ESTIMAÇÃO POR MÍNIMA DISTÂNCIA DE HELLINGER

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. **RONALDO DIAS** e aprovada pela Comissão Julgadora.

Campinas, 18 de agosto de 1988.

Prof. Dr. ______________________
JOSÉ ANTONIO CORDEIRO
(Orientador)

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestre em **ESTATÍSTICA**.

Tese de Mestrado apresentada em 15.08.1988 no Instituto de Matemática, Estatística e Ciência da Computação da Universidade Estadual de Campinas (UNICAMP) para a banca:

Prof. Dr. José Antonio Cordeiro (Orientador)

Prof. Dr. Jorge Alberto Achcar

Prof. Dr. Ademir Petenate

RONALDO DIAS

Aos meus pais

pelo apoio e incetivo que recebi desde o inicio de meus estudos.

A G R A D E C I M E N T O S

Primeiramente ao Prof. J.A. Cordeiro, agradeço-o duplamente. Não só pelas idéias que geraram esta tese, mas também pela amizade e companheirismo que se desenvolveram durante estes dezoito meses que estive na Unicamp.

Aos professores Mauro Marques e Oscar Bustos, presto meus agradecimentos. Ao primeiro pelo interesse e suporte técnico no que diz respeito a livros, e artigos publicados em revistas internacionais, nem sempre disponíveis na biblioteca do IMECC. Ao segundo, que apesar da distância, manteve o mesmo incetivo e o apoio moral de quando fui seu aluno.

Um agradecimento muito especial vai para minha esposa Nancy. Pela grande compreensão, pela paciência e carinho nos momentos dificéis, pelas sugestões e discussões sobre o assunto deste trabalho e pela excelente datilografia do mesmo.

Finalmente gostaria de agradecer aos amigos José Ramos, Myriam Otoni e Yara Rehder, pela ajuda dada com respeito aos softwares usados na Unicamp.

RONALDO DIAS

INDICE

INTRODUCAO

ESTIMACAO POR MINIMA DISTANCIA

A princípio vamos descrever brevemente o que é o problema de estimação paramétrica.

Sejam $X_1, X_2, \ldots, X_n$, n observações de um certo fenômeno aleatório e que a distribuição conjunta de $X_1, X_2, \ldots, X_n$ é conhecida, exceto por um parâmetro $\theta \in \Theta \subset \mathbb{R}^d$, $d \geq 1$. Θ será chamado de espaço paramétrico.

Admitamos que a função de distribuição conjunta F_θ de $X_1, X_2, \ldots, X_n$ pertença a uma família $\Gamma = \{ F_\theta ; \theta \in \Theta \}$, e $\psi: \Theta \longrightarrow \mathbb{R}^d$ seja uma função suave e bijetiva. O problema de estimação paramétrica consiste em estimar $\psi(\theta)$ por meio de funções $T(X_1, \ldots, X_n)$ que são variáveis aleatórias só dependentes das observações.

Vários métodos de estimação foram desenvolvidos ao longo dos anos. Um destes métodos ficou particularmente conhecido após os trabalhos de R. A. Fisher o qual foi denominado por Máxima Verossimilhança (ver Bickel & Doksum(1977)).

O estimador de máxima verossimilhança (EMV) consiste no seguinte (ver Bustos(1981)):

<u>Definição 1:</u> Seja $n \geq 1$ fixo; chama-se função de verossimilhança a $L: \Theta \times \mathbb{R}^n \longrightarrow \mathbb{R}$ definida por

$$L(\theta, x_1, \ldots, x_n) = f_\theta(x_1, \ldots, x_n)$$

onde $\theta \in \Theta$ e f_θ é a densidade de $F_\theta \in \Gamma$ em relação a uma certa medida μ σ-finita que domina Γ.

Para cada $\underset{\sim}{x} = (x_1, \ldots, x_n)' \in \mathbb{R}^n$, supõe-se que exista um $\hat{\theta}(\underset{\sim}{x}) \in \Theta$ tal que

$$L(\hat{\theta}(\underset{\sim}{x}), \underset{\sim}{x}) \geq L(\theta, \underset{\sim}{x}), \ \theta \in \Theta$$

Chama-se estimador de máxima verossimilhança (EMV) de $\psi(\theta)$ à função

$$T_n(X_1, \ldots, X_n) = \psi(\hat{\theta}(\underset{\sim}{X}))$$

e $\hat{\theta}(\underset{\sim}{X})$ é chamado o EMV de θ.

Dentre os procedimentos de estimação, estamos interessados numa classe em particular, que é conhecida como estimação por mínima distância entre funções de distribuições.

É oportuno aqui definir o que se pode entender por distância entre funções de distribuições.

Em matemática encontramos a definição de distância entre elementos de um conjunto abstrato que é extremamente satisfatória para as situações nas quais vamos incorrer.

Definição 2: Seja M um conjunto qualquer não vazio. Chama-se métrica ou distância entre pontos de M uma função $\delta: M \times M \to \mathbb{R}$ tal que:

(a) $\delta(A,B) \geq 0$, quaisquer que sejam $A,B \in M$;

(b) $\delta(A,B)=0 \Leftrightarrow A=B$;

(c) $\delta(A,B)=\delta(B,A)$, quaisquer que sejam $A,B \in M$;

(d) $\delta(A,C) \leq \delta(A,B) + \delta(B,C)$, quaisquer que sejam $A,B,C \in M$.

A estimação por Mínima Distância foi desenvolvida nos anos 50 por Wolfowitz, ver Wolfowitz(1953). Este método providencia estimativas consistentes para parâmetros desconhecidos, sob a suposição de amostragem independente de uma população fixa. O método pode ser descrito por:

Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma distribuição G,

$$G_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x),$$

a função distribuição empírica e $I(X_i \leq x)$ é o indicador do evento $[X_i \leq x]$.

Mais ainda, seja $\Gamma = \{F_\theta;\ \theta \in \Theta\}$, onde $\Theta \subset \mathbb{R}^d$, $d \geq 1$ que denotará a família de distribuições parametrizadas. Por questão de simplicidade, nesta seção vamos considerar somente as distribuições univariadas, embora o método possa ser estendido para dimensões maiores. Como na definição 2 denotaremos por $\delta(.,.)$ a medida de distância entre duas funções de distribuições.

Agora daremos algumas definições de distância entre funções de distribuições (ver Parr(1981)).

1.- Distancia ponderada de Smirnov-Kolmogorov

$$\delta_1(G_n,F_\theta) = \sup_{-\infty<z<\infty} |G_n(z) - F_\theta(z)| \sqrt{\eta(F_\theta(z))} \qquad (0.1.1)$$

quando $\eta(u)\equiv 1$, obtemos a estatística de Kolmogorov-Smirnov.

2.- Distancia ponderada de Cramer- von Mises

$$\delta_2(G_n,F_\theta) = \int_{-\infty}^{\infty} [G_n(z)-F_\theta(z)]^2\eta(F_\theta(z))dF_\theta(z) \qquad (0.1.2)$$

quando $\eta(u)\equiv 1$, obtemos a estatística de von-Mises.

3.- Distancia de Kuiper

$$\delta_3(G_n,F_\theta) = \sup_{-\infty<z<\infty} (G_n(z)-F_\theta(z)) - \inf_{-\infty<z<\infty} (G_n(z)-F_\theta(z)) \qquad (0.1.3)$$

4.- Distancia Qui-quadrado

$$-\infty = z_0 < z_1 < z_2 < \dots < z_{k-1} < z_k = \infty$$

$$\delta_4(G_n,F_\theta) = n\sum_{j=1}^{k} \frac{[(G_n(z_j)-G_n(z_{j-1})) - (F_\theta(z_j)-F_\theta(z_{j-1}))]^2}{F_\theta(z_j) - F_\theta(z_{j-1})} \qquad (0.1.4)$$

5.- Distancia baseada na funcao caracteristica empirica

$$\phi_n(t) = \int_{-\infty}^{\infty} e^{itz}dG_n(z); \; -\infty < t < \infty$$

Se pusermos $\phi(t,\theta) = \int_{-\infty}^{\infty} e^{itz}dF_\theta(z)$ então obtemos distâncias típicas baseadas nas funções características, tais como

$$\delta_{5,p}(G_n,F_\theta) = \int_{-\infty}^{\infty} |\phi_n(t)-\phi(t,\theta)|^p \, w(t)dt \qquad (0.1.5.1)$$

$$\delta_{5,\Lambda}(G_n,F_\theta) = \sup_t |\phi_n(t)-\phi(t,\theta)| \sqrt{w(t)} \qquad (0.1.5.2)$$

Neste trabalho estamos interessados numa distância em particular, a saber

6.- Distancia de Hellinger

Suponha que $\Gamma = \langle F_\theta;\ \theta \in \Theta \rangle$ seja dominada por uma certa medida μ, σ-finita. Então definimos a distância de Hellinger por:

$$\delta(G_n, F_\theta) = \left[\int_{-\infty}^{\infty} (f_\theta^{1/2}(z) - g_n^{1/2}(z))^2 d\mu(z) \right]^{1/2} \qquad (0.1.6)$$

onde f_θ e g_n são, respectivamente, as densidades de F_θ e G_n relativamente a μ.

