partial\pi_i \partial\theta_j \partial\theta_k} \ln f(x|\boldsymbol{\theta}, \boldsymbol{\pi}) \right| \leq M_2(x)$, para $i = 1, \ldots, q$, $j, k = 1, \ldots, p$;

(iii) $\forall \boldsymbol{\pi} \in B_{\boldsymbol{\pi}_1}$, $\left| \frac{\partial^3}{\partial\pi_i \partial\pi_j \partial\theta_k} \ln f(x|\boldsymbol{\theta}, \boldsymbol{\pi}) \right| \leq M_3(x)$, para $i, j = 1, \ldots, q$, $k = 1, \ldots, p$;

onde $A_{\boldsymbol{\theta}_1}$ e $B_{\boldsymbol{\pi}_1}$ são como em (C3) e $X \sim f(\cdot|\boldsymbol{\theta}_1, \boldsymbol{\pi}_1)$.

Observação 3.0.9 Estas condições são parte daquelas usadas para provar as propriedades assintóticas do estimador de máxima verossimilhança usual, quando o parâmetro a ser estimado é um vetor de $(p+q)$ componentes reais; em particular as matrizes $I_{\boldsymbol{\theta}}$ e $I_{\boldsymbol{\theta}, \boldsymbol{\pi}}$, da condição (C4), são submatrizes da matriz de informação de Fisher.

Por último, se $(X_1, \ldots, X_n)$ é uma amostra aleatória da distribuição $F_{\boldsymbol{\theta}_1, \boldsymbol{\pi}_1} \in \mathcal{F}$, satisfazendo as condições de regularidade, então para cada $(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) \in \Theta \times \Pi$, seja

$$\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) = \frac{\partial}{\partial \boldsymbol{\theta}} \overline{\mathcal{L}}_n(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) = [a_i]_{1\times p},$$

onde $a_i = \overline{\mathcal{L}}_{\theta_i}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) = \frac{\partial}{\partial \theta_i} \overline{\mathcal{L}}_n(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$, para $i = 1, \ldots, p$.

Sejam, também,

$$\overline{\mathcal{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) = \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \overline{\mathcal{L}}_n(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) = [k_{ij}]_{p\times p},$$

onde $k_{ij} = \overline{\mathcal{L}}_{\theta_i \theta_j}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) = \frac{\partial^2}{\partial \theta_j \partial \theta_i} \overline{\mathcal{L}}_n(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$, para $i, j = 1, \ldots, p$

$$\text{e} \quad \overline{\mathcal{L}}_{\boldsymbol{\theta}, \boldsymbol{\pi}}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) = \frac{\partial^2}{\partial \boldsymbol{\pi} \partial \boldsymbol{\theta}} \overline{\mathcal{L}}_n(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) = [c_{ij}]_{p\times q},$$

onde $c_{ij} = \overline{\mathcal{L}}_{\theta_i \pi_j}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) = \frac{\partial^2}{\partial \pi_j \partial \theta_i} \overline{\mathcal{L}}_n(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$, para $i = 1, \ldots, p$, $j = 1, \ldots, q$.

3.1 Consistência forte

Nesta seção provamos a consistência forte do estimador de pseudo máxima verossimilhança $\widehat{\boldsymbol{\theta}}_n(\widehat{\boldsymbol{\pi}}_n)$.

Como no caso uniparamétrico, resultados preliminares são apresentados nos lemas seguintes.

Lema 3.1.1 *Seja* $(X_1, \ldots, X_n)$ *uma amostra aleatória da distribuição* $F_{\boldsymbol{\theta}_0,\boldsymbol{\pi}_0} \in \mathcal{F}$. *Sob as condições de regularidade (C1) - (C5), temos;*

(i) $\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) = \dfrac{1}{n}\sum_{j=1}^{n} \dfrac{\partial}{\partial \boldsymbol{\theta}} \ln f(X_j|\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) \xrightarrow{\text{q.c.}} \mathbf{0}$;

(ii) $\overline{\mathcal{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) = \dfrac{1}{n}\sum_{j=1}^{n} \dfrac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln f(X_j|\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) \xrightarrow{\text{q.c.}} -I_{\boldsymbol{\theta}_0}$;

(iii) $\overline{\mathcal{L}}_{\boldsymbol{\theta}\boldsymbol{\pi}}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) = \dfrac{1}{n}\sum_{j=1}^{n} \dfrac{\partial^2}{\partial \boldsymbol{\pi}\partial \boldsymbol{\theta}} \ln f(X_j|\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) \xrightarrow{\text{q.c.}} -I_{\boldsymbol{\theta}_0\boldsymbol{\pi}_0}$.

Prova: Das condições de regularidade (C3) e (C4) decorrem:

$$E_{\boldsymbol{\theta}_0,\boldsymbol{\pi}_0}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(X_1|\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)\right] = \mathbf{0}\ ; \tag{3.1}$$

$$\begin{aligned}
&E_{\boldsymbol{\theta}_0,\boldsymbol{\pi}_0}\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln f(X_1|\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)\right] \\
&= -E_{\boldsymbol{\theta}_0,\boldsymbol{\pi}_0}\left\{\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(X_1|\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)\right]^t \left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(X_1|\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)\right]\right\} \\
&= -I_{\boldsymbol{\theta}_0}\ ;
\end{aligned} \tag{3.2}$$

$$\begin{aligned}
&E_{\boldsymbol{\theta}_0,\boldsymbol{\pi}_0}\left[\frac{\partial^2}{\partial \boldsymbol{\pi}\partial \boldsymbol{\theta}} \ln f(X_1|\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)\right] \\
&= -E_{\boldsymbol{\theta}_0,\boldsymbol{\pi}_0}\left\{\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(X_1|\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)\right]^t \left[\frac{\partial}{\partial \boldsymbol{\pi}} \ln f(X_1|\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)\right]\right\} \\
&= -I_{\boldsymbol{\theta}_0,\boldsymbol{\pi}_0}\ ;
\end{aligned} \tag{3.3}$$

onde $I_{\boldsymbol{\theta}_0}$ e $I_{\boldsymbol{\theta}_0,\boldsymbol{\pi}_0}$ são matrizes de componentes reais.

Logo, como $X_1, \ldots, X_n$ são variáveis aleatórias independentes e identicamente distribuídas, então, pela lei forte dos grandes números de Kolmogorov e as expressões acima, prova-se (i), (ii) e (iii). □

Lema 3.1.2 *Seja $(X_1, \ldots, X_n)$ uma amostra aleatória da distribuição $F_{\boldsymbol{\theta}_0, \boldsymbol{\pi}_0} \in \mathcal{F}$, $\widetilde{\boldsymbol{\theta}}_n = \widetilde{\boldsymbol{\theta}}_n(X_1, \ldots, X_n)$ e $\widehat{\boldsymbol{\pi}}_n = \widehat{\boldsymbol{\pi}}_n(X_1, \ldots, X_n)$ dois estimadores fortemente consistentes para $\boldsymbol{\theta}_0$ e $\boldsymbol{\pi}_0$, respectivamente, então, para todo $\delta > 0$ e sob as condições de regularidade (C1), (C2) e (C5), temos com probabilidade 1 que, para n suficientemente grande:*

(i) $|\overline{\mathcal{L}}_{\theta_i\theta_j\theta_k}(\widetilde{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\pi}}_n)| < EM_1(X_1) + \delta$, *para* $i, j, k = 1, \ldots, p$;

(ii) $|\overline{\mathcal{L}}_{\theta_i\theta_j\pi_k}(\widetilde{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\pi}}_n)| < EM_2(X_1) + \delta$, *para* $i, j = 1, \ldots, p$, $k = 1, \ldots, q$;

(iii) $|\overline{\mathcal{L}}_{\theta_i\pi_j\pi_k}(\widetilde{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\pi}}_n)| < EM_3(X_1) + \delta$, *para* $i = 1, \ldots, p$, $j, k = 1, \ldots, q$,

onde M_1, M_2 e M_3 são como na condição de regularidade (C5).

Prova: É análoga à do Lema 2.1.2. □

Observação 3.1.3 O lema ainda vale se $(\widetilde{\theta}_n)_{n\geq 1}$ e $(\widehat{\pi}_n)_{n\geq 1}$ são seqüências tais que, com probabilidade 1, $(\widetilde{\theta}_n, \widehat{\boldsymbol{\pi}}_n) \in A_{\boldsymbol{\theta}_0} \times B_{\boldsymbol{\pi}_0}$ para n suficientemente grande, onde $A_{\boldsymbol{\theta}_0}$ e $B_{\boldsymbol{\pi}_0}$ são como na condição de regularidade (C5).

Lema 3.1.4 *Sejam $(X_1, \ldots, X_n)$ uma amostra aleatória da distribuição $F_{\boldsymbol{\theta}_0, \boldsymbol{\pi}_0} \in \mathcal{F}$ e $\widehat{\boldsymbol{\pi}}_n = \widehat{\boldsymbol{\pi}}_n(X_1, \ldots, X_n)$ um estimador fortemente consistente de $\boldsymbol{\pi}_0$. Então, sob as condições de regularidade (C1) - (C5), temos:*

(i) $\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\theta_0, \widehat{\pi}_n) \xrightarrow{\text{q.c.}} \mathbf{0}$;

(ii) $\overline{\mathcal{L}}_{\boldsymbol{\theta\theta}}(\theta_0, \widehat{\pi}_n) \xrightarrow{\text{q.c.}} -I_{\boldsymbol{\theta}_0}$;

(iii) $\overline{\mathcal{L}}_{\boldsymbol{\theta\pi}}(\theta_0, \widehat{\pi}_n) \xrightarrow{\text{q.c.}} -I_{\boldsymbol{\theta}_0, \boldsymbol{\pi}_0}$.

Prova: Usando os mesmos argumentos da prova do Lema 2.1.3, podemos provar a convergência das componentes correspondentes, isto é, para $i, j = 1, \ldots, p$ e $k = 1, \ldots, q$:

$$\overline{\mathcal{L}}_{\theta_i}(\theta_0, \widehat{\pi}_n) \xrightarrow{\text{q.c.}} 0 \; ;$$
$$\overline{\mathcal{L}}_{\theta_i\theta_j}(\theta_0, \widehat{\pi}_n) \xrightarrow{\text{q.c.}} -E\left[\frac{\partial}{\partial\theta_i}\ln f(X_1|\theta_0,\pi_0)\frac{\partial}{\partial\theta_j}\ln f(X_1|\theta_0,\pi_0)\right] \; ;$$
$$\overline{\mathcal{L}}_{\theta_i\pi_k}(\theta_0, \widehat{\pi}_n) \xrightarrow{\text{q.c.}} -E\left[\frac{\partial}{\partial\theta_i}\ln f(X_1|\theta_0,\pi_0)\frac{\partial}{\partial\pi_k}\ln f(X_1|\theta_0,\pi_0)\right] \; .$$

Isso prova o lema. □

Lema 3.1.5 *Seja $A \subseteq \mathbb{R}^p$ um retângulo aberto. Se $g : A \to \mathbb{R}$ possui derivadas parciais de terceira ordem, g_{ijk}, em todos os pontos de A, então, dados $\mathbf{x} = (x_1, \ldots, x_p)$ e $\mathbf{x}_0 = (x_{01}, \ldots, x_{0p})$ em A:*

$$\begin{aligned} g(\mathbf{x}) &= g(\mathbf{x}_0) + \sum_{i=1}^{p}(x_i - x_{0i})g_i(\mathbf{x}_0) + \frac{1}{2}\sum_{i=1}^{p}(x_i - x_{0i})^2 g_{ii}(\mathbf{x}_0) \\ &+ \sum_{i=1}^{p-1}\sum_{j>i}^{p}(x_i - x_{0i})(x_j - x_{0j})g_{ij}(\mathbf{x}_0) + \frac{1}{6}\sum_{i=1}^{p}(x_i - x_{0i})^3 g_{iii}(\mathbf{x}^*_{iii}) \\ &+ \frac{1}{2}\sum_{i=1}^{p-1}\sum_{j>i}^{p}(x_i - x_{0i})^2(x_j - x_{0j})g_{iij}(\mathbf{x}^*_{iij}) \\ &+ \frac{1}{2}\sum_{i=1}^{p-1}\sum_{j>i}^{p}(x_i - x_{0j})(x_j - x_{0j})^2 g_{ijj}(\mathbf{x}^*_{ijj}) \\ &+ \sum_{i=1}^{p-2}\sum_{j>i}^{p-1}\sum_{k>j}^{p}(x_i - x_{0i})(x_j - x_{0j})(x_k - x_{0k})g_{ijk}(\mathbf{x}^*_{ijk}) \, , \end{aligned} \tag{3.4}$$

onde, para $i, j, k = 1, \ldots, p$, $\mathbf{x}^*_{ijk}$ pertence a algum segmento de extremos $\mathbf{z} = (z_1, \ldots, z_p)$ e $\mathbf{w} = (w_1, \ldots, w_p)$, com z_i, $w_i \in \{x_{0i}, x_i\}$, $i = 1, \ldots, p$.