7.- Distancia de Kulback-Leibler

$$\delta(F_\theta, G) = \int \ln(f_\theta(z)/g(z)).f_\theta(z)\ d\mu(z) \qquad (0.1.7)$$

onde f_θ e g são, respectivamente, as densidades de F_θ e G relativamente a μ

Os estimadores de Mínima Distância são obtidos minimizando-se a distância entre a distribuição empírica e a família Γ.

Os estimadores que cumprem tal objetivo são fortemente consistentes,isto é, se temos uma sequência $\langle T_n \rangle$ de estimadores de θ por mínima distância, então $T_n \longrightarrow \theta$ q.c.. Entretanto a classe de estimadores - MD (mínima distância) providencia muitos exemplos de não normalidade assintótica de algumas distâncias. Em particular, as distâncias do "tipo-sup" tais como δ_1 e δ_3 são não normais, enquanto que outras, sob condições de regularidade, são distribuídas normalmente, (ver Parr(1981)).

Podemos encontrar pelo menos dois fatos que tornam a estimação MD um método atrativo. Um deles é que este método é de fácil implementação, pois dado um conjunto de dados, um modelo paramétrico e uma distância entre as funções de distribuição resta apenas construir uma rotina de minimização para contruir o estimador.

O segundo fato (o mais importante) é que os estimadores MD leva em consideração quando $G \notin \Gamma$, isto é, quando o modelo paramétrico conjecturado é incorreto (ver Simpson(1987)). Nestes casos, os estimadores do tipo MD são consistentes para o valor de θ_o tal que $\inf_{\theta\in\Theta} \delta(G,F_\theta) = \delta(G,F_{\theta_o})$. Assim, a estimação por MD seleciona a melhor aproximação em Γ em relação a $\delta(.,.)$.

Quando uma parcela das observações se destaca por ter um comportamento surpreendentemente diferente das restantes (ver Barnett & Lewis(1982), este método tende a produzir estimadores robustos. Na verdade, quando $\delta(.,.)$ é a distância de Hellinger o estimador resultante é assintóticamente equivalente ao estimador de máxima verossimilhança, e mais, é robusto no sentido de dar pouco peso a observações que podem ser consideradas como "outliers",(Ver Simpson(1987)).

CAPITULO I

INFERENCIA BASEADA NA DISTANCIA DE HELLINGER

§ 1: Definicoes Basicas

A partir de agora, salvo menção em contrário, nossa definição de distância entre funções de distribuição será dada pela definição (0.1.6), ou seja, a distância de Hellinger.

Da definição (0.1.6) claramente vemos que,

$$\delta^2(G_n,F_\theta) = 2 - 2 \int_{\mathbb{R}} (\sqrt{f_\theta \cdot g_n}\,)\, d\mu \qquad (1.1.1)$$

e assim podemos definir uma quantidade que se mostrará extremamente útil no decorrer deste trabalho.

<u>Definição I.1:</u> Para quaisquer par de funções de distribuição F_1 e F_2 tal que F_1 e F_2 possuam densidades f_1 e f_2 respectivamente, em relação a alguma medida ν σ-finita. Chamaremos o número

$$\rho = \rho(F_1,F_2) = \int_{\mathbb{R}} \sqrt{f_1 \cdot f_2}\, d\nu$$

de **afinidade** entre F_1 e F_2 (ver Matusita(1951)).

<u>Lema I.1:</u> Sejam F_1 e F_2 funções de distribuições quaisquer. Então:

(i) $0 \leq \rho(F_1,F_2) \leq 1$

(ii) $\delta^2(F_1,F_2) = 2(1 - \rho(F_1,F_2)) \leq 2$

(iii) $\delta^2(F_1,F_2) \leq \int_{\mathbb{R}} |f_1-f_2|\, d\nu \leq 2\delta(F_1,F_2)$

(iv) Sejam (F_n) uma sequência de funções de distribuição, então:

$\delta(F_n,F_\theta) \longrightarrow 0$ quando $n \rightarrow \infty$ se, e somente se $|F_n(E)-F_\theta(E)| \longrightarrow 0$, quando $n \rightarrow \infty$, uniformemente em $E \subseteq \mathbb{R}$.

É imediato que $\delta(F_n,F_\theta) \rightarrow 0$, quando $n \rightarrow \infty$, implica que $\rho(F_n,F_\theta) \rightarrow 1$ e vice-versa

Prova do Lema I.1:

(i) Por construção $\rho(F_1,F_2) \geq 0$.

Agora suponha $\delta(F_1,F_2) = 0$, daí $F_1 \equiv F_2$ e portanto $\rho(F_1,F_2)=1$; se tivermos $\delta(F_1,F_2) > 0$ então $\delta^2(F_1,F_2) = 2(1-\rho(F_1,F_2)) > 0$, logo $\rho(F_1,F_2) < 1$.

(ii) De (i) temos $0 \leq \rho(F_1,F_2) \leq 1$, daí $0 \leq 2(1-\rho(F_1,F_2)) \leq 2$.

(iii) Primeiro vamos provar que $\delta^2(F_1,F_2) \leq \int_{\mathbb{R}} |f_1 - f_2|\, d\nu$.

$|\sqrt{f_1} - \sqrt{f_2}| \leq |\sqrt{f_1} + \sqrt{f_2}| \rightarrow (\sqrt{f_1} - \sqrt{f_2})^2 \leq |\sqrt{f_1} - \sqrt{f_2}|.|\sqrt{f_1} + \sqrt{f_2}|$

Daí, $(\sqrt{f_1} - \sqrt{f_2})^2 \leq |f_1 - f_2|$ e portanto,

$$\int_{\mathbb{R}} (\sqrt{f_1} - \sqrt{f_2})^2 \leq \int |f_1 - f_2|\, d\nu$$

Mas,

$$\int_{\mathbb{R}} |f_1 - f_2|\, d\nu = \int_{\mathbb{R}} |\sqrt{f_1} - \sqrt{f_2}|.|\sqrt{f_1} + \sqrt{f_2}|\, d\nu$$

$$\leq \left[\int_{\mathbb{R}} (\sqrt{f_1} - \sqrt{f_2})^2 d\nu . \int_{\mathbb{R}} (\sqrt{f_1} + \sqrt{f_2})^2 d\nu \right]^{1/2}$$

$$= \delta(F_1,F_2) \left[\int_{\mathbb{R}} f_1 d\nu + \int_{\mathbb{R}} f_2 d\nu + 2\int_{\mathbb{R}} \sqrt{f_1 f_2}\, d\nu \right]^{1/2}$$

$$= \delta(F_1,F_2) \left[2 + 2\rho(F_1,F_2) \right]^{1/2}$$

$$\leq \delta(F_1,F_2)\, [2 + 2]^{1/2} = 2.\delta(F_1,F_2).$$

a primeira desigualdade decorrendo da Desigualdade de Cauchy-Schwartz.

(iv) Seja $E \subseteq \mathbb{R}$, um conjunto mensurável qualquer tal que $F_i(E) = \int_E f_i\, d\nu$, $i=1,2$.