Se as derivadas parciais de segunda ordem são contínuas e, em particular, se as de terceira ordem são limitadas, podemos escrever (3.4) como:

$$\begin{aligned} g(\mathbf{x}) &= g(\mathbf{x}_0) + g'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)^t + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)g''(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)^t \\ &+ \frac{1}{6}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} c_{ijk}(x_i - x_{0i})(x_j - x_{0j})(x_k - x_{0k})g_{ijk}(\mathbf{x}^*_{ijk}) \, , \end{aligned} \tag{3.5}$$

onde $g'(\mathbf{x}_0)$ *e* $g''(\mathbf{x}_0)$ *são as matrizes de derivadas parciais de primeira e segunda ordem, respectivamente, avaliadas em* $\mathbf{x}_0$*, isto é,* $g'(\mathbf{x}_0) = (g_i(\mathbf{x}_0))_{1\times p}$ *e* $g''(\mathbf{x}_0) = (g_{ij}(\mathbf{x}_0))_{p\times p}$*; e* $0 \leq c_{ijk} \leq 1$*, para* $i, j, k = 1, \ldots, p$.

Prova: Por indução. Pela fórmula de Taylor com resto de Lagrange, (3.4) vale para $p = 1$.

Suponhamos que (3.4) vale para p. Provemos para $p+1$. Para isto, consideremos $\mathbf{x}_0 = (x_{01}, \ldots, x_{0p+1})$ e $\mathbf{x} = (x_1, \ldots, x_{p+1})$ em A. Seja $h : A \to \mathbb{R}$, definida por $h(\mathbf{y}) = g(\mathbf{y}, x_{p+1})$; então da hipótese de indução aplicada a h, decorre que:

$$\begin{aligned} g(\mathbf{x}, x_{p+1}) &= g(\mathbf{x}_0, x_{p+1}) + \sum_{i=1}^{p}(x_i - x_{0i})g_i(\mathbf{x}_0, x_{p+1}) \\ &+ \frac{1}{2}\sum_{i=1}^{p}(x_i - x_{0i})^2 g_{ii}(\mathbf{x}_0, x_{p+1}) \\ &+ \sum_{i=1}^{p-1}\sum_{j>i}^{p}(x_i - x_{0i})(x_j - x_{0j})g_{ij}(\mathbf{x}_0, x_{p+1}) \\ &+ \frac{1}{6}\sum_{i=1}^{p}(x_i - x_{0i})^3 g_{iii}(\mathbf{x}^*_{iii}, x_{p+1}) \\ &+ \frac{1}{2}\sum_{i=1}^{p-1}\sum_{j>i}^{p}(x_i - x_{0i})^2(x_j - x_{0j})g_{iij}(\mathbf{x}^*_{iij}, x_{p+1}) \\ &+ \frac{1}{2}\sum_{i=1}^{p-1}\sum_{j>i}^{p}(x_i - x_{0i})(x_j - x_{0j})^2 g_{ijj}(\mathbf{x}^*_{ijj}, x_{p+1}) \\ &+ \sum_{i=1}^{p-2}\sum_{j>i}^{p-1}\sum_{k>j}^{p}(x_i - x_{0i})(x_j - x_{0j})(x_k - x_{0k})g_{ijk}(\mathbf{x}^*_{ijk}, x_{p+1}) \;, \end{aligned} \tag{3.6}$$

onde, para $i, j, k = 1, \ldots, p$, $\mathbf{x}^*_{ijk}$ pertence a algum segmento de extremos $\mathbf{y} = (y_1, \ldots, y_p)$ e $\mathbf{w} = (w_1, \ldots, w_p)$, com z_i, $w_i \in \{x_{0i}, x_i\}$, para $i = 1, \ldots, p$.

Agora, usando a fórmula de Taylor com resto de Lagrange, as funções $g(\mathbf{x}_0, \cdot)$, $g_i(\mathbf{x}_0, \cdot)$ e $g_{ij}(\mathbf{x}_0, \cdot)$, para $i, j = 1, \ldots, p$, obtemos:

$$\begin{aligned} g(\mathbf{x}_0, x_{p+1}) &= g(\mathbf{x}_0) + (x_{p+1} - x_{0_{p+1}})g_{p+1}(\mathbf{x}_0) + \frac{1}{2}(x_{p+1} - x_{0_{p+1}})^2 g_{p+1,p+1}(\mathbf{x}_0) \\ &+ \frac{1}{6}(x_{p+1} - x_{0_{p+1}})^3 g_{p+1,p+1,p+1}(\mathbf{x}_0, x^*) \;, \end{aligned}$$

onde x^* está entre $x_{0_{p+1}}$ e x_{p+1};

$g_i(\mathrm{x}_0, x_{p+1}) = g(\mathrm{x}_0) + (x_{p+1} - x_{0_{p+1}})g_{i,p+1}(\mathrm{x}_0) + \frac{1}{2}(x_{p+1} - x_{0_{p+1}})^2 g_{i,p+1,p+1}(\mathrm{x}_0, x_i^*)$,

onde x_i^* está entre $x_{0,p+1}$ e x_{p+1};

$g_{ij}(\mathrm{x}_0, x_{p+1}) = g_{ij}(\mathrm{x}_0) + (x_{p+1} - x_{0p+1})g_{i,j,p+1}(\mathrm{x}_0, x_{ij}^*)$,

onde x_{ij}^* está entre $x_{0,p+1}$ e x_{p+1}.

Substituindo estas últimas expressões em (3.6), podemos verificar que (3.4) vale para $p+1$.

Por último, para verificar (3.5), basta observar que se as derivadas parciais de segunda ordem são contínuas, em particular se as de terceira ordem são limitadas, então $g_{ij}(\mathrm{x}_0) = g_{ji}(\mathrm{x}_0)$, para $i, j = 1, \ldots, p$. Assim,

$$
\begin{aligned}
&\sum_{i=1}^{p}(x_i - x_{0i})g_i(\mathrm{x}_0) + \frac{1}{2}\sum_{i=1}^{p}(x_i - x_{0i})^2 g_{ii}(\mathrm{x}_0) \\
&+ \sum_{i=1}^{p-1}\sum_{j>i}^{p}(x_i - x_{0i})(x_j - x_{0j})g_{ij}(\mathrm{x}_0) \\
&= g'(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0)^t + \frac{1}{2}\Bigg[\sum_{i=1}^{p}(x_i - x_{0i})^2 g_{ii}(\mathrm{x}_0) \\
&+ \sum_{i=1}^{p-1}\sum_{j>i}^{p}(x_i - x_{0i})(x_j - x_{0j})g_{ij}(\mathrm{x}_0) \\
&+ \sum_{i=1}^{p-1}\sum_{j>i}^{p}(x_i - x_{0i})(x_j - x_{0j})g_{ji}(\mathrm{x}_0)\Bigg] \\
&= g'(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0)^t + \frac{1}{2}(\mathrm{x} - \mathrm{x}_0)g''(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0)^t .
\end{aligned}
$$

Daí decorre (3.5). □

Lema 3.1.6 *Se $A_{p\times p}$ é uma matriz simétrica e definida positiva. Então $\forall \mathrm{x} \in \mathbb{R}^p$, temos:*

$$\lambda_p\|\mathrm{x}\|^2 \leq \mathrm{x}A\mathrm{x}^t \leq \lambda_1\|\mathrm{x}\|^2 ,$$

onde $\lambda_1 > \cdots > \lambda_p$ são os autovalores de A e $\|x\|^2 = \sum_{j=1}^{p} x_j^2$.

Prova: Como A é simétrica, existe uma matriz ortonormal $C_{p\times p}$ tal que $C^tAC = D$, onde $D =$ diagonal $(\lambda_1, \ldots, \lambda_p)$, sendo $\lambda_1, \ldots, \lambda_p$ os autovalores de A; e como A é definida positiva, $\lambda_1 > 0, \dots, \lambda_p > 0$. Sem perda de generalidade, podemos supor que $\lambda_1 > \cdots > \lambda_p$. Então, $\forall x \in \mathbb{R}^p : \mathbf{x}A\mathbf{x}^t = \mathbf{x}CDC^t\mathbf{x}^t = \mathbf{y}D\mathbf{y}^t = \sum_{j=1}^p \lambda_j y_j^2$, onde $\mathbf{y} = (y_1, \dots, y_p) = \mathbf{x}C$; e portanto $\lambda_p \|\mathbf{x}C\|^2 = \lambda_p \|\mathbf{y}\|^2 = \lambda_p \sum_{j=1}^p y_j^2 \le \sum_{j=1}^p \lambda_j y_j^2 = \mathbf{x}A\mathbf{x}^t \le \lambda_1 \sum_{j=1}^p y_j^2 = \lambda_1 \|\mathbf{y}\|^2 = \lambda_1 \|xC\|^2$. Daqui decorre o lema, pois sendo C ortonormal, $\|\mathbf{x}C\|^2 = \mathbf{x}C(\mathbf{x}C)^t = \|x\|^2$. □

Lema 3.1.7 *Sejam $(X_1, \dots, X_n)$ uma amostra aleatória de $F_{\boldsymbol{\theta}_0, \boldsymbol{\pi}_0} \in \mathcal{F}$, com $\mathcal{F}$ satisfazendo as condições de regularidade (C1) - (C5), e $\widehat{\boldsymbol{\pi}}_n = \widehat{\boldsymbol{\pi}}_n(X_1, \dots, X_n)$ um estimador fortemente consistente de $\boldsymbol{\pi}_0$. Temos, com probabilidade 1, que para cada $\varepsilon > 0$ suficientemente pequeno:*

$$\forall \boldsymbol{\theta} \in \partial B(\boldsymbol{\theta}_0, \varepsilon) : \overline{\mathcal{L}}_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\pi}}_n) < \overline{\mathcal{L}}_n(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}}_n) ,$$

para n suficientemente grande; onde

$$B(\boldsymbol{\theta}_0, \varepsilon) = \{\boldsymbol{\theta} \in \mathbb{R}^p : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \varepsilon\}$$

e

$$\partial B(\boldsymbol{\theta}_0, \varepsilon) = \{\boldsymbol{\theta} \in \mathbb{R}^p : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = \varepsilon\} .$$

Prova: Sejam $A_{\boldsymbol{\theta}_0}$ e $B_{\boldsymbol{\pi}_0}$ retângulos abertos contendo $\boldsymbol{\theta}_0$ e $\boldsymbol{\pi}_0$, respectivamente, satisfazendo a condição de regularidade (C5) - (i). Seja $\varepsilon_1 > 0$ tal que $\forall \varepsilon \in (0, \varepsilon_1)$, $\partial B(\boldsymbol{\theta}_0, \varepsilon) \subseteq A_{\boldsymbol{\theta}_0}$.

Como $\widehat{\boldsymbol{\pi}}_n \xrightarrow{\text{q.c.}} \boldsymbol{\pi}_0$, segue, com probabilidade 1, que $\widehat{\boldsymbol{\pi}}_n \in B_{\boldsymbol{\pi}_0}$ para n suficientemente grande. Assim, pela condição de regularidade (C5) - (i) e (3.5) do Lema 3.1.5, segue com probabilidade 1 que, para $\varepsilon \in (0, \varepsilon_1)$, $\boldsymbol{\theta} \in \partial B(\boldsymbol{\theta}_0, \varepsilon)$ e n suficientemente grande:

$$\overline{\mathcal{L}}_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\pi}}_n) - \overline{\mathcal{L}}_n(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}}_n) = S_{1n} + S_{2n} + S_{3n} , \tag{3.7}$$

onde:

$$S_{1n} = \overline{\mathcal{L}}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}}_n)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^t ;$$

$$\begin{aligned} S_{2n} &= \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\overline{\mathcal{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}}_n)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^t \ ; \\ S_{3n} &= \frac{1}{6}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} c_{ijk}(\theta_i-\theta_{0i})(\theta_j-\theta_{0j})(\theta_k-\theta_{0k})\overline{\mathcal{L}}_{\theta_i\theta_j\theta_k}(\boldsymbol{\theta}^*_{ijk},\widehat{\boldsymbol{\pi}}_n) \ ; \end{aligned}$$

com $0 \le c_{ijk} \le 1$ e $\boldsymbol{\theta}^*_{ijk} \in B(\boldsymbol{\theta}_0,\varepsilon)$, para $i,j,k = 1,\ldots,p$.

Provaremos abaixo que, com probabilidade 1, para $\varepsilon > 0$ suficientemente pequeno, tem-se para $\theta \in \partial B(\boldsymbol{\theta}_0,\varepsilon)$ e n suficientemente grande:

$$|S_{1n}| < p^3\varepsilon^3 \ ; \tag{3.8}$$

$$S_{2n} < -\frac{1}{4}\lambda_p\varepsilon^2 \ . \tag{3.9}$$

onde $\lambda_p > 0$, é o menor autovalor da matriz $I_{\boldsymbol{\theta}_0}$;

$$|S_{3n}| < p^3\varepsilon^3 EM_1(X_1) \ , \tag{3.10}$$

onde M_1 é como na condição de regularidade (C5).

Segue de (3.6) e das desigualdades anteriores que, se $0 < \varepsilon < \frac{1}{4}\lambda_p p^{-3}[1+EM_1(X_1)]^{-1}$ e $\varepsilon < \varepsilon_1$, então, com probabilidade 1, para $\boldsymbol{\theta} \in \partial B(\boldsymbol{\theta}_0,\varepsilon)$ e n suficientemente grande,

$$\overline{\mathcal{L}}_n(\boldsymbol{\theta},\widehat{\boldsymbol{\pi}}_n) - \overline{\mathcal{L}}_n(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}}_n) = S_{1n}+S_{2n}+S_{3n} < 0 \ ,$$

o que prova o lema.

Prova de (3.7): Pelo Lema 3.1.1 (i), $S_{1n} = \overline{\mathcal{L}}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}}_n)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^t \xrightarrow{\text{q.c.}} 0$, daí segue (3.7). □

Prova de (3.9): Pelo Lema 3.1.1 (ii), $S_{2n} = \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\overline{\mathcal{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}}_n)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^t \xrightarrow{\text{q.c.}} -\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)I_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^t$. Então, pela condição de regularidade (C4) - (i) e o Lema 3.1.4, segue para $\varepsilon > 0$ suficientemente pequeno e $\theta \in \partial B(\boldsymbol{\theta}_0,\varepsilon)$, que com probabilidade 1:

$$\begin{aligned} S_{2n} &< -\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)I_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^t + \frac{1}{4}\lambda_p\varepsilon^2 \\ &< -\frac{1}{2}\lambda_p\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^2 + \frac{1}{4}\lambda_p\varepsilon^2 = -\frac{1}{4}\lambda_p\varepsilon^2 \ . \end{aligned}$$

Isso prova (3.9). □

Prova de (3.10): Pela Observação 3.1.3 vale o Lema 3.1.2 (i), logo, para $0 < \varepsilon < \varepsilon_1$ e $\boldsymbol{\theta} \in \partial B(\boldsymbol{\theta}_0, \varepsilon)$, temos, com probabilidade 1, que para n suficientemente grande:

$$|S_{3n}| < \frac{1}{6} \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} \varepsilon^3 [EM_1(X_1) + 5EM_1(X_1)] = p^3 \varepsilon^3 EM_1(X_1) \ . \qquad \square$$

Isso prova o lema. $\square$

O resultado principal da seção é enunciado e provado a seguir.

Teorema 3.1.8 *Seja* $(X_1, \ldots, X_n)$ *uma amostra aleatória de* $F_{\boldsymbol{\theta}_0, \pi_0} \in \mathcal{F}$, *com* $\mathcal{F}$ *satisfazendo as condições de regularidade (C1) - (C5) e* $\widehat{\boldsymbol{\pi}}_n = \widehat{\boldsymbol{\pi}}_n(X_1, \ldots, X_n)$ *um estimador fortemente consistente de* $\boldsymbol{\pi}_0$. *Então, existe uma seqüência* $(\widetilde{\boldsymbol{\theta}}_n)_{n \geq 1}$ *tal que com probabilidade 1:*

(i) $\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\widetilde{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\pi}}_n) = \mathbf{0}$, *para* n *suficientemente grande;*

(ii) $\widetilde{\boldsymbol{\theta}}_n \underset{n \uparrow \infty}{\longrightarrow} \boldsymbol{\theta}_0$.

Prova: Da condição de regularidade (C2), segue, para $\varepsilon > 0$ e $n \in \mathbb{N}_+$, que existe $\widetilde{\boldsymbol{\theta}}_{n,\varepsilon} \in \overline{B}(\boldsymbol{\theta}_0, \varepsilon)$, satisfazendo

$$\overline{\mathcal{L}}_n(\widetilde{\boldsymbol{\theta}}_{n,\varepsilon}, \widehat{\boldsymbol{\pi}}_n) = \max\{\overline{\mathcal{L}}_n(\boldsymbol{\theta}, \widehat{\boldsymbol{\pi}}_n) : \boldsymbol{\theta} \in B(\boldsymbol{\theta}_0, \varepsilon)\} \ .$$

Daí segue, pelo Lema 3.1.7 e da condição de regularidade (C2), que para $\varepsilon > 0$, suficientemente pequeno, temos com probabilidade 1:

$$\widetilde{\boldsymbol{\theta}}_{n,\varepsilon} \in B(\boldsymbol{\theta}_0, \varepsilon) \quad \text{e} \quad \overline{\mathcal{L}}_{\boldsymbol{\theta}}(\widetilde{\boldsymbol{\theta}}_{n,\varepsilon}, \widehat{\boldsymbol{\pi}}_n) = \mathbf{0} \ , \tag{3.11}$$

para n suficientemente grande.

Seja $\varepsilon_0 > 0$ satisfazendo (3.11); e definamos para n suficientemente grande $\widetilde{\boldsymbol{\theta}}_n$, de modo que:

$$\|\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| = \inf\{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| : \boldsymbol{\theta} \in \overline{B}(\boldsymbol{\theta}_0, \varepsilon_0) \quad \text{e} \quad \overline{\mathcal{L}}_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \widehat{\boldsymbol{\pi}}_n) = \mathbf{0}\} \ . \tag{3.12}$$

Então, pela continuidade das funções norma e $\overline{\mathcal{L}}_{\boldsymbol{\theta}}$, temos com probabilidade 1 que:

$$\widetilde{\boldsymbol{\theta}}_n \in \overline{B}(\boldsymbol{\theta}_0, \varepsilon_0) \quad \text{e} \quad \overline{\mathcal{L}}_{\boldsymbol{\theta}}(\widetilde{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\pi}}_n) = \mathbf{0} \ . \tag{3.13}$$

Ou seja, $\widetilde{\boldsymbol{\theta}}_n$ satisfaz (i) da tese.

Por último, $\widetilde{\boldsymbol{\theta}}_n \xrightarrow[n\uparrow\infty]{} \boldsymbol{\theta}_0$ com probabilidade 1. Com efeito, seja $0 < \varepsilon < \varepsilon_0$. Segue de (3.11) que, com probabilidade 1 e para n suficientemente grande,

$$\widetilde{\boldsymbol{\theta}}_{n,\varepsilon} \in B(\boldsymbol{\theta}_0, \varepsilon) \subseteq B(\boldsymbol{\theta}_0, \varepsilon_0) \quad \text{e} \quad \overline{\mathcal{L}}_{\boldsymbol{\theta}}(\widetilde{\boldsymbol{\theta}}_{n,\varepsilon}, \widehat{\boldsymbol{\pi}}_n) = \mathbf{0} \ ;$$

e portanto por (3.12), $\|\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \leq \|\widetilde{\boldsymbol{\theta}}_{n,\varepsilon} - \boldsymbol{\theta}_0\| < \varepsilon$. Ou seja, $\widetilde{\boldsymbol{\theta}}_n \xrightarrow[n\uparrow\infty]{} \boldsymbol{\theta}_0$ com probabilidade 1. □

Corolário 3.1.9 *Consideremos as hipóteses do Teorema 3.1.8. Suponhamos também que para cada $n \in \mathbb{N}_+$, a equação $\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \widehat{\boldsymbol{\pi}}_n) = \mathbf{0}$, vista como função de $\boldsymbol{\theta}$, tem uma única raiz. Então, o estimador de pseudo máxima verossimilhança, $\widehat{\boldsymbol{\theta}}_n(\widehat{\boldsymbol{\pi}}_n)$, é fortemente consistente para $\boldsymbol{\theta}_0$.*

Prova: Segue da condição de regularidade (C2) e o Teorema 3.1.8. □

3.2 Distribuição assintótica

Nesta seção estudamos a distribuição assintótica do estimador de pseudo máxima verossimilhança no caso multiparamétrico. Como nas seções anteriores, resultados preliminares são apresentados.

Lema 3.2.1 *Seja $(X_1, \ldots, X_n)$ uma amostra aleatória da distribuição $F_{\boldsymbol{\theta}_0, \boldsymbol{\pi}_0} \in \mathcal{F}$, onde $\mathcal{F}$ satisfaz as condições de regularidade (C1), (C2) e (C5). Sejam $\widetilde{\boldsymbol{\theta}}_{in} = \widetilde{\boldsymbol{\theta}}_{in}(X_1, \ldots, X_n)$, $i = 1, 2$, dois estimadores fortemente consistentes de $\boldsymbol{\theta}_0$; $\widehat{\boldsymbol{\pi}}_{in} = \widehat{\boldsymbol{\pi}}_{in}(X_1, \ldots, X_n)$, $i = 1, 2$, dois estimadores fortemente consistentes de $\boldsymbol{\pi}_0$. Temos:*

(i) $(\widetilde{\boldsymbol{\theta}}_{1n} - \boldsymbol{\theta}_0)\overline{\mathcal{L}}_{\theta_i\theta_j\theta_k}(\widetilde{\boldsymbol{\theta}}_{2n}, \widehat{\boldsymbol{\pi}}_{1n}) \xrightarrow{\text{q.c.}} \mathbf{0}_{1\times p}$, *para $i, j, k = 1, \ldots, p$;*

(ii) $(\widetilde{\boldsymbol{\pi}}_{1n} - \boldsymbol{\pi}_0)\overline{\mathcal{L}}_{\theta_i\theta_j\theta_k}(\widetilde{\theta}_0, \widehat{\boldsymbol{\pi}}_{2n}) \xrightarrow{\text{q.c.}} \mathbf{0}_{1\times q}$, *para* $i = 1, \ldots, p$, $j, k = 1, \ldots, q$.