$|F_1(E) - F_2(E)| = |\int_E (f_1-f_2)d\nu| \leq |\int_{\mathbb{R}} (f_1-f_2)d\nu| \leq \int_{\mathbb{R}} |f_1-f_2|d\nu$

por (iii) temos

$$\int_{\mathbb{R}} |f_1-f_2|d\nu \leq 2\delta(F_1,F_2)$$

Daí,

$$|F_1(E) - F_2(E)| \leq 2\delta(F_1,F_2)$$

Logo, para todo $E \subseteq \mathbb{R}$

$$|F_n(E) - F_\theta(E)| \leq 2\delta(F_n,F_\theta)$$

Portanto, $\delta(F_n,F_\theta) \longrightarrow 0$, quando $n \longrightarrow \infty$, implica $|F_n(E) - F_\theta(E)| \longrightarrow 0$ uniformemente em E, pois $\delta(F_n,F_\theta)$ não depende de E.

Por outro lado, $|F_n(E) - F_\theta(E)| \longrightarrow 0$, quando $n \longrightarrow \infty$, significa que $\int_E |f_n - f_\theta| d\nu \longrightarrow 0$. Mas, como já vimos em (iii),

$$(\sqrt{f_n} - \sqrt{f_\theta})^2 \leq |f_n - f_\theta|$$

daí,

$$\int_E (\sqrt{f_n} - \sqrt{f_\theta})^2 d\nu \le \int_E |f_n - f_\theta| d\nu, \quad \text{qualquer que seja } E \subseteq \mathbb{R}.$$

Portanto,

$|F_n(E) - F_\theta(E)| \longrightarrow 0$, quando $n \to \infty$, implica

$$\int_E (\sqrt{f_n} - \sqrt{f_\theta})^2 d\nu \longrightarrow 0, \text{ qualquer que seja } E \subseteq \mathbb{R}$$

Assim, vale também para $E = \mathbb{R}$ e temos $\delta(F_n, F_\theta) \longrightarrow 0$, quando $n \to \infty$.

De posse destes resultados, podemos estudar relevantes problemas da inferência paramétrica. Para tornar mais compreensível o que pretendemos expor, vamos dividir estes problemas em dois grandes casos, a saber,

- Caso Discreto: Quando a distribuição F_θ é discreta e finita;
- Caso Geral: Quando F_θ é absolutamente contínua e quando F_θ tem suporte enumerável.

§2: Caso Discreto

Vamos supor que F seja uma função de distribuição discreta com probabilidades $p_1, p_2, \ldots, p_k$ para os eventos $E_1, E_2, \ldots, E_k$ que são observados em um experimento. A distribuição empírica baseada em n observações, isto é, $\left(\frac{n_1}{n}, \frac{n_2}{n}, \ldots, \frac{n_k}{n}\right)$ será denotada por S_n. Então por definição,

$$\delta^2(F,S_n) = \sum_{i=1}^{k}\left(\sqrt{\frac{n_i}{n}} - \sqrt{p_i}\right)^2 = 2\left(1 - \sum_{i=1}^{k}\sqrt{\frac{n_i}{n}}\cdot\sqrt{p_i}\right) \tag{1.2.1}$$

Ponha $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$ então pode-se mostrar que

$$\delta^2(F,S_n) \le \frac{\chi^2}{n} \tag{1.2.2}$$

pois para cada i=1,...,k temos:

$$\left(\sqrt{\frac{n_i}{n}} - \sqrt{p_i}\right) = \frac{n_i - np_i}{\sqrt{n}\,(\sqrt{n_i}+\sqrt{np_i})} = \left(\frac{n_i - np_i}{\sqrt{n}\,.\,\sqrt{np_i}}\right).\left(\frac{\sqrt{np_i}}{\sqrt{n_i} + \sqrt{np_i}}\right)$$

$$= \left(\frac{n_i - np_i}{n\sqrt{p_i}}\right).\left(\frac{\sqrt{n_i} + \sqrt{np_i}}{\sqrt{np_i}}\right)^{-1}$$

$$= \left(\frac{n_i - np_i}{n\sqrt{p_i}}\right).\left(1 + \sqrt{\frac{n_i}{np_i}}\right)^{-1}$$

Portanto,

$$\left(\sqrt{\frac{n_i}{n}} - \sqrt{p_i}\right)^2 = \frac{(n_i - np_i)^2}{n^2 p_i}\left[1 + \sqrt{\frac{n_i}{np_i}}\right]^{-2},$$

e daí decorre o resultado.

<u>Lema I.2:</u> Quando uma variável aleatória tem função de distribuição discreta F_θ, então para todo $\varepsilon > 0$, temos

$$\mathbb{P}\left\{\delta^2(S_n,F_\theta) < \varepsilon\left[\frac{k-1}{n}\right]\right\} \geq 1 - \frac{1}{\varepsilon}$$

Prova: Seja $p_1, p_2, \ldots p_{k'} > 0$ e $p_{k'+1} = \ldots = p_k = 0$ para $k' \leq k$, então

$$\delta^2(S_n,F_\theta) = \sum_{i=1}^{k}\left(\sqrt{\frac{n_i}{n}} - \sqrt{p_i}\right)^2 \leq \sum_{i=1}^{k'}\frac{1}{p_i}\left(\frac{n_i}{n} - p_i\right)^2 + \sum_{i=k'+1}^{k}\frac{n_i}{n}$$

Assim,

$$\mathbb{E}(\delta^2(S_n,F_\theta)) \leq \sum_{i=1}^{k'}\frac{1}{p_i}\mathbb{E}\left(\frac{n_i}{n} - p_i\right)^2 = \sum_{i=1}^{k'}\frac{1 - p_i}{n} = \frac{k' - 1}{n}$$

$$\leq \frac{k - 1}{n}$$

aqui $\mathbb{E}(.)$ denota a esperança com respeito a F_θ. Agora pela Desigualdade de Markov

$$\mathbb{P}\left\{\delta^2(S_n,F_\theta) < \mathbb{E}(\delta^2(S_n,F_\theta))\varepsilon\right\} \geq 1 - \frac{1}{\varepsilon},\ \forall\ \varepsilon > 0.$$

Agora, quando n é grande χ^2 é distribuída assintóticamente como uma qui-quadrado com k-1 graus de liberdade. Portanto da relação (1.2.2), obtemos

Teorema I.1: Quando uma variável aleatória tem função de distribuição F_θ discreta, com densidade $p_i > 0$, $\forall$ $i=1,\ldots,k$ e n é suficientemente grande, nós obtemos a seguinte igualdade assintótica

$$\mathbb{P}\left\{\delta^2(S_n,F_\theta) < \varepsilon^2\right\} \cong \mathbb{P}\left\{ \chi^2_{k-1} < 4n\varepsilon^2\right\}$$

$\forall$ $\varepsilon>0$ e χ^2_{k-1} é uma variável aleatória com distribuição qui-quadrado com k-1 graus de liberdade

Prova do Teorema I.1:

$$\delta^2(S_n,F_\theta) = \frac{1}{n}\sum_{i=1}^{k}\frac{(n_i-np_i)^2}{np_i}\left[1+\sqrt{\frac{n_i}{np_i}}\,\right]^{-2}$$

Daí,

$$\delta^2(S_n,F_\theta) \le \frac{\chi^2}{n},$$

onde

$$\chi^2 = \sum_{i=1}^{k}\frac{(n_i-np_i)^2}{np_i} \qquad (1.2.3)$$

Portanto,

$$\mathbb{P}\left\{\delta^2(S_n,F_\theta) < \varepsilon^2\right\} \ge \mathbb{P}\left\{\frac{\chi^2}{n} < \varepsilon^2\right\}$$

Agora, para n suficientemente grande

$$\delta^2(S_n,F_\theta) = \frac{1}{n}\sum_{i=1}^{k}\frac{(n_i-np_i)^2}{np_i} \cong \frac{\chi^2_{k-1}}{4n},$$

já que $\left[1+\sqrt{\frac{n_i}{np_i}}\,\right]^{-2} \cong 1/4.$

Assim, $\mathbb{P}\left\{\delta^2(S_n,F_\theta) < \varepsilon^2\right\} \cong \mathbb{P}\left\{\chi^2_{k-1} < 4n\varepsilon^2\right\}$, $\forall$ $\varepsilon > 0$ e n grande.

O Lema I.2 e o Teorema I.1 dizem que possuimos uma cota inferior para a probabilidade de que a distância entre a função de distribuição empírica e a função de distribuição discreta parametrizada da qual provém S_n não exceda um número positivo pré-fixado e que temos uma distribuição conhecida e tabelada para esta cota inferior. Estes resultados podem ser de muito interesse, por exemplo num teste de bondade de ajuste. Isto é, estamos interessados em decidir se uma variável aleatória tem função distribuição $F_\theta \in \Gamma = \{F_\theta$ discreta e finita; $\theta \in \Theta\}$ ou então se esta variável aleatória tem função distribuição F tal que o $\inf_{F\in\Gamma} \delta(F,F_\theta) \geq \epsilon > \eta$, para algum número real $\eta > 0$, pré-fixado. Para isto temos o seguinte resultado:

<u>Teorema I.2:</u> Sejam F_θ a função de distribuição discreta com probabilidades $p_1, p_2, \ldots, p_k$ para os eventos $E_1, E_2, \ldots, E_k$ que são observados em um experimento, S_n a função de distribuição empírica e η um número real positivo. Então temos:

$$\mathbb{P}\left\{ \inf_{F\in\Gamma} \delta(S_n,F) \leq \eta \right\} \geq 1 - \frac{k-1}{n\eta^2}; \text{ quando } F_\theta \in \Gamma$$

e

$$\mathbb{P}\left\{ \inf_{F\in\Gamma} \delta(S_n,F) > \eta \right\} \geq 1 - \frac{k-1}{n(\epsilon-\eta)^2}; \text{ quando } \inf_{F\in\Gamma} \delta(F,F_\theta) \geq \epsilon > \eta$$

Prova: Quando $F_\theta \in \Gamma$, obtemos claramente

$$\inf_{F\in\Gamma} \delta(F,S_n) \le \delta(F_\theta,S_n)$$

portanto de acordo com o Lema I.2, temos

$$\mathbb{P}\left\{ \inf_{F\in\Gamma} \delta(F,S_n) \le \eta \right\} \ge \mathbb{P}\left\{\delta(F_\theta,S_n) < \eta\right\} \ge 1 - \frac{k-1}{n\eta^2}$$

Quando $\inf_{F\in\Gamma} \delta(F_\theta,F) \ge \varepsilon > \eta$, obtemos

$$\inf_{F\in\Gamma} \delta(F,S_n) \ge \inf_{F\in\Gamma}\left\{\delta(F,F_\theta) - \delta(F_\theta,S_n)\right\} \ge \varepsilon - \delta(F_\theta,S_n),$$

portanto

$$\mathbb{P}\left\{ \inf_{F\in\Gamma} \delta(F,S_n) > \eta\right\} \ge \mathbb{P}\left\{\delta(S_n,F_\theta) \le \varepsilon - \eta\right\} \ge 1 - \frac{k-1}{n(\varepsilon-\eta)^2}$$

Assim temos a seguinte regra de decisão:
(ver Matusita(1955))

- Quando $\inf_{F\in\Gamma} \delta(S_n,F) \le \eta$, decida que $F_\theta \in \Gamma$;

- Quando $\inf_{F\in\Gamma} \delta(S_n,F) > \eta$, decida que $\inf_{F\in\Gamma} \delta(F_\theta,F) \ge \varepsilon > \eta$.

A função risco desta regra de decisão é uniformemente limitada por $\frac{k-1}{n\eta^2}$ ou $\frac{k-1}{n(\varepsilon-\eta)^2}$, dado que a função perda toma o valor 0 quando a decisão é correta e um valor menor que 1, quando a decisão está errada. Portanto temos que o risco tende a zero quando o tamanho da amostra cresce.

Desde que $\delta^2(F,G) = 2(1 - \rho(F,G))$ onde F e G são funções de distribuição, podemos definir a regra de decisão em função da afinidade ρ:

- Quando $\sup_{F\in\Gamma} \rho(F,S_n) \geq 1 - \frac{\eta^2}{2}$, decida que $F_\theta \in \Gamma$;

- Quando $\sup_{F\in\Gamma} \rho(F,S_n) < 1 - \frac{\eta^2}{2}$, decida $\inf_{F\in\Gamma} \delta(F,F_\theta) \geq \epsilon > \eta$.

Com esta mesma regra de decisão podemos estudar o problema de independência.

Sejam X,Y duas variáveis aleatórias e $p_1,p_2,\ldots p_k$, $q_1,q_2,\ldots,q_l$ as probabilidades dos eventos $A_1,A_2,\ldots,A_k$ e $B_1,B_2,\ldots B_l$ respectivamente. Denote por F_θ a distribuição conjunta de X e Y e por Γ_1 o conjunto de todas as distribuições sobre os eventos (A_i,B_j) tal que as probabilidades de (A_i,B_j) determinadas por elas, possam ser escritas na forma $p_i \cdot q_j$ com $\sum_{i=}^{k} p_i = \sum_{j=}^{l} q_j = 1$, com $p_i \geq 0$ e $q_j \geq 0$. Note que para aplicar a regra de decisão acima, temos somente que mostrar que $\sup_{F\in\Gamma_1} \rho(F,S_n)$ pode ser calculado.

Sejam F a função de distribuição do vetor (X,Y) cuja densidade pode ser escrita como $p_i . q_j$ e S_n a função de distribuição empírica cuja densidade tem a forma p_{ij}. Ponha $x_i=\sqrt{p_i}$, $y_j=\sqrt{q_j}$ e $a_{ij}=\sqrt{p_{ij}}$. Então temos

$$\rho(F,S_n) = \sum_{i=1}^{k} \sum_{j=1}^{l} \sqrt{p_i} . \sqrt{q_j} . \sqrt{p_{ij}} = \sum_{i,j} a_{ij} . x_i . y_j$$

Em notação matricial,

$$\underset{\sim}{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}, \qquad \underset{\sim}{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_l \end{pmatrix}, \qquad A = \begin{pmatrix} a_{11} & \cdots & a_{1l} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kl} \end{pmatrix}$$

Assim,

$$\rho(F,S_n) = \langle A'\underset{\sim}{x}, \underset{\sim}{y} \rangle \text{ ou } \langle \underset{\sim}{x}, A\underset{\sim}{y} \rangle$$

onde <.,.> é o produto interno usual.

Desde que

$$\| \underset{\sim}{x} \| = \left(\sum_{i=1}^{k} x_i^2 \right)^{1/2} = 1 \quad \text{e} \quad \| \underset{\sim}{y} \| = \left(\sum_{j=1}^{l} y_j^2 \right)^{1/2} = 1$$

é verdade que,

$$\langle A'\underset{\sim}{x}, \underset{\sim}{y} \rangle = \langle \underset{\sim}{x}, A\underset{\sim}{y} \rangle \le \| A'\underset{\sim}{x} \| \text{ ou } \| A\underset{\sim}{y} \|$$

A expressão $\langle A'\underset{\sim}{x}, \underset{\sim}{y} \rangle = \langle \underset{\sim}{x}, A\underset{\sim}{y} \rangle$ atinge seu máximo se, e somente se, a direção de $A'\underset{\sim}{x}$ coincide com a de $\underset{\sim}{x}$ (ou a de $\underset{\sim}{y}$ com $A\underset{\sim}{y}$). Isto decorre da Desigualdade de Cauchy-Schwartz.

Portanto encontramos a seguinte relação:

$$\max_{F \in \Gamma_1} \rho(F,S_n) = \max_{\substack{\|x\|=1 \\ \forall \tilde{i}, x_i \ge 0}} \| A' \underset{\sim}{x} \|$$

Agora A.A' é simétrica positiva semi-definida, então A.A' pode ser diagonalizada por uma matriz ortogonal V. Isto é,

$$V.A.A'.V' = \Lambda = \mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_k),$$

onde λ_i são os auto-valores, todos não-negativos e sua soma é 1. Mais ainda,

$$\sum_{i=1}^{k} \lambda_i = \mathrm{traço}(A.A') = \sum_{i,j} a_{ij}.a_{ji} = 1.$$

Seja λ_{max}, o maior valor dos autovalores $\lambda_1,\lambda_2,\ldots,\lambda_k$, então

$$\max_{\substack{\|x\|=1 \\ x_i \geq 0}} \| A'\underset{\sim}{x} \|^2 = \lambda_{max}$$

Consequentemente,

$$\max_{F\in\Gamma_1} \rho(F,S_n) = \sqrt{\lambda_{max}}$$

Assim o problema de decisão torna-se num certo sentido mais fácil, pois basta encontrar o maior auto-valor da matriz A'A, o que é computacionalmente favorável, já que existem softwares bastante eficientes para este tipo de problema. Na maioria dos casos (dependendo da estrutura de A'A), estes programas são mais rápidos do que aqueles que maximizam funções.