Prova: Segue do Lema 3.1.2. □

Lema 3.2.2 *Seja* $(X_1, \ldots, X_n)$ *uma amostra aleatória da distribuição* $F_{\boldsymbol{\theta}_0, \pi_0} \in \mathcal{F}$, *onde* $\mathcal{F}$ *satisfaz as condições de regularidade (C1) - (C5). Seja* $\widehat{\boldsymbol{\pi}}_n = \widehat{\boldsymbol{\pi}}_n(X_1, \ldots, X_n)$ *tal que*

$$\sqrt{n}(\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\theta_0, \boldsymbol{\pi}_0), \widehat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0) \overset{\mathcal{D}}{\mapsto} \mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2) \sim N_{p+q}(0, \Sigma) ,$$

onde $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^t & \Sigma_{22} \end{pmatrix}$.

Então, temos:

(i) $\Sigma_{11} = I_{\boldsymbol{\theta}_0}$;

(ii) $\Sigma_{12} = O_{p\times q}$.

Prova: Pelas condições de regularidade (C3) (i) e (C4) (i) segue que

$$\{\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(X_n|\theta_0, \boldsymbol{\pi}_0)\}_{n\geq 1}$$

é uma seqüência de vetores aleatórios independentes e identicamente distribuídos, com vetor de médias $E_{\boldsymbol{\theta}_0, \pi_0}[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(X_1|\theta_0, \boldsymbol{\pi}_0)] = \mathbf{0}$ e matriz de covariâncias

$$E_{\boldsymbol{\theta}_0, \pi_0}\{[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(X_1|\theta_0, \boldsymbol{\pi}_0)]^t[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(X_1|\theta_0, \boldsymbol{\pi}_0)]\} = I_{\boldsymbol{\theta}_0} ,$$

que é uma matriz de componentes reais. Portanto, pelo teorema central do limite para vetores aleatórios, $\sqrt{n}\,\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\theta_0, \pi_0) \overset{\mathcal{D}}{\mapsto} \mathbf{W}_1 \sim N_p(0, I_{\boldsymbol{\theta}_0})$. Por outro lado, segue da hipótese e do teorema de Cramér Wold que $\sqrt{n}\,\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\theta_0, \boldsymbol{\pi}_0) \overset{\mathcal{D}}{\mapsto} \mathbf{W}_1 \sim N_p(\mathbf{0}, \Sigma_{11})$. Logo, $\Sigma_{11} = I_{\boldsymbol{\theta}_0}$. Isso prova (i).

Para provar (ii), consideremos que

$$\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0p}) , \quad \boldsymbol{\pi}_0 = (\pi_{01}, \ldots, \pi_{0q}) , \quad \widehat{\boldsymbol{\pi}}_n = (\widehat{\pi}_{n0}, \ldots, \widehat{\pi}_{nq}) ,$$

$$I_{\boldsymbol{\theta}_0} = (a_{ij})_{p\times p} , \quad \Sigma_{22} = (b_{ij})_{q\times q} \quad \text{e} \quad \Sigma_{12} = (\sigma_{ij})_{p\times q} .$$

Sejam $j \in \{1, \ldots, p\}$ e $k \in \{1, \ldots, q\}$. Mostraremos que $\sigma_{jk} = 0$. Pelo Teorema 2.1.5, existe uma seqüência $(\widetilde{\theta}_{nj})_{n \geq 1}$ tal que com probabilidade 1 tenhamos $\widetilde{\theta}_{nj} \xrightarrow[n \uparrow \infty]{} \theta_{0j}$ e $\mathcal{L}_{\theta_j}(\theta_{01}, \ldots, \theta_{0j-1}, \widetilde{\theta}_{nj}, \theta_{0j+1}, \ldots, \theta_{0p}, \pi_{01}, \ldots, \pi_{0q}) = 0$, para n suficientemente grande. Então, procedendo como na prova do Teorema 2.1.5 para verificar (2.16), podemos mostrar que:

$$\sqrt{n}(\widetilde{\theta}_{nj} - \theta_0, \widehat{\pi}_{n_k} - \pi_0) \overset{\mathcal{D}}{\mapsto} U_{j_k} \sim N_2(\mathbf{0}, C_{j_k}) \ , \tag{3.14}$$

onde $C_{j_k} = \begin{pmatrix} a_{jj}^{-1} & a_{jj}^{-1}\sigma_{jk} \\ a_{jj}^{-1}\sigma_{ik} & \sigma_{jk} \end{pmatrix}$, com $a_{jj} > 0$, pois $I_{\boldsymbol{\theta}_0}$ é definida positiva.

De (3.14), podemos concluir que $\widetilde{\theta}_{nj}$ é assintoticamente normal e eficiente para o problema de estimar θ_j, sendo os restantes parâmetros conhecidos. Portanto, pelos argumentos da prova de (ii) do Teorema 2.1.5, (3.14) também implica que $\sigma_{jk} = 0$.

Assim, provamos que $\forall j \in \{1, \ldots, p\}$ e $\forall k \in \{1, \ldots, q\}$, $\sigma_{jk} = 0$. Portanto, $\Sigma_{12} = \mathbf{0}_{p \times q}$.

□

Lema 3.2.3 *Seja $A \subset \mathbb{R}^p$ um retângulo aberto. Se $g : A \to \mathbb{R}$ possui derivadas parciais de segunda ordem, g_{ij}, em todos os pontos de A, então, dados $\mathbf{x} = (x_1, \ldots, x_p)$ e $\mathbf{x}_0 = (x_{01}, \ldots, x_{op})$ em A:*

$$\begin{aligned} g(\mathbf{x}) &= g(\mathbf{x}_0) + \sum_{i=1}^{p}(x_i - x_{0i})g_i(\mathbf{x}_0) + \frac{1}{2}\sum_{i=1}^{p}(x_i - x_{0i})^2 g_{ii}(\mathbf{x}_{ii}^*) \\ &\quad + \sum_{i=1}^{p-1}\sum_{j>i}^{p}(x_i - x_{0i})(x_j - x_{0j})g_{ij}(\mathbf{x}_{ij}^*) \ , \end{aligned} \tag{3.15}$$

onde, para $i, j = 1, \ldots, p$, $\mathbf{x}_{ij}^*$ pertence a algum segmento de extremos $\mathbf{z} = (z_1, \ldots, z_p)$ e $\mathbf{w} = (w_1, \ldots, w_p)$, com $z_i, w_i \in \{x_{0i}, x_i\}$, $i = 1, \ldots, p$.

Em particular, podemos escrever (3.15) como:

$$\begin{aligned} g(\mathbf{x}) &= g(\mathbf{x}_0) + \sum_{i=1}^{p}(x_i - x_{0i})g_i(\mathbf{x}_0) \\ &\quad + \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p} c_{ij}(x_i - x_{0i})(x_j - x_{0j})g_{ij}(\mathbf{x}_{ij}^*) \ , \end{aligned} \tag{3.16}$$

onde $0 \leq c_{ij} \leq 1$, *para* $i, j = 1, \ldots, p$, *ou, equivalentemente,*

$$g(\mathrm{x}) = g(\mathrm{x}_0) + g'(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0)^t + \frac{1}{2}(\mathrm{x} - \mathrm{x}_0)C(\mathrm{x} - \mathrm{x}_0)^t \ , \tag{3.17}$$

onde $C = [c_{ij} g_{ij}(\mathrm{x}^*_{ij})]_{p \times p}$.

Prova: Análoga à do Lema 3.1.5. □

Observação 3.2.4 Claramente, os pontos x^*_{ij}, onde são avaliadas as derivadas de segunda ordem, também satisfazem, para $i, j = 1, \ldots, p$, $\mathrm{x}^*_{ij} \to \mathrm{x}_0$ quando $\mathrm{x} \to \mathrm{x}_0$.

Lema 3.2.5 *Seja* $(X_1, \ldots, X_n)$ *uma amostra aleatória da distribuição* $F_{\boldsymbol{\theta}_0, \boldsymbol{\pi}_0} \in \mathcal{F}$, *com* $\mathcal{F}$ *satisfazendo as condições de regularidade (C1) - (C5). Sejam* $(\widetilde{\theta}_n)_{n \geq 1}$ *e* $(\widehat{\pi}_n)_{n \geq 1}$, *tais que* $\widetilde{\boldsymbol{\theta}}_n \xrightarrow{\text{q.c.}} \boldsymbol{\theta}_0$ *e* $\widehat{\boldsymbol{\pi}}_n \xrightarrow{\text{q.c.}} \boldsymbol{\pi}_0$. *Então:*

(i) $\sqrt{n}\,\overline{\mathcal{L}}^t_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}}_n) - \sqrt{n}\,\overline{\mathcal{L}}^t_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) - \mathbf{Y}_n \sqrt{n}(\widehat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0)^t \xrightarrow{\text{q.c.}} \mathbf{0}^t_{1 \times p}$, *onde* $\mathbf{Y}_n \xrightarrow{\text{q.c.}} -I_{\boldsymbol{\theta}_0, \boldsymbol{\pi}_0}$;

(ii) $\sqrt{n}\,\overline{\mathcal{L}}^t_{\boldsymbol{\theta}}(\widetilde{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\pi}}_n) - \sqrt{n}\,\overline{\mathcal{L}}^t_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}}_n) - \mathbf{V}_n \sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^t \xrightarrow{\text{q.c.}} \mathbf{0}^t_{1 \times p}$, *onde* $\mathbf{V}_n \xrightarrow{\text{q.c.}} -I_{\boldsymbol{\theta}_0}$;

(iii) $-\sqrt{n}\,\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\widetilde{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\pi}}_n) + \sqrt{n}\,\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) + \sqrt{n}(\widehat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0)\mathbf{Y}^t_n + \sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\mathbf{V}_n \xrightarrow{\text{q.c.}} \mathbf{0}_{1 \times p}$.

Prova: Sejam $A_{\boldsymbol{\theta}_0}$ e $B_{\boldsymbol{\pi}_0}$ retângulos satisfazendo a condição de regularidade (C5). Como $(\widetilde{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\pi}}_n) \xrightarrow{\text{q.c.}} (\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)$ e $(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) \in A_{\boldsymbol{\theta}_0} \times B_{\boldsymbol{\pi}_0}$, segue, com probabilidade 1, que $(\widetilde{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\pi}}_n) \in A_{\boldsymbol{\theta}_0} \times B_{\boldsymbol{\pi}_0}$ para n suficientemente grande. Logo, por (3.17) do Lema 3.2.3, temos, com probabilidade 1 e para n suficientemente grande, que se $k \in \{1, \ldots, p\}$:

- $$\begin{aligned} \sqrt{n}\,\overline{\mathcal{L}}_{\theta_k}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}}_n) &= \sqrt{n}\,\overline{\mathcal{L}}_{\theta_k}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) + \sqrt{n}\,\overline{\mathcal{L}}_{\theta_k \boldsymbol{\pi}}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0)\,(\widehat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0)^t \\ &\quad + \sqrt{n}(\widehat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0)C_n(\widehat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0)^t \ , \end{aligned} \tag{3.18}$$

onde $C_n = [c_{nij} \overline{\mathcal{L}}_{\theta_k \pi_i \pi_j}(\boldsymbol{\theta}_0, \boldsymbol{\pi}^*_{nij})]_{q \times q}$, com $0 \leq c_{nij} \leq 1$ e $\boldsymbol{\pi}^*_{nij} \to \boldsymbol{\pi}_0$ quando $\widehat{\boldsymbol{\pi}}_n \to \boldsymbol{\pi}_0$, para $i, j = 1, \ldots, p$;

- $$\begin{aligned} \sqrt{n}\,\overline{\mathcal{L}}_{\theta_k}(\widetilde{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\pi}}_n) &= \sqrt{n}\,\overline{\mathcal{L}}_{\theta_k}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}}_n) + \sqrt{n}\,\overline{\mathcal{L}}_{\theta_k \boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_n)\,(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^t \\ &\quad + \sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)D_n(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^t \ , \end{aligned} \tag{3.19}$$

onde $D_n = [d_{ij}\overline{\mathcal{L}}_{\theta_k\boldsymbol{\theta}}(\boldsymbol{\theta}^*_{nij}, \widehat{\boldsymbol{\pi}}_n)]_{p\times p}$, com $0 \leq d_{ij} \leq 1$ e $\boldsymbol{\theta}^*_{nij} \longrightarrow \boldsymbol{\theta}_0$ quando $\widetilde{\boldsymbol{\theta}}_n \longrightarrow \boldsymbol{\theta}_0$, para $i, j = 1, \ldots, p$.