De maneira análoga podemos utilizar os resultados anteriores para o problema de duas amostras (ver Matusita(1955)).

Sejam X e Y duas variáveis aleatórias (não necessariamente independentes) que tenham funçõesde distribuição discreta F e G respectivamente cujas densidades existem e são $p_1,\dots p_k$ e $q_1,\dots,q_k$ sobre os eventos $E_1,\dots,E_k$, isto é, $p_i = \mathbb{P}(X \in E_i)$ e $q_i = \mathbb{P}(Y \in E_i)$, para $i=1,\dots k$. Mais ainda, sejam $(n_1,\dots,n_k)$, $\sum_{i=1}^{k} n_i = n$ e $(m_1,\dots,m_k)$, $\sum_{i=1}^{k} m_i = m$, observações sobre X e Y respectivamente.

Então queremos decidir a partir de dois conjuntos de observações $(n_1,\dots,n_k)$ e $(m_1,\dots,m_k)$ se $F = G$ ou se $\delta(F,G) \geq \epsilon > 0$, onde ϵ é um número real pré-fixado. Neste problema nós também estamos interessados na questão de F e G estarem ou não próximas uma da outra.

Denote S_n como a distribuição empírica $\left(\frac{n_1}{n} ,\dots, \frac{n_k}{n} \right)$ e S_m como a distribuição empírica $\left(\frac{m_1}{m} ,\dots, \frac{m_k}{m}\right)$. Então nós temos:

Teorema I.3: (Matusita(1955))

Sejam η, $\varepsilon \in \mathbb{R}$ tal que $0 < \eta < \varepsilon$.

(a) Quando $\delta(F,G) = 0$, tem-se:

$$\mathbb{P}\left\{ \delta(S_n, S_m) < \eta \right\} \geq 1 - \frac{k-1}{\eta^2} \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}} \right)^2$$

(b) Quando $\delta(F,G) \geq \varepsilon$, tem-se

$$\mathbb{P}\left\{ \delta(S_n, S_m) \geq \eta \right\} \geq 1 - \frac{k-1}{(\varepsilon-\eta)^2} \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}} \right)^2$$

Mais ainda, se X e Y são independentes, tem-se

(c) Quando $\delta(F,G) = 0$

$$\mathbb{P}\left\{ \delta(S_n, S_m) < \eta \right\} \geq 1 - \frac{4(k-1)}{\eta^2} \left(\frac{1}{n} + \frac{1}{m} \right) + \frac{16(k-1)^2}{\eta^4 . n . m}$$

(d) Quando $\delta(F,G) \geq \varepsilon$, tem-se

$$\mathbb{P}\left\{ \delta(S_n, S_m) \geq \eta \right\} \geq 1 - \frac{4(k-1)}{(\varepsilon-\eta)^2} \left(\frac{1}{n} + \frac{1}{m} \right) + \frac{16(k-1)^2}{(\varepsilon-\eta)^4 n . m}$$

Neste caso (a) é mais preciso do que (c) se, e somente se, $3(m+n) - 2\sqrt{m.n} > \dfrac{16(k-1)}{\eta^2}$ e (b) é mais preciso do que (d) se, e somente se, $3(m+n) - 2\sqrt{m.n} > \dfrac{16(k-1)}{(\varepsilon-\eta)^2}$.

Prova: Quando $\delta(F,G) = 0$, então

$$\delta(S_n,S_m) \le \delta(F,S_n) + \delta(S_m,G)$$

daí, usando a Desigualdade de Markov

$$\begin{aligned}
\mathbb{P}\left\{ \delta(S_n,S_m) < \eta \right\} &\ge \mathbb{P}\left\{ \delta(S_n,F) + \delta(G,S_m) < \eta \right\} \\
&\ge 1 - \frac{1}{\eta^2}\, \mathbb{E}\,(\delta(S_n,F) + \delta(G,S_m))^2 \\
&\ge 1 - \frac{1}{\eta^2}\left[\sqrt{\mathbb{E}(\delta(S_n,F)^2)} + \sqrt{\mathbb{E}(\delta(G,S_m)^2)} \right]^2 \\
&\ge 1 - \left(\frac{k-1}{\eta^2}\right)\cdot\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right)^2
\end{aligned}$$

e mais ainda, se X e Y são independentes

$$\begin{aligned}
\mathbb{P}\left\{ \delta(S_n,S_m) < \eta \right\} &\ge \mathbb{P}\left\{ \delta(S_n,F) < \frac{\eta}{2} \,,\, \delta(G,S_m) < \frac{\eta}{2} \right\} \\
&= \mathbb{P}\left\{ \delta(S_n,F) < \frac{\eta}{2} \right\}\cdot \mathbb{P}\left\{ \delta(G,S_m) < \frac{\eta}{2} \right\} \\
&= \left(1 - \frac{4(k-1)}{n.\eta^2} \right)\cdot\left(1 - \frac{4(k-1)}{m.\eta^2} \right) \\
&= 1 - \frac{4(k-1)}{\eta^2}\left(\frac{1}{n} + \frac{1}{m}\right) + \frac{16(k-1)^2}{\eta^4 m.n}
\end{aligned}$$

e quando $\delta(F,G) \ge \epsilon$ então

$$\begin{aligned}
\delta(S_n,S_m) &\ge \delta(F,G) - \delta(S_n,F) - \delta(G,S_m) \\
&\ge \epsilon - \delta(S_n,F) - \delta(G,S_m)
\end{aligned}$$

e analogamente, nós obtemos (b) e (d) (ver Matusita(1955)).

No caso em que sabemos que X e Y são independentes e F e G são tais que ${}_{(F)}\chi^2_{k-1}$ e ${}_{(G)}\chi^2_{k-1}$ como obtidos em (1.2.3) têm distribuição χ^2 com k-1 graus de liberdade, podemos fazer uso do seguinte resultado:

Quando $\delta(F,G) = 0$,

$$\mathbb{P}\left\{ \delta(S_n,S_m) < \eta \right\} \geq \mathbb{P}\left\{ \frac{1}{2n}\,{}_{(F)}\chi^2_{k-1} + \frac{1}{2m}\,{}_{(G)}\chi^2_{k-1} < \eta^2 \right\}$$

e

$$\mathbb{P}\left\{ \delta(S_n,S_m) < \eta \right\} \geq \mathbb{P}\left\{ {}_{(F)}\chi^2_{k-1} < n.\eta^2 \right\}.\ \mathbb{P}\left\{ {}_{(G)}\chi^2_{k-1} < m.\eta^2 \right\}$$

Quando $\delta(F,G) > \varepsilon$,

$$\mathbb{P}\left\{ \delta(S_n,S_m) \geq \eta \right\} \geq \mathbb{P}\left\{ \frac{1}{2n}\,{}_{(F)}\chi^2_{k-1} + \frac{1}{2m}\,{}_{(G)}\chi^2_{k-1} < (\varepsilon-\eta)^2 \right\}$$
$$\geq 1 - \frac{k-1}{2(\varepsilon-\eta)^2}\left(\frac{1}{n} + \frac{1}{m}\right)$$

e

$$\mathbb{P}\left\{ \delta(S_n,S_m) \geq \eta \right\} \geq \mathbb{P}\left\{ {}_{(F)}\chi^2_{k-1} \leq n.(\varepsilon-\eta)^2 \right\}.\mathbb{P}\left\{ {}_{(G)}\chi^2_{k-1} \leq m.(\varepsilon-\eta)^2 \right\}$$

§3: Caso Geral

3.1 - Quando F_θ é absolutamente contínua

Aqui estamos interessados, particularmente, no problema de observar n variáveis aleatórias $X_1,\dots,X_n$ independentes e identicamente distribuidas com densidade pertencente a uma família paramétrica $\Omega = \{ f_\theta: \theta \in \Theta \}$, $\Theta \subseteq \mathbb{R}^d$, $d \geq 1$. E procurar um procedimento que minimize a distância de Hellinger entre distribuições f_θ e $\hat{g}_n$,
onde $\hat{g}_n$ é um estimador não paramétrico adequado da densidade de X_i, $i=1,\dots,n$. Denotaremos este estimador por $\hat{\theta}_n^H(X_1,\dots,X_n)=\hat{\theta}_n^H$, o número real tal que:

$$\min_{\theta\in\Theta} \delta(F_\theta,G) = \delta(F_{\hat{\theta}_n^H},G) = \| \sqrt{f_{\hat{\theta}_n^H}} - \sqrt{\hat{g}_n} \|$$

onde G_n é a função de distribuição empírica e $\| \cdot \|$ é a norma no L^2, o espaço das funções de quadrado integrável.

Observe que o estimador $\hat{\theta}_n^H$ guarda uma forte relação heurística com o estimador de máxima verossimilhança (EMV) $\hat{\theta}_n$, quando $\hat{g}_n$ provém de alguma $f_{\theta_o} \in \Omega$. Pois com esta suposição e n suficientemente grande o EMV estaria próximo de θ_o e o estimador da densidade $\hat{g}_n$ estaria próximo de f_{θ_o}, sob o ponto de vista formal quando θ está suficientemente próximo de θ_o, verificamos que:

$$\max_{\theta\in\Theta} \sum_{i=1}^{n} \log f_\theta(x_i) = \sum_{i=1}^{n} \log f_{\theta^*}(x_i) \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \max_{\theta\in\Theta} \frac{1}{n} \sum_{i=1}^{n} \log f_\theta(x_i) = \frac{1}{n}\sum_{i=1}^{n} \log f_{\theta^*}(x_i) \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \max_{\theta\in\Theta} \int \log f_\theta(\underset{\sim}{x}) dG_n(\underset{\sim}{x}) = \int \log f_{\theta^*}(\underset{\sim}{x}) dG_n(\underset{\sim}{x}) \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \max_{\theta\in\Theta} \int \log \left[\frac{f_\theta(\underset{\sim}{x})}{\hat{g}_n(\underset{\sim}{x})} \right] dG_n(\underset{\sim}{x}) = \int \log \left[\frac{f_{\theta^*}(\underset{\sim}{x})}{\hat{g}_n(\underset{\sim}{x})} \right] dG_n(\underset{\sim}{x})$$

$$\Leftrightarrow \quad \max_{\theta\in\Theta} \int \log \left[\frac{f_\theta(\underset{\sim}{x})}{\hat{g}_n(\underset{\sim}{x})} \right] \hat{g}_n(\underset{\sim}{x})\, d\underset{\sim}{x} = \int \log \left[\frac{f_{\theta^*}(\underset{\sim}{x})}{\hat{g}_n(\underset{\sim}{x})} \right] \hat{g}_n(\underset{\sim}{x})\, d\underset{\sim}{x}$$

Mas,

$$\int \log \left[\frac{f_\theta(\underset{\sim}{x})}{\hat{g}_n(\underset{\sim}{x})} \right] \hat{g}_n(\underset{\sim}{x})\, d\underset{\sim}{x} = 2 \int \log \left[1 + \left(\frac{f_\theta^{1/2}(\underset{\sim}{x})}{\hat{g}_n^{1/2}(\underset{\sim}{x})} - 1 \right) \right] \hat{g}_n(\underset{\sim}{x})\, d\underset{\sim}{x}$$

expandindo em Série de Taylor,

$$\log \left[1 + \left(\frac{f_\theta^{1/2}(\underset{\sim}{x})}{\hat{g}_n^{1/2}(\underset{\sim}{x})} - 1 \right) \right] \cong \left(\frac{f_\theta^{1/2}(\underset{\sim}{x})}{\hat{g}_n^{1/2}(\underset{\sim}{x})} - 1 \right) - \frac{1}{2} \left(\frac{f_\theta^{1/2}(\underset{\sim}{x})}{\hat{g}_n^{1/2}(\underset{\sim}{x})} - 1 \right)^2$$

Obtemos então

$$\int \log \left[\frac{f_\theta(\underset{\sim}{x})}{\hat{g}_n(\underset{\sim}{x})} \right] \hat{g}_n(\underset{\sim}{x})\, d\underset{\sim}{x} \cong$$

$$\cong 2\int \left[\left(\frac{f_\theta^{1/2}(\underset{\sim}{x})}{\hat{g}_n^{1/2}(\underset{\sim}{x})} - 1 \right) - \frac{1}{2} \left(\frac{f_\theta^{1/2}(\underset{\sim}{x})}{\hat{g}_n^{1/2}(\underset{\sim}{x})} - 1 \right)^2 \right] \hat{g}_n(\underset{\sim}{x})\, d\underset{\sim}{x}$$

Observe que,

$$2\left[\left(\frac{f_\theta^{1/2}}{\hat{g}_n^{1/2}} - 1 \right) \hat{g}_n - \frac{1}{2} \left(\frac{f_\theta^{1/2}}{\hat{g}_n^{1/2}} - 1 \right)^2 \hat{g}_n \right] =$$

$$= 2 \left(f_\theta^{1/2} - \hat{g}_n^{1/2} \right) \hat{g}_n^{1/2} - \left(f_\theta^{1/2} - \hat{g}_n^{1/2} \right)^2$$

$$= 2\, \hat{g}_n^{1/2} . f_\theta^{1/2} - 2\, \hat{g}_n^{1/2} = f_\theta^{1/2} + 2\, f_\theta^{1/2} . g_n^{1/2} - \hat{g}_n$$

$$= 4\, \hat{g}_n^{1/2} . f_\theta^{1/2} - 3\, \hat{g}_n - f_\theta$$

Portanto,

$$\int \log \left[\frac{f_\theta(\underset{\sim}{x})}{\hat{g}_n(\underset{\sim}{x})} \right] \hat{g}_n(\underset{\sim}{x})\, d\underset{\sim}{x} \cong$$

$$\cong \int 4\, \hat{g}_n^{1/2}(\underset{\sim}{x}) - f_\theta^{1/2}(\underset{\sim}{x}) - 3\, \hat{g}_n(\underset{\sim}{x}) - f_\theta(\underset{\sim}{x})\, d\underset{\sim}{x}$$

$$= -2\, \delta^2(F_\theta, G_n)$$

Assim, é razoável supor que o estimador $\hat{\theta}_n^H$ será assintoticamente eficiente sob f_θ.

Podemos pensar $\hat{\theta}_n^H$ como um valor em $\hat{g}_n$ de um funcional T. Seja $\mathcal{F}$ a classe de todas as densidades com respeito à medida de Lebesgue na reta. O funcional T do tipo MDH (mínima distância de Hellinger) é aquele que está definido em $\mathcal{F}$ e satisfaz:

Para toda $g \in \mathcal{F}$,

$$\delta(F_{T(g)}, G) = \| f_{T(g)}^{1/2} - g^{1/2} \| = \min_{t \in \Theta} \| f_t^{1/2} - g^{1/2} \| \qquad (1.3.0)$$

O teorema a seguir garante a existência e a continuidade do estimador T do tipo MHD. Isto é, existe um funcional T(g) que satisfaz (1.3.1) e T(g) é consistente para $\theta \in \Theta$.

<u>Teorema I.4:</u> (Beran (1977))

Suponha que $\Theta \subseteq \mathbb{R}^d$ seja compacto e $d \geq 1$. Admita que $\theta_1 \neq \theta_2 \Rightarrow f_{\theta_1} \neq f_{\theta_2}$ num conjunto que tenha medida de Lebesgue positiva e para quase todo x, $f_\theta(x)$ é contínua em θ. Então:

(i) Para toda $g \in$, existe $T(g) \in \Theta$ satisfazendo (1.3.0);

(ii) Se T(g) é único, o funcional T é contínuo em g na topologia induzida pela métrica de Hellinger;

(iii) $T(f_\theta) = \theta$ é único para todo $\theta \in \Theta$.

Prova: Veja a belíssima demonstração em Beran(1977).

Neste mesmo artigo de 1977, Beran faz as seguintes afirmações:

(i) O teorema I.4 é também útil para famílias paramétricas $\{ f_\theta : \theta \in \Theta \}$, onde Θ não é compacto mas pode ser imerso dentro de um conjunto compacto.