De (3.18), dos Lemas 3.1.1 (iii) e 3.2.1 (ii), segue que, para $k \in \{1, \ldots, p\}$:

$$\sqrt{n}\,\overline{\mathcal{L}}_{\theta_k}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}}_n) - \sqrt{n}\,\overline{\mathcal{L}}_{\theta_k}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0) - \mathbf{Y}_{nk}\sqrt{n}(\widehat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0)^t \xrightarrow{\text{q.c.}} 0 \ , \tag{3.20}$$

onde $\mathbf{Y}_{nk} \xrightarrow{\text{q.c.}} -I^{(k)}_{\boldsymbol{\theta}_0,\boldsymbol{\pi}_0}$, sendo $I^{(k)}_{\boldsymbol{\theta}_0,\boldsymbol{\pi}_0}$ a k-ésima fila da matriz $I_{\boldsymbol{\theta}_0,\boldsymbol{\pi}_0}$.

Sejam $\mathbf{Y}_n$ e $\mathbf{V}_n$ matrizes com k-ésima linha $\mathbf{Y}_{n_k}$ e $\mathbf{V}_{n_k}$, respectivamente. Enão, de (3.19), dos Lemas 3.1.4 (ii) e 3.2.1 (i), segue que, para $k \in \{1, \ldots, p\}$:

$$\sqrt{n}\,\overline{\mathcal{L}}_{\theta_k}(\widetilde{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\pi}}_n) - \sqrt{n}\,\overline{\mathcal{L}}_{\theta_k}(\boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}}_n) - \mathbf{V}_{nk}\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^t \xrightarrow{\text{q.c.}} 0 \ , \tag{3.21}$$

onde $\mathbf{V}_{nk} \xrightarrow{\text{q.c.}} -I^{(k)}_{\boldsymbol{\theta}_0}$, sendo $I^{(k)}_{\boldsymbol{\theta}_0}$ a k-ésima fila da matriz $I_{\boldsymbol{\theta}_0}$.

De (3.20) segue (i), de (3.21) (ii), e (iii) segue de (i) e (ii).

Isso prova o lema. $\square$

A seguir enunciamos e provamos o resultado principal desta seção.

Teorema 3.2.6 *Seja* $(X_1, \ldots, X_n)$ *uma amostra aleatória da distribuição* $F_{\boldsymbol{\theta}_0,\boldsymbol{\pi}_0} \in \mathcal{F}$, *com* $\mathcal{F}$ *satisfazendo as condições de regularidade (C1) - (C5), e* $\widehat{\boldsymbol{\pi}}_n = \widehat{\boldsymbol{\pi}}_n(X_1, \ldots, X_n)$ *um estimador fortemente consistente de* $\boldsymbol{\pi}_0$, *tal que*

$$\sqrt{n}(\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\pi}_0), \widehat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0) \overset{\mathcal{D}}{\longmapsto} \mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2) \sim N_{p+q}(\mathbf{0}, \Sigma) \ , \tag{3.22}$$

onde $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12}^t \\ \Sigma_{12}^t & \Sigma_{22} \end{pmatrix}$.

Seja $(\widetilde{\theta}_n)_{n\geq 1}$, *tal que* $\widetilde{\boldsymbol{\theta}}_n \xrightarrow{\text{q.c.}} \boldsymbol{\theta}_0$ *e, com probabilidade 1,* $\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\widetilde{\theta}_n, \widehat{\boldsymbol{\pi}}_n) = \mathbf{0}$, *para* n *suficientemente grande. Então, temos:*

$$\sqrt{n}(\widetilde{\theta}_n - \boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0) \overset{\mathcal{D}}{\longmapsto} \mathbf{V} \sim N_{p+q}(0_{p+q}, \Sigma^*) \ , \tag{3.23}$$

onde $\Sigma^* = \begin{pmatrix} I^{-1}_{\boldsymbol{\theta}_0} + I^{-1}_{\boldsymbol{\theta}_0} I_{\boldsymbol{\theta}_0,\pi_0}\Sigma_{22} I^t_{\boldsymbol{\theta}_0,\pi_0} I^{-1}_{\boldsymbol{\theta}_0} & -I^{-1}_{\boldsymbol{\theta}_0} I_{\boldsymbol{\theta}_0,\pi_0}\Sigma_{22} \\ -\Sigma_{22} I^t_{\boldsymbol{\theta}_0,\pi_0} I^{-1}_{\boldsymbol{\theta}_0} & \Sigma_{22} \end{pmatrix}$.

Prova: Pelas propriedades das seqüências $(\widetilde{\theta}_n)_{n\geq 1}$ e $(\widehat{\pi}_n)_{n\geq 1}$ segue, do Lema 3.2.5 (iii) que:

$$\sqrt{n}[\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\theta_0, \pi_0) + (\widehat{\pi}_n - \boldsymbol{\pi}_0)\mathbf{Y}_n^t] + \sqrt{n}(\widetilde{\theta}_n - \boldsymbol{\theta}_0)\mathbf{V}_n^t \xrightarrow{\text{q.c.}} \mathbf{0}_{1\times p} \;, \tag{3.24}$$

onde $\mathbf{Y}_n^t \xrightarrow{\text{q.c.}} -I^t_{\boldsymbol{\theta}_0\boldsymbol{\pi}_0}$ e $\mathbf{V}_n^t \xrightarrow{\text{q.c.}} -I^t_{\boldsymbol{\theta}_0} = -I_{\boldsymbol{\theta}_0}$.

Logo, segue de (3.24), de (3.23) e do teorema de Slutsky que

$$\sqrt{n}(\widetilde{\theta}_n - \boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0)$$

é assintoticamente equivalente a

$$\mathbf{U}_n = \sqrt{n}(\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\theta_0, \pi_0)I^{-1}_{\boldsymbol{\theta}_0} - (\widehat{\pi}_n - \boldsymbol{\pi}_0)I^t_{\boldsymbol{\theta}_0,\pi_0}I^{-1}_{\boldsymbol{\theta}_0},\;\; \widehat{\pi}_n - \boldsymbol{\pi}_0) \;.$$

Assim, basta determinar a distribuição assintótica de $\mathbf{U}_n$. Para isto usamos o teorema de Cramér Wold.

$\forall \mathbf{t} = (\mathbf{t}_1, \mathbf{t}_2) \in \mathbb{R}^{p+q}$, segue de (3.23) e do Lema 3.2.2 que:

$$\begin{aligned}
\mathbf{t}\mathbf{U}_n^t &= \mathbf{t}_1 I^{-1}_{\boldsymbol{\theta}_0}\sqrt{n}\,\overline{\mathcal{L}}^t_{\boldsymbol{\theta}}(\theta_0, \boldsymbol{\pi}_0) + (-\mathbf{t}_1 I^{-1}_{\boldsymbol{\theta}_0} I_{\boldsymbol{\theta}_0,\pi_0} + \mathbf{t}_2)\sqrt{n}(\widehat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0)^t \\
&\overset{\mathcal{D}}{\mapsto} \mathbf{t}_1 I^{-1}_{\boldsymbol{\theta}_0}\mathbf{W}_1^t + (-\mathbf{t}_1 I^{-1}_{\boldsymbol{\theta}_0} I_{\boldsymbol{\theta}_0,\pi_0} + \mathbf{t}_2)\mathbf{W}_2^t \sim N(0, \sigma^*) \;,
\end{aligned}$$

onde:

$$\begin{aligned}
\sigma^* &= \mathbf{t}_1 I^{-1}_{\boldsymbol{\theta}_0} I_{\boldsymbol{\theta}_0}(\mathbf{t}_1 I^{-1}_{\boldsymbol{\theta}_0})^t + (-\mathbf{t}_1 I^{-1}_{\boldsymbol{\theta}_0} I_{\boldsymbol{\theta}_0,\pi_0} + \mathbf{t}_2)\Sigma_{22}(-\mathbf{t}_1 I^{-1}_{\boldsymbol{\theta}_0} I_{\boldsymbol{\theta}_0,\pi_0} + \mathbf{t}_2)^t \\
&= \mathbf{t}_1(I^{-1}_{\boldsymbol{\theta}_0} + I^{-1}_{\boldsymbol{\theta}_0} I_{\boldsymbol{\theta}_0,\pi_0}\Sigma_{22} I^t_{\boldsymbol{\theta}_0,\pi_0} I^{-1}_{\boldsymbol{\theta}_0})\mathbf{t}_1^t - \mathbf{t}_1(I^{-1}_{\boldsymbol{\theta}_0} I_{\boldsymbol{\theta}_0,\pi_0}\Sigma_{22})\mathbf{t}_2^t \\
&\quad -\mathbf{t}_2(\Sigma_{22} I^t_{\boldsymbol{\theta}_0,\pi_0} I^{-1}_{\boldsymbol{\theta}_0})\mathbf{t}_1^t + \mathbf{t}_2\Sigma_{22}\mathbf{t}_2^t \\
&= (\mathbf{t}_1, \mathbf{t}_2)\Sigma^*(\mathbf{t}_1, \mathbf{t}_2)^t \\
&= \mathbf{t}\Sigma^*\mathbf{t}^t \;.
\end{aligned}$$

Isto é, temos que $\forall \mathbf{t} \in \mathbb{R}^{p+q}$,

$$\mathbf{t}\mathbf{U}_n^t \overset{\mathcal{D}}{\mapsto} N(0, \mathbf{t}\Sigma^*\mathbf{t}^t) \;.$$

Portanto, $\mathbf{U}_n \overset{\mathcal{D}}{\mapsto} N_{p+q}(\mathbf{0}_{p+q}, \Sigma^*)$.

Isto prova o teorema. □

Depois deste teorema podemos enunciar condições sob as quais o estimador de pseudo máxima verossimilhança, $\widehat{\theta}_n$, tem uma distribuição assintótica normal, isto é, temos o seguinte corolário:

Corolário 3.2.7 *Consideremos as hipóteses do Teorema 3.2.6. Assumamos também que para cada $n \in \mathbb{N}_+$, a equação $\overline{\mathcal{L}}_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \widehat{\boldsymbol{\pi}}_n) = 0$, vista como função de $\boldsymbol{\theta}$, tem uma única raiz. Então, se $\widehat{\boldsymbol{\theta}}_n = \widehat{\theta}_n(\widehat{\pi}_n)$ é o estimador de pseudo máxima verossimilhança associado a $\widehat{\boldsymbol{\pi}}_n$, temos que:*

$$\sqrt{n}(\widetilde{\theta}_n - \boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}}_n - \boldsymbol{\pi}_0) \overset{\mathcal{D}}{\mapsto} \mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2) \sim N_{p+q}(0_{p+q}, \Sigma^*) \ , \tag{3.25}$$

onde Σ^* é como no Teorema 3.2.6.

Em particular,

$$\sqrt{n}(\widehat{\theta}_n - \boldsymbol{\theta}_0) \overset{\mathcal{D}}{\mapsto} \mathbf{V}_1 \sim N_p(0, \Sigma_{11}^*) \ , \tag{3.26}$$

onde $\Sigma_{11}^* = I_{\boldsymbol{\theta}_0}^{-1} + I_{\boldsymbol{\theta}_0}^{-1} I_{\boldsymbol{\theta}_0,\pi_0} \Sigma_{22} I_{\boldsymbol{\theta}_0,\pi_0}^t I_{\boldsymbol{\theta}_0}^{-1}$.

Prova: Decorre do Corolário 3.1.9 e do Teorema 3.2.6. □

Observação 3.2.8 Como no caso uniparamétrico (ver Observação 2.2.7), quando $\Pi = \{\pi_0\}$, com π_0 conhecido, os resultados dos Teoremas 3.1.8 e 3.2.6 se reduzem aos resultados usuais onde todos os parâmetros são estimados pelo método de máxima verossimilhança.

Capítulo 4

Aplicação – Convolução de distribuições discretas

Neste capítulo, apresentamos uma aplicação do método de pseudo máxima verossimilhança a um modelo de convolução discreto, desenvolvido por Gong e Samaniego (1978).

4.1 Modelos de convolução com ruído

Seja X uma variável aleatória igual à soma de duas variáveis independentes Y e Z. Então, a distribuição de X, F_X, é a convolução das distribuições de Y e Z, que denotamos por F_Y e F_Z, respectivamente, e pode ser vista como um modelo para sinais com ruído.

Existem muitas situações onde ruído está presente. Por exemplo, dados obtidos por um contador Geiger podem ser vistos como ocorrências devidas à presença de uma substância radioativa e ocorrências devidas ao ruído ou estática.

Problemas de estimação para distribuições de sinais com ruído tem sido examinados por vários autores. Dentre outros, Gaffey (1959) construiu um estimador consistente para a distribuição de um componente de convolução contínua sob a suposição de que a distribuição correspondente ao ruído seja conhecida. Sclove e Van Ryzin (1969) obtem estimadores pelo método de momentos e deduzem as suas correspondentes variâncias assintóticas, para uma variedade de distribuições multiparamétricas de sinais com ruído.

Devido à natureza inconveniente da função de verossimilhança para estes modelo, a estimação por máxima verossimilhança para estes modelos tem encontrado resistência substancial. No entanto, tem havido algum progresso em problemas uniparamétricos. Por exemplo, Samaniego (1976) prova o seguinte resultado:

Teorema 4.1.1 *Seja X uma variável aleatória assumindo valores inteiros não negativos, cuja distribuição está indexada por um parâmetro positivo θ. Se a função de probabilidade de X, P_θ, é diferenciável em relação a θ, então*

$$\forall n,\ \forall \theta,\ \frac{\partial}{\partial \theta} P_\theta(X = n) = P_\theta(X = n - 1) - P_\theta(X = n) \ ,$$

se, e somente se, a distribuição de X é a convolução de uma distribuição de Poisson com parâmetro θ e a distribuição de uma variável aleatória, assumindo valores inteiros não negativos e que independe de θ.

Do teorema precedente segue o seguinte resultado:

Corolário 4.1.2 *Sejam Y e Z duas variáveis aleatórias independentes, com distribuições F_Y e F_Z, respectivamente, sendo F_Y uma distribuição de Poisson com parâmetro θ e Z assumindo valores inteiros não negativos. Seja X com distribuição F_X igual a convolução entre F_Y e F_Z. Então, se $X_1, \ldots, X_n$ é uma amostra aleatória da distribuição F_X, a equação de verossimilhança $\frac{\partial}{\partial\theta}\overline{\mathcal{L}}_n(X_1, \ldots, X_n; \theta) = 0$, pode ser escrita como*

$$\sum_{j=1}^{n} \frac{P_\theta(X_j = x_{i-1})}{P_\theta(X_j = x_i)} = n \ , \tag{4.1}$$

onde $\mathbf{x} = (x_1, \ldots, x_n)$ é o vetor de observações.

Do corolário anterior segue que o comportamento das razões dos quocientes de probabilidades tais como

$$\frac{P_\theta(X = a)}{P_\theta(X = b)} \ , \tag{4.2}$$

para $a < b$, é relevante para o problema de estimação.

A propriedade da razão paramétrica monótona decrescente (RPMD) é definida em Samaniego (1976), significando que as razões da forma (4.2) são decrescentes.

Para distribuições de Poisson com RPMD a equação de verossimilhança (4.1) tem como máximo uma solução; e o estimador de máxima verossimilhança pode ser encontrado numericamente.

Entre os modelos que têm a propriedade RPMD está a convolução Poisson-binomial, a qual denotamos por $\mathcal{P}(\theta) * \mathcal{B}(N, \pi)$, com π conhecido. Resultados para o último modelo, com ambos os parâmetros conhecidos são obtidos em Gong e Samaniego (1978); e resultados similares para convoluções de distribuições binomial são obtidas em Samaniego (1977). Distribuições de convolução de Pascal são estudadas em Gong e Samaniego (1978).

4.2 Estimação por pseudo máxima verossimilhança no modelo $\mathcal{P}(\theta) * \mathcal{B}(N, \pi)$

Em muitas situações nas quais um modelo de convolução é adequado, existe um parâmetro de interesse particular, os outros podendo ser considerados como parâmetros de incômodo ("nuisance"). Nestes casos, quando os parâmetros de incômodo são substituídos por estimadores consistentes (são obtidos com facilidade), então os estimadores de pseudo máxima verossimilhança são também facilmente obtidos. A convolução $\mathcal{P}(\theta) * \mathcal{B}(N, \pi)$ é um bom exemplo de um problema no qual a estimação por máxima verossimilhança de (θ, π) é muito complicada, mas a estimação por pseudo máxima verossimilhança de um ou outro parâmetro é facilmente implementada.

4.2.1 "Regularidade" dos modelos de sinais com ruído

Gong e Samaniego (1978) demonstram que o modelo de convolução com sinais tendo uma distribuição de Poisson e ruído com distribuição normal, satisfaz as condições de regularidade sob as quais foi desenvolvida a teoria assintótica do estimador de máxima verossimilhança; e estabelecem que outros três modelos são regulares: sinais com distribuição de Poisson e ruído com distribuição binomial, sinais com distribuição binomial e ruído com distribuição de Poisson ou normal.

Supondo satisfeitas as condições de regularidade, aplicaremos a teoria desenvolvida no Capítulo 2, para o modelo de convolução com sinais tendo distribuição de Poisson e ruído com distribuição binomial. Antes apresentamos algumas notações que serão usadas até o final do capítulo.

$X_1, \ldots, X_n$ será uma amostra aleatória de $X = Y + Z$, com Y e Z independentes, tendo Y uma distribuição de Poisson com parâmetro θ_0, desconhecido, e Z com distribuição binomial de parâmetros N, conhecido, e π_0, desconhecido, ou seja, $Y \sim \mathcal{P}(\theta_0)$, $Z \sim \mathcal{B}(N, \pi_0)$ e $X \sim \mathcal{P}(\theta_0) * \mathcal{B}(N, \pi_0)$. Consideramos θ_0 como parâmetro de interesse e π_0 como de incômodo.

Como no Capítulo 2:

$$\overline{\mathcal{L}}_n(\theta_0, \pi_0) = \frac{1}{n} \sum_{j=1}^{n} \frac{\partial}{\partial \theta} \ln f(X_j | \theta_0, \pi_0)$$
$$\overline{\mathcal{L}}_\theta(\theta_0, \pi_0) = \frac{\partial}{\partial \theta} \overline{\mathcal{L}}_n(\theta_0, \pi_0) ,$$

onde $f(\cdot | \theta_0, \pi_0)$ é a função de probabilidade de X.

Além disso, consideramos também que $\widehat{\pi}_n = \widehat{\pi}_n(X_1, \ldots, X_n)$ é um estimador de π_0 obtido pelo método dos momentos.

Por último,

$$\overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j \quad \text{e} \quad S_n^2 = \frac{1}{n} \sum_{j=1}^{n} (X_j - \overline{X}_n)^2 .$$

A seguir, apresentamos alguns resultados relacionados com a estimação por pseudo máxima verossimilhança do parâmetro θ_0 no modelo de convolução $\mathcal{P}(\theta) * \mathcal{B}(N, \pi)$.

Lema 4.2.1 $\widehat{\pi}_n$ *é fortemente consistente e, com probabilidade 1, para n suficientemente* *grande, temos que:*

$$\hat{\pi}_n = [(\overline{X}_n - S_n^2)/N]^{1/2} .$$

Prova: Pela definição de X, temos que

$$E(X) = \theta_0 + N\pi_0 \tag{4.3}$$

$$E(X^2) = \theta_0 + N\pi_0(1-\pi_0) + (\theta_0 + N\pi_0)^2 . \tag{4.4}$$

Assim, pelo método dos momentos, temos que o estimador de π_0 é

$$\hat{\pi}_n = [(\overline{X}_n - S_n^2)/N]^{1/2} , \tag{4.5}$$

se $\overline{X}_n - S_n^2 \geq 0$.

Por outro lado, pela lei forte dos grandes números de Kolmogorov, temos que

$$(\overline{X}_n - S_n^2)/N \xrightarrow{\text{q.c.}} \pi_0^2 . \tag{4.6}$$

O lema decorre de (4.5) e (4.6).

Lema 4.2.2 *Seja V uma variável aleatória assumindo valores inteiros não negativos, tendo função de probabiidad $f_V(\cdot|\theta)$, $\theta \in \mathbb{R}_+$, com a propriedade RPMD. Seja W uma variável aleatória com distribuição $\mathcal{B}(1,\pi)$, com π conhecido, e independente de V. Então, $U = V + W$ tem função de probabilidade $f_U(\cdot|\theta)$ com a propriedade RPMD.*

Prova: Para todo $x \in \mathbb{N}$:

$$\begin{aligned} f_U(x|\theta) &= P_\theta(U = x) \\ &= P_\theta(W=0)P_\theta(V=x) + P_\theta(W=1)P_\theta(V=x-1) \\ &= (1-\pi)f_V(x/\theta) + \pi f_V(x-1|\theta) . \end{aligned}$$

Logo, para $a < b$,

$$\frac{f_U(a|\theta)}{f_U(b|\theta)} = \frac{\frac{(1-\pi)f_V(a|\theta)}{f_V(b-1|\theta)} + \frac{\pi f_V(a-1|\theta)}{f_V(b-1|\theta)}}{\frac{(1-\pi)f_V(b|\theta)}{f_V(b-1|\theta)} + \pi} .$$

Daqui, segue o lema, pois $f_V(\cdot|\theta)$ tem a propriedade RPMD. □

Lema 4.2.3 *Para cada $\pi \in (0,1)$, $f(\cdot|\theta,\pi)$ tem a propriedade RPMD.*

Prova: $f(\cdot|\theta,\pi)$ é a função de probabilidade de $X = Y + Z$, onde Y e Z são independentes, $Y \sim \mathcal{P}(\theta)$ e $Z \sim B(N,\pi)$.

Como para todo $(x_1, x_2) \in \mathbb{N} \times \mathbb{N}$ e para todo $\theta \in \mathbb{R}_+$,

$$\frac{f_Y(x_1|\theta)}{f_Y(x_2|\theta)} = \frac{\theta^{x_1 - x_2} x_2!}{x_1!} \; .$$

Então, $f_Y(\cdot|\theta)$ tem a propriedade RPMD. Logo, se $N = 1$, segue pelo Lema 4.2.~~3~~2 que $f(\cdot|\theta) = f_X(\cdot|\theta)$ tem a propriedade RPMD, isto é, o lema vale para $N = 1$. Por outro lado, $Z = \sum_{j=1}^{N} Z_j$, onde as Z_j são independentes e com distribuição $\mathcal{B}(1,\pi)$, $j = 1, \ldots, N$. Portanto, usando indução e o Lema 4.2.2, prova-se o lema. □

Lema 4.2.4 *Para todo $x \in \mathbb{N}$ e todo $(\theta, \pi) \in \mathbb{R}_+ \times (0,1)$,*

$$\frac{\partial}{\partial\theta} f(x|\theta,\pi) = f(x-1|\theta;\pi) - f(x|\theta,\pi) \; ,$$

onde $f(-1|\theta,\pi) := 0$.