(ii) Em problemas de bondade de ajuste, um "plot" residual de $f^{1/2}_{\hat{\theta}^H_n}(x) - \hat{g}^{1/2}_n(x)$, $\hat{\theta}^H_n$ sendo o estimador tipo MHD é um bom ponto de partida para considerar questões como identificar observações divergentes do grosso dos dados para investigações posteriores. Verificar se a família paramétrica é plausível para explicar o conjunto de observações.

(iii) Ainda sobre o problema de bondade de ajuste, a estatística $\delta^2(F_{\hat{\theta}^H_n}, G_n) = \| f^{1/2}_{\hat{\theta}^H_n} - \hat{g}^{1/2}_n \|^2$ parece particularmente adequada, desde que ela estima a distância de Hellinger ao quadrado entre a atual densidade $\hat{g}_n$ e a densidade mais próxima na família paramétrica $\{ f_\theta : \theta \in \Theta \}$.

(iv) É importante notar que a estatística $\| f^{1/2}_{\hat{\theta}^H_n} - \hat{g}^{1/2}_n \|^2$ não é muito afetada no ajuste na presença de poucos "outliers", (ver secção 4, Beran(1977)). É justamente esta insensibilidade que torna a estatística útil em decidir se o grosso dos dados pode ser razoavelmente ajustado por uma f_θ.

3.2 - Quando F_θ é discreta com suporte enumerável

Considere o esquema geral de estimação numa classe de distribuições paramétricas $\Gamma = \{ F_\theta : \theta \in \Theta \}$. Assuma $\Theta \subseteq \mathbb{R}^d$, $d \geq 1$ e que Γ é dominada por uma certa medida μ. Denote $\| \cdot \|$ a norma no L^2.

Seja f_θ a densidade de F_θ em relação a uma medida μ σ-finita e g_n um estimador não-paramétrico da densidade baseado numa amostra aleatória $X_1, \ldots, X_n$.

Se $g_n(x) \geq 0$ e $\sum_{x=0}^{\infty} g_n(x) = 1$, então facilmente obtemos,

$$\delta(F_\theta, G_n) = \| g_n^{1/2} - f_\theta^{1/2} \|^2 = 2 - 2 \sum g_n^{1/2} . f_\theta^{1/2} \qquad (1.3.1)$$

Para dados de contagem, a função de densidade empírica g_n é dada por

$$g_n(x) = \frac{N_x}{n} \, , \; x=0,1,2,\ldots \qquad (1.3.2)$$

onde $N_x = \sum_{i=1}^{n} I(X_i = x)$ e $I(X_i = x)$ é o indicador do evento $[X_i=x]$, isto é, N_x é a freqüência de x entre $X_1, \ldots, X_n$.

Portanto, o estimador de θ do tipo MDH que minimiza a distância de Hellinger, maximiza

$$\rho_{n,\theta} = \sum_{x=0}^{\infty} g_n^{1/2}(x) . f_\theta^{1/2}(x) \qquad (1.3.3)$$

Se f_θ é diferenciável em θ, maximizar (1.3.3) é equivalente a encontrar um zero da seguinte equação de estimação:

$$\rho_{n,\theta}^{-1} \cdot \sum_{x=0}^{\infty} g_n^{1/2}(x) . f_\theta^{1/2}(x) . l_\theta(x) = 0 \qquad (1.3.4)$$

onde $l_\theta(x) = \frac{\partial}{\partial\theta} \ln f_\theta(x)$.

Isto ocorre porque queremos maximizar $\rho_{n,\theta}$ em Θ. Daí,

$$\frac{\partial}{\partial\theta}\rho_{n,\theta} = \frac{\partial}{\partial\theta}\sum_{x=0}^{\infty} g_n^{1/2}(x) . f_\theta^{1/2}(x) = \frac{1}{2}\sum_{x=0}^{\infty} g_n^{1/2}(x) . f_\theta^{1/2-1}(x) \frac{\partial}{\partial\theta} f_\theta(x)$$

$$= \frac{1}{2}\sum_{x=0}^{\infty}\sqrt{g_n(x) f_\theta(x)} \; . \; l_\theta(x)$$

Mas,

$$\frac{1}{2}\sum_{x=0}^{\infty}\sqrt{g_n(x) . f_\theta(x)} . l_\theta(x) = 0, \text{ quando } \theta = \hat{\theta}_n = \max_{\theta\in\Theta} \rho_{n,\theta}$$

Logo,

$$\rho_{n,\theta}^{-1} \cdot \sum_{x=0}^{\infty}\sqrt{g_n(x) . f_\theta(x)} \; . \; l_\theta(x) = 0, \text{ quando } \theta = \hat{\theta}_n.$$

Em oposição a (1.3.4), temos a equação de estimação de máxima verossimilhança:

$$\sum_{x=0}^{\infty} g_n(x) . l_\theta(x) = 0 \qquad (1.3.5)$$

Obtemos (1.3.5) da seguinte forma:

Sejam $X_1,\dots,X_n$ uma amostra aleatória de uma distribuição F_θ, x_i um dos possíveis valores de cada X_i, $i=1,\dots,n$ e $x_i \in \{0,1,2,\dots\}$.

Defina $L(\theta,x_1,\dots,x_n) = \prod_{i=1}^{n} f_\theta(x_i)$. Daí,

$$L(\theta,x_1,\dots,x_n) = [f_\theta(0)]^{N_0}\dots\dots[f_\theta(x)]^{N_x}\dots$$

$$= \prod_{x=0}^{\infty} [f_\theta(x)]^{N_x}$$

Logo,

$$\ln L(\theta,x_1,\dots x_n) = \sum_{x=0}^{\infty} N_x \ln f_\theta(x)$$

Daí,

$\frac{\partial}{\partial\theta} \ln L(\theta,x_1,\dots x_n) = \sum_{x=0}^{\infty} N_x \frac{\partial}{\partial\theta} f_\theta(x) = 0$, quando $\theta = \hat{\theta}$, é o EMV.

$$\frac{\partial}{\partial\theta} \ln L(\theta,x_1,\dots,x_n) = \sum_{x=0}^{\infty} \frac{N_x}{n} \cdot l_\theta(x) = \sum_{x=0}^{\infty} g_n(x) = 0$$

quando $\theta = \hat{\theta}$.

Observe que quando g_n provém de uma distribuição F_θ então as equações (1.3.4) e (1.3.5) concordam no limite, quando $n \longrightarrow \infty$. Porém, quando g_n provém de uma $F_{\theta^*} \neq F_\theta$, a equação (1.3.4) leva isto em consideração, muito mais que a equação (1.3.5), pois a esperança de l_θ é tomada em relação à $\rho_{n,\theta}^{-1} \cdot \sqrt{g_n \cdot f_\theta}$ ao invés de g_n. Pode-se ver isto da seguinte forma:

Seja $f_\theta*$ a densidade de $F_\theta*$, (1.3.4) e (1.3.5) podem ser pensadas como esperanças de uma v.a., $l_\theta(Z)$ tal que em (1.3.4) $\mathbb{P}(\ Z = x\ ) = \rho^{-1}_{n,\theta}\ .\ \sqrt{g_n(x).f_\theta(x)}$, daí,

$$\mathbb{E}(\ l_\theta(Z)\ ) = \sum_{x=0}^{\infty} \rho^{-1}_{n,\theta}\ .\ \sqrt{g_n(x).f_\theta(x)}.\ l_\theta(x)$$

e, em (1.3.5), $\mathbb{P}\ (\ Z = x) = g_n(x)$, logo

$$\mathbb{E}(\ l_\theta(Z)\ ) = \sum_{x=0}^{\infty} g_n(x).l_\theta(x)$$

Agora suponha que cada X_i tenha densidade $f_\theta* \neq f_\theta$, para todo $i=1,\ldots,n$. Então, claramente, (1.3.5) leva em consideração somente as informações provenientes de $f_\theta*$, enquanto que (1.3.4) considera também as informações provenientes de f_θ.