Prova: $f(\cdot|\theta,\pi)$ é a função de probabilidade da distribuição de convolução $\mathcal{P}(\theta) * \mathcal{B}(N,\pi)$. Então, $\forall x \in \mathbb{N}$ e $\theta \in (0,1)$:

$$f(x|\theta,\pi) = \sum_{y=0}^{a_x} \binom{N}{y} \pi^y (1-\pi)^{N-y} e^{-\theta} \theta^{x-y}/(x-y)! \; ,$$

onde $a_x = \min\{x, N\}$.

$$\begin{aligned}
\frac{\partial}{\partial\theta} f(x|\theta,\pi) &= \sum_{y=0}^{a_x} \binom{N}{y} \pi^y (1-\pi)^{N-y} e^{-\theta} (x-y) \theta^{x-y-1}/(x-y)! \\
&\quad - \sum_{y=0}^{a_x} \binom{N}{y} \pi^y (1-\pi)^{N-y} e^{-\theta} \theta^{x-y}/(x-y)! \\
&= \sum_{y=0}^{a_{x-1}} \binom{N}{y} \pi^y (1-\pi)^{N-y} e^{-\theta} \theta^{x-1-y}/(x-1-y)! - f(x|\theta,\pi) \\
&= f(x-1|\theta,\pi) - f(x|\theta,\pi) \; ,
\end{aligned}$$

onde definimos $f(-1|\theta,\pi) = 0$.

Isso prova o lema. □

Lema 4.2.5 *Para cada $\pi \in (0,1)$ e $n \in \mathbb{N}_+$, a equação $\overline{\mathcal{L}}_\theta(\theta,\pi) = 0$, vista como função de θ, tem como máximo uma raiz.*

Prova: Como $\overline{\mathcal{L}}_\theta(\theta,\pi) = \frac{1}{n}\sum_{j=1}^{n} \frac{\partial}{\partial\theta} \ln f(X_j|\theta,\pi)$, então, pelo Lema 4.2.4, a equação $\overline{\mathcal{L}}_\theta(\theta,\pi) = 0$ é equivalente a

$$\sum_{j=1}^{n} \frac{f(X_j - 1|\theta,\pi)}{f(X_j|\theta,\pi)} = n \ . \qquad (4.7)$$

Daqui, pelo Lema 4.2.3, decorre o lema. □

Com os resultados anteriores e supondo satisfeitas as condições de regularidade dadas no Capítulo 2, podemos estabelecer o seguinte resultado:

Teorema 4.2.6 *Seja $\widehat{\theta}_n = \widehat{\theta}_n(\widehat{\pi}_n)$ o estimador de pseudo máxima verossimilhança de θ_0 associado a $\widehat{\pi}_n$, então $\widehat{\theta}_n$ é fortemente consistente.*

Prova: Segue do Lema 4.2.1, do Lema 4.2.5 e do Corolário 2.1.7. □

A seguir apresentamos alguns resultados relacionados com a distribuição assintótica do estimador de pseudo máxima verossimilhança.

Lema 4.2.7 *Sejam $\mu = E(X)$ e $\sigma^2 = \mathrm{Var}(X)$, então*

$$\sqrt{n}\left(\overline{\mathcal{L}}_\theta(\theta_0,\pi_0), \overline{X}_n - \mu, n^{-1}\sum_{j=1}^{n}(X_j-\mu)^2 - \sigma^2\right) \overset{\mathcal{D}}{\mapsto} \mathbf{V} = (V_1, V_2, V_3) \sim N_3(\mathbf{0},\Gamma),$$

onde $\Gamma = (\Gamma_{ij})_{3\times 3}$, com

$$\Gamma_{11} = I_{\theta_0} \ , \quad \Gamma_{12} = \Gamma_{21} = \Gamma_{13} = \Gamma_{31} = 1 \ ,$$

$$\Gamma_{22} = \theta_0 + N\pi_0(1-\pi_0) \ ,$$

$$\Gamma_{23} = \Gamma_{32} = \theta_0 + N\pi_0(1-\pi_0)(1-2\pi_0) \ ,$$

$$\Gamma_{33} = \theta_0 + 2[\theta_0 + N\pi_0(1-\pi_0)]^2 + N\pi_0(1-\pi_0)[1 - 6\pi_0(1-\pi_0)] \ .$$

Prova: Como as X_j, $j = 1, \ldots, n$, são variáveis independentes e identicamente distribuídas e, além disso, $E\left[\frac{\partial}{\partial\theta}\ln f(X|\theta_0, \pi_0)\right] = 0$, $E(\overline{X}_n) = \mu$ e $E\left[n^{-1}\sum_{j=1}^{n}(X_j - \mu)^2\right] = \sigma^2$. Segue do teorema central do limite multivariado que:

$$\sqrt{n}\left(\mathcal{L}_\theta(\theta_0, \pi_0), \overline{X}_n - \mu, n^{-1}\sum_{j=1}^{n}(X_j - \mu)^2 - \sigma^2\right) \overset{\mathcal{D}}{\mapsto} \mathbf{V} = (V_1, V_2, V_3) \sim N_3(\mathbf{0}, \Gamma),$$

onde $\Gamma = (\Gamma_{ij})_{3\times 3}$, com

$$\begin{aligned}
\Gamma_{11} &= \mathrm{Var}\left[\frac{\partial}{\partial\theta}\ln f(X|\theta_0, \pi_0)\right] = E\left[\frac{\partial}{\partial\theta}\ln f(X|\theta_0, \pi_0)\right]^2 = I_{\theta_0} \\
&= \sum_{x=0}^{\infty}\left[\frac{\partial}{\partial\theta}\ln f(x|\theta_0, \pi_0)\right]^2 f(x|\theta_0, \pi_0) \\
&= \sum_{x=0}^{\infty}\left[\frac{\partial}{\partial\theta}f(x|\theta_0, \pi_0)\right]^2 / f(x|\theta_0, \pi_0) \\
&= \sum_{x=0}^{\infty}\left[f(x-1|\theta_0, \pi_0) - f(x|\theta_0, \pi_0)\right]^2 / f(x|\theta_0, \pi_0) \ .
\end{aligned}$$

Nesta última igualdade usamos o Lema 4.2.5.

$$\begin{aligned}
\Gamma_{12} &= \Gamma_{21} = \mathrm{Cov}\left(\frac{\partial}{\partial\theta}\ln f(X|\theta_0, \pi_0), X\right) \\
&= E\left[X\frac{\partial}{\partial\theta}\ln f(X|\theta_0, \pi_0)\right] \\
&= \sum_{x=0}^{\infty}\left[\frac{\partial}{\partial\theta}\ln f(x|\theta_0, \pi_0)\right] x f(x|\theta_0, \pi_0) \\
&= \sum_{x=0}^{\infty}\left[\frac{\partial}{\partial\theta}f(x|\theta_0, \pi_0)\right] x \\
&= \sum_{x=0}^{\infty}\left[f(x-1|\theta_0, \pi_0) - f(x|\theta_0, \pi_0)\right] x \\
&\quad \text{(a última igualdade segue do Lema 4.2.5)} \\
&= \sum_{x=1}^{\infty} x f(x-1)/\theta_0, \pi_0) - \sum_{x=0}^{\infty} x f(x|\theta_0, \pi_0) \\
&\quad \text{(isto último porque } f(-1|\theta_0, \pi_0) = 0 \quad \text{e} \\
&\quad \sum_{x=0}^{\infty} x f(x|\theta_0, \pi_0) = \mu \in \mathbb{R}), \\
&\quad \sum_{x=0}^{\infty}(x+1) f(x|\theta_0, \pi_0) - \sum_{x=0}^{\infty} x f(x|\theta_0, \pi_0) = \sum_{x=0}^{\infty} f(x|\theta_0, \pi_0) = 1 \ .
\end{aligned}$$

$$
\begin{aligned}
\Gamma_{13} &= \Gamma_{31} = \mathrm{Cov}\left(\frac{\partial}{\partial\theta}\ln f(X|\theta_0,\pi_0),(X-\mu)^2\right) \\
&= E\left[(X-\mu)^2\frac{\partial}{\partial\theta}\ln f(X|\theta_0,\pi_0)\right] \\
&= \sum_{x=0}^{\infty}(x-\mu)^2\left[\frac{\partial}{\partial\theta}\ln f(x|\theta_0,\pi_0)\right]xf(x|\theta_0,\pi_0) \\
&= \sum_{x=0}^{\infty}(x-\mu)^2\frac{\partial}{\partial\theta}f(x|\theta_0,\pi_0) \\
&= \sum_{x=0}^{\infty}(x-\mu)^2\left[f(x-1|\theta_0,\pi_0)-f(x|\theta_0,\pi_0)\right] \\
&\quad \text{(esta última igualdade deve-se ao Lema 4.2.5)} \\
&= \sum_{x=1}^{\infty}(x-\mu)^2 f(x-1)/\theta_0,\pi_0) - \sum_{x=0}^{\infty}(x-\mu)^2 f(x|\theta_0,\pi_0) \\
&\quad \text{(a última igualdade segue de } f(-1|\theta_0,\pi_0)=0 \text{ e porque} \\
&\quad \sum_{x=0}^{\infty}(x-\mu)^2 f(x|\theta_0,\pi_0)=\sigma^2\in \mathbb{R}_+) \\
&= \sum_{x=0}^{\infty}[(x-\mu)+1]^2 f(x|\theta_0,\pi_0)-\sigma^2 \\
&= \sum_{x=0}^{\infty}(x-\mu)^2 f(x|\theta_0,\pi_0) - 2\sum_{x=0}^{\infty}(x-\mu)f(x|\theta_0,\pi_0) \\
&\quad +\sum_{x=0}^{\infty} f(x|\theta_0,\pi_0)-\sigma^2 = 1\ .
\end{aligned}
$$

$$
\begin{aligned}
\Gamma_{22} &= \mathrm{Var}(X) = \sigma^2 = \theta_0 + N\pi_0(1-\pi_0)\ , \\
\Gamma_{23} &= \mathrm{Cov}(X-\mu,(X-\mu)^2) \\
&= \mathrm{Cov}([Y-\mu_Y]+[Z-\mu_2],[Y-\mu_Y]^2+2[Y-\mu_Y][Z-\mu_Z]+[Z-\mu_Z]^2) \\
&= E[Y-\mu_Y]^2+E[Z-\mu_Z]^3 \\
&= \theta_0+N\pi_0(1-\pi_0)(1-2\pi_0)\ .
\end{aligned}
$$

Aqui foi usada a propriedade de bilinearidade da covariância e o fato que Y e Z são independentes, sendo $Y \sim \mathcal{P}(\theta_0)$ e $Z \sim \mathcal{B}(N,\pi_0)$.

$$
\Gamma_{33} = \mathrm{Var}(X-\mu)^2
$$

$$
\begin{aligned}
&= \operatorname{Var}([Y-\mu_Y]^2 + 2[Y-\mu_Y][Z-\mu_Z] + [Z-\mu_Z]^2) \\
&= E[Y-\mu_Y]^4 - E^2[Y-\mu_Y]^2 + 2E[Y-\mu_Y]^2 E[Z-\mu_Z]^2 \\
&\quad + E[Z-\mu_Z]^4 - E^2[Z-\mu_Z]^2 \\
&= 3N^2\pi_0^2(1-\pi_0)^2 + N\pi_0(1-\pi_0)[1-6\pi_0(1-\pi_0)] - N^2\pi_0^2(1-\pi_0)^2 \\
&\quad + 2N\pi_0(1-\pi_0)\theta_0 + (\theta_0 + 3\theta_0^2) - \theta_0^2 \\
&= 2N^2\pi_0^2(1-\pi_0)^2 + 2N\pi_0(1-\pi_0)\theta_0 + 2\theta_0^2 + \theta_0 \\
&\quad + N\pi_0(1-\pi_0)[1-6\pi_0(1-\pi_0)] \\
&= 2[N\pi_0(1-\pi_0) + \theta_0]^2 + \theta_0 + N\pi_0(1-\pi_0)[1-6\pi_0(1-\pi_0)]\ .
\end{aligned}
$$

Isso prova o lema. □

Lema 4.2.8

$$
\sqrt{n}(\overline{\mathcal{L}}_\theta(\theta_0,\pi_0), \widehat{\pi}_n - \pi_0) \overset{\mathcal{D}}{\mapsto} \mathbf{W} = (W_1, W_2) \sim N_2(\mathbf{0}, \Sigma)\ ,
$$

onde

$$
\Sigma = \begin{pmatrix} \Gamma_{11} & 0 \\ 0 & t^2(\Gamma_{22} - 2\Gamma_{23} + \Gamma_{33}) \end{pmatrix}\ ,
$$

com $t = (2N\pi_0)^{-1}$ *e* Γ_{11}, Γ_{22}, Γ_{33} *e* Γ_{23} *como no Lema 4.2.7.*

Prova: Pelo Lema 4.2.1 e o teorema de Slutsky, basta obter a distribuição limite de

$$
\begin{aligned}
&\sqrt{n}\left(\overline{\mathcal{L}}_\theta(\theta_0,\pi_0), [(\overline{X}_n - S_n^2)/N]^{1/2} - \pi_0\right) \\
&= \sqrt{n}\left(g(\overline{\mathcal{L}}_\theta(\theta_0,\pi_0), \overline{X}_n, S_n^2) - g(0,\mu,\sigma^2)\right)\ ,
\end{aligned} \tag{4.8}
$$

onde $\forall (t_1,t_2,t_3) \in \mathbb{R}^3$, $t_2 - t_3 \geq 0$:

$$
g(t_1,t_2,t_3) = (t_1, [(t_2-t_3)/N]^{1/2})\ ,
$$

sendo g diferenciável com

$$
g'(t_1,t_2,t_3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (2N)^{-1}[(t_2-t_3)/N]^{-1/2} & -(2N)^{-1}[(t_2-t_3)/N]^{-1/2} \end{pmatrix}\ .
$$

Por outro lado, $\sqrt{n}(\overline{\mathcal{L}}_\theta(\theta_0,\pi_0), \overline{X}_n - \mu, S_n^2 - \sigma^2)$ é assintoticamente equivalente a

$$\sqrt{n}\left(\overline{\mathcal{L}}_\theta(\theta_0,\pi_0), \overline{X}_n - \mu, n^{-1}\sum_{j=1}^{n}(X_j-\mu)^2 - \sigma^2\right) .$$

Portanto, pelo teorema de Slutsky e o Lema 4.2.7:

$$\sqrt{n}\left(\overline{\mathcal{L}}_\theta(\theta_0,\pi_0), \overline{X}_n - \mu, S_n^2 - \sigma^2\right) \overset{\mathcal{D}}{\mapsto} \mathbf{V} = (V_1, V_2, V_3) \sim N_3(\mathbf{0}, \Gamma) , \tag{4.9}$$

onde $\Gamma = (\Gamma_{ij})_{3\times 3}$ é como no Lema 4.2.7.

Logo, usando o método delta (Serfling, 1980) junto com (4.8) e (4.9), provamos o Lema. □

Com os resultados anteriores podemos estabelecer a distribuição assintótica do estimador de pseudo máxima verossimilhança.

Teorema 4.2.9 *Se $\widehat{\theta}_n = \widehat{\theta}_n(\widehat{\pi}_n)$ é o estimador de pseudo máxima verossimilhança associado a $\widehat{\pi}_n$, então:*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \overset{\mathcal{D}}{\mapsto} Z \sim N(0,\sigma^2) ,$$

onde $\sigma^2 = I_{\theta_0}^{-1} + I_{\theta_0}^{-2} I_{\theta_0,\pi_0}\sigma_{22}$, com

$$\begin{aligned}
I_{\theta_0} &= E\left[\frac{\partial}{\partial\theta}\ln f(X|\theta_0,\pi_0)\right]^2 \\
&= \sum_{x=0}^{\infty}[f(x-1|\theta_0,\pi_0) - f(x|\theta_0,\pi_0)]^2/f(x|\theta_0,\pi_0) \qquad (4.10)\\
I_{\theta_0,\pi_0} &= E\left[\frac{\partial^2}{\partial\theta\partial\pi}\ln f(X|\theta_0,\pi_0)\right] \\
&= \sum_{x=0}^{\infty}[f(x-1|N,\theta_0,\pi_0) - f(x|N,\theta_0,\pi_0)][f(x-1|N-1,\theta_0,\pi_0) \\
&\quad -f(x|N-1,\theta_0,\pi_0)]/f(x|N,\theta_0,\pi_0) \qquad (4.11)\\
\sigma_{22} &= t^2(\Gamma_{22} - 2\Gamma_{23} + \Gamma_{33}) , \qquad (4.12)
\end{aligned}$$

com t, Γ_{22}, Γ_{23} e Γ_{33} como no Lema 4.2.8.

Prova: Segue do Lema 4.2.8 e dos Teoremas 4.2.6 e 2.2.5. □

4.3 Eficiência assintótica relativa do estimador de pseudo máxima verossimilhança no modelo $\mathcal{P}(\theta) * \mathcal{B}(N, \pi)$

Gong e Samaniego (1978) comparam as variâncias assintóticas do estimador de pseudo máxima verossimilhança (EPMV) com a correspondente ao estimador de máxima verossimilhança (EMV) e ao estimador pelo método dos momentos (EMM), do parâmetro θ_0. As variâncias anteriores são:

$$
\begin{aligned}
\sigma^2_{EPMV} &= I_{\theta_0}^{-1} + I_{\theta_0}^{-2} I^2_{\theta,\pi_0} t^2 [\Gamma_{22} - 2\Gamma_{23} + \Gamma_{33}] \ ; & (4.13)\\
\sigma^2_{EMV} &= I_{\pi_0}(I_{\theta_0} I_{\pi_0} - I^2_{\theta_0,\pi_0})^{-1} \ ; & (4.14)\\
\sigma^2_{EMM} &= (1 - Nt)^2\Gamma_{22} + 2Nt(1 - Nt)\Gamma_{23} + (Nt)^2\Gamma_{33} \ . & (4.15)
\end{aligned}
$$

A dificuldade para comparar diretamente as três expressões está na impossibilidade de representar as componentes da matriz de informação de Fisher, I_{θ_0}, I_{θ_0,π_0} de maneira distinta das séries (4.10) e (4.11), respectivamente , e a componente I_{π_0} de uma maneira distinta da série

$$
\begin{aligned}
I_{\pi_0} &= E\left[\frac{\partial}{\partial \pi} \ln f(X|\theta_0, \pi_0)\right]^2 \\
&= N^2 \sum_{x=0}^{\infty} [f(x-1|N-1, \theta_0, \pi_0] - f(x|N-1, \theta_0, \pi_0)]^2 / f(x|N, \theta_0, \pi_0) \ . \quad (4.16)
\end{aligned}
$$

Os autores aproximam as informações I_{θ_0}, I_{θ_0,π_0} e I_{π_0} pela soma dos termos iniciais das séries (4.10), (4.11) e (4.16), desenvolvendo também limites para os erros E_{11}, E_{12} e E_{22}, de tais aproximações.

Para o erro E_{11}, na aproximação de (4.10): notemos que $g(x) = \frac{f(x-1|N,\theta,\pi)}{f(x|N,\theta,\pi)}$ é decrescente em π, pois procedendo como no Lema 4.2.4, podemos verificar que $f(\cdot|N, \theta, \pi)$ tem a propriedade RPMD para N e θ fixados arbitrariamente. Então, segue do Lema 4.2.5 que $\phi_\theta(x) = g(x) - 1$ também é decrescente em π. Logo, se $x \geq 1$, com $L > N + \theta$, temos que:

$$\begin{aligned}
|\phi_\theta(x)| &\leq \max\{|\phi_\theta(x|N,\theta,0)|, |\phi(x|N,\theta,1)|\} \\
&= \max\left\{\frac{|x-\theta|}{\theta}, \frac{|x-N-\theta|}{\theta}\right\} \\
&= \frac{x-\theta}{\theta}.
\end{aligned}$$

Portanto, temos para o erro E_{11} que

$$\begin{aligned}
E_{11} &= I_{\theta_0} - \sum_{x=0}^{L} \phi_\theta^2(x|N,\theta_0,\pi_0) f(x|N,\theta_0,\pi_0) \\
&= \sum_{x=L+1}^{\infty} \phi_\theta^2(x) f(x|N,\theta_0,\pi_0) \\
&\leq \sum_{x=L+1}^{\infty} \left(\frac{x-\theta}{\theta}\right)^2 f(x|N,\theta_0,\pi_0) \\
&= E\left(\frac{X-\theta}{\theta}\right)^2 - \sum_{x=0}^{L}\left(\frac{x-\theta}{\theta}\right)^2 f(x) \qquad (4.17)
\end{aligned}$$

Limitantes similares são estabelecidos para os erros

$$E_{12} = I_{\theta_0,\pi_0} - \sum_{x=0}^{L} \phi_\theta(x)\phi_\pi(x)$$

e

$$E_{22} = I_{\pi_0} - \sum_{x=0}^{L} \phi_\pi^2(x) .$$

Com tais limitantes, aproximaram I_{θ_0}, I_{θ_0,π_0} e I_{π_0}, com erros E_{ij} menores ou iguais a 10^{-7}; e estas aproximações foram usadas para calcular a eficiência assintótica do EPMV relativa ao EMV e a do EMM relativa ao EPMV. Os valores dos parâmetros examinados foram os seguintes:

$$\theta_0 = 0,1;\ 0,25;\ 0.5:\ 1:\ 10$$

$$\pi_0 = 0,1;\ 0,9$$

$$N = 1;\ 10.$$

Para os valores anteriores dos parâmetros, encontrou-se:

$$0,57288 \leq EAR(EPMV/EMV) < 1 \ ;$$

$$0,42712 \leq EAR(EMM/EPMV) < 1 \ .$$

Destes cálculos, pode-se conjeturar que este estimador de pseudo máxima verossimilhança de θ_0 é uniformemente assintoticamente mais eficiente do que o estimado pelo método dos momentos.

Referências

Bahadur, R.R. (1964). On the Fisher's bound for asymptotic variances. *Annals of Mathematical Statistics*, **35**:1545–1552.

Basu, D. (1977). On the elimination of nuisance parameters. *Journal of American Statistics Association*, **72**:355–366.

Bowman, K.O. and Shenton, L.R. (1988). *Properties of estimators for the gamma distribution.* Marcel Dekker, New York. of the parameters of the gamma distribution and their bias. *Technometrics*, **11**:683–690.

Chiang, Ch. and Cox, Ch. (1985). Pseudo maximum likelihood estimation for the Dirichlet-multinomial distribution. *Communication in Statistics. Theory and Methods*, **14**(10):2293–2311.

Cox, D.R. (1975). Partial likelihood. *Biometrika*, **62**, 269–276.

Cramér, H. (1946). *Mathematical methods of statistics.* Princeton University Press, Princeton, N.J.

Gaffey, W.R. (1959). A consistent estimator of a component of a convolution. *Annals of Mathematical Statistics*, **30**:198–205.

Gong, G.H. and Samaniego, F.J. (1978). Pseudo maximum likelihood estimation: theory and applications. *Technical Report No. 7*, Department of Mathematics, University of California, Davis.

Gong, G.H. and Samaniego, F.J. (1981). Pseudo maximum likelihood estimation: theory and applications. *Annals of Statistics*, **9**: 861–869.

James, B.R. (1981). *Probabilidade - um curso em nível intermediário.* Instituto de Matemática Pura e Aplicada - CNPq.

Lehmann, E.L. (1983). *Theory of point estimation.* John Willey, New York.

Parke, W.R. (1986). Pseudo maximum likelihood estimation: the asymptotic distribution. *Annals of Mathematical Statistics*, **14**:335–357.

Samaniego, F.J. (1976). A characterization of convuled Poisson distributions with application to estimation. *Journal of the American Statistical Association*, **71**:475–479.

Samaniego, F.J. (1980). Maximum likelihood estimation for binomially distributed signals in discrete noise. *Journal of the American Statistical Association*, **75**:117–121.

Serfling, R.J. (1980). *Approximation theorems of mathematical statistics.* Willey, New York.