Mais ainda, se por exemplo F_θ tem informação de Fisher finita, então $f_\theta^{1/2}(x).|l_\theta(x)| \longrightarrow 0$, quando $x \rightarrow \infty$. Isto significa que uma contagem (observação) pouco provável tem pequeno impacto no estimador por MDH. Pois neste caso:

$$\mathbb{E}(l_\theta(\underset{\sim}{X})) < +\infty \quad \Rightarrow \quad \mathbb{E}(|l_\theta(\underset{\sim}{X})|) < +\infty \quad \Leftrightarrow \quad \sum |l_\theta(x)|.f_\theta(x) < +\infty$$

$\Rightarrow |l_\theta(x)|f_\theta(x) \longrightarrow 0$, quando $x \rightarrow \infty$ e μ é uma medida dominante em

$\Gamma = \{\ F_\theta\text{:}\ \theta \in \Theta\ \}$.

Portanto, $f_\theta^{1/2}(x).|l_\theta(x)| \longrightarrow 0$, quando $x \rightarrow \infty$.

3.2.1 - Propriedades Assintóticas

Estabelecer a distribuição assintótica do estimador MDH é importante para se fazer comparações teóricas com outros estimadores e para tornar válidas as inferências aproximadas.

Primeiro é necessário discutir a consistência do estimador MDH, isto é, a continuidade de MDH na forma funcional dada por Beran(1977).

Nesta seção, vamos utilizar uma notação que será mais adequada para o entendimento dos conceitos que aqui serão expostos.

Ponha,

$$H(\theta,G) = \delta^2(F_\theta,G) = \| f_\theta^{1/2} - g^{1/2} \|^2$$

onde g é a densidade de G (ou a parte absolutamente contínua de G relativa à família Γ).

O funcional T do tipo MDH é aquele que resolve

$$H(T(G),G) = \min_{t\in\Theta} H(t,G), \text{ se existe solução}$$

No teorema I.4 provou-se a existência e a continuidade de T no caso contínuo, para Θ compacto. Mostrou-se também que o resultado se aplica quando Θ está imerso em um conjunto compacto $\overline{\Theta}$ e que $H(.,G)$ é contínuo em $\overline{\Theta}$. Porém tais imersões podem ser complicadas em situações multi-paramétricas e a distância de Hellinger nem sempre se estende continuamente.

Veja por exemplo o caso da Binomial Negativa bi-paramétrica descrita por Collings e Margolin(1985) como

$$f_\theta(x) = \frac{\Gamma(x+c^{-1})}{x!\,\Gamma(c^{-1})} \left(\frac{cm}{1+cm} \right)^x \left(\frac{1}{1+cm} \right)^{c^{-1}}, \; x=0,1,2,\ldots \quad (1.3.6)$$

onde $\theta = (m,c)$, $0 < m < +\infty$ e $0 \le c < +\infty$.

Note que $c \to 0$, f_θ corresponde a densidade de Poisson com média m, (ver Simpson(1987)).

Agora quando $m \to \infty$ com $c \to 0$, $H(\theta,G) \longrightarrow 2$, para G fixa. Isto porque:

$$c \to 0 \Rightarrow f_\theta(x) \longrightarrow \frac{e^{-m}\, m^x}{x!}, \quad x=0,1,\ldots\ ,$$

Daí,

$$\| f_\theta^{1/2} - g^{1/2} \|^2 = 2(1 - \sum_{x=0}^{\infty} \sqrt{g(x).f_\theta(x)})$$

$$= 2\left[1 - \sum_{x=0}^{\infty} \sqrt{g(x)} \ . \ \frac{e^{-m/2}\, m^{x/2}}{\sqrt{x!}} \right]$$

Mas,

$$\sum_{x=0}^{\infty} \sqrt{g(x)} \ . \ \frac{e^{-m/2}\, m^{x/2}}{\sqrt{x!}} \longrightarrow 0, \text{ quando } m \to \infty$$

Por outro lado, pode-se mostrar que se $f_\theta(0) \to 1$, quando $m \to \infty$ com $m.c=k$ (constante), então $H(\theta,G) \longrightarrow 2 - 2.g^{1/2}(0)$. Isto porque se $f_\theta(0) \to 1$ quando $m \to \infty$ com $m.c=k$, então

$$1 - f_\theta(0) \longrightarrow 0 \Rightarrow \sum_{x=1}^{\infty} f_\theta(x) \longrightarrow 0 \overset{f_\theta>0}{\Rightarrow} f_\theta(x) \longrightarrow 0, \ \forall\ x>0.$$

$$\text{Agora, } \rho_{n,\theta} = \sum_{x=0}^{\infty} \sqrt{g(x).f_\theta(x)} = \sqrt{g(0).f_\theta(0)} + \sum_{x=1}^{\infty} \sqrt{g(x).f_\theta(x)} \Rightarrow$$

$$\Rightarrow \lim_{\substack{m\to\infty \\ mc=k}} \rho_{n,\theta} = \sqrt{g(0)} + \lim_{\substack{m\to\infty \\ mc=k}} \sum_{x=1}^{\infty} \sqrt{g(x).f_\theta(x)}$$

Mas, $\forall\ x>0$, $f_\theta(x) \le 1$, daí $|f_\theta(x).g(x)| \le g(x)$ e $f_\theta(x) \longrightarrow 0$.

Portanto, aplicando o Teorema da Convergência Dominada e usando o fato que a função $\sqrt{\ }$ é contínua, temos

$$\lim_{\substack{m \to \infty \\ mc = k}} \rho_{n,\theta} = \sqrt{g(0)}$$

Logo,

$$H(\theta, G) \longrightarrow 2 - 2\, g^{1/2}(0).$$

Assim $H(\cdot,\cdot G)$ não se estende continuamente aos pontos limites de Θ e a compactificação de Θ via sua aplicação conforme na esfera (ver Apostol(1957),pag.11) não satisfaz a afirmação (i) feita por Beran(1977) (veja seção 3.1).

O teorema de Simpson(1987) estende a existência e continuidade de Beran.

Seja $\mathcal{G}$ a classe das distribuições G para as quais

$$\inf_{t \in \Theta - \mathbb{C}} H(t,G) > H(\theta^{*},G) \text{ para algum compacto } \mathbb{C} \subset \Theta \text{ e } \theta^{*} \in \mathbb{C}$$

Se Θ é compacto, tome $\mathbb{C} = \Theta$.

$_{\wedge}$ inclui qualquer distribuição não singular com respeito a Γ. Pois se existisse $G_o \in {}_{\wedge}$ tal que

$$G_o(E^c) = F_\theta(E) = 0,\ F_\theta \in \Gamma \text{ com } E \subset \mathbb{R}, \text{ mensurável}$$

Então, $H(\theta, G_o) = 2$, $\forall\, \theta \in \Theta$, daí,

$$2 = \inf_{t \in \Theta - \mathbb{C}} H(t, G_o) > H(\theta^{*}, G_o) = 2 \quad \text{(contradição!)}$$

pela definição da classe $\mathcal{G}$.

<u>Teorema I.5:</u> (Simpson(1987))

Suponha $f_\theta(x)$ contínua em θ para cada x. Então para cada $G \in \mathcal{G}$,

(a) T(G) existe;

(b) Se T(G) é único, então $\| g_n^{1/2} - g^{1/2} \| \longrightarrow 0$, implica que $T(G_n) \longrightarrow T(G)$ quando $n \to \infty$.

Prova:

(a) T(G) existe

Caso 1: Θ é compacto, então tome $\Theta = \mathbb{C}$, daí por H(.,G) ser contínua, temos que:

$\exists\ t^* \in \mathbb{C}=\Theta;\quad \min_{t\in\Theta} H(t,G) = H(t^*,G),\quad$ daí $\quad T(G) = t^*$

Caso 2: Θ não é compacto, mas sabemos que existe $\mathbb{C}$ compacto e $\forall\ \theta^* \in \mathbb{C}$ temos $\inf_{t\in\Theta-\mathbb{C}} H(t,G) > H(\theta^*,G)$ e como $\mathbb{C}$ é compacto existe $t^* \in \mathbb{C}$ tal que $\min_{t\in\mathbb{C}} H(t,G) = H(t^*,G)$ e portanto $\min_{t\in\Theta} H(t,G) = H(t^*,G)$, daí $T(G) = t^*$.

(b) Se T(G) é único, então $\| g_n^{1/2} - g^{1/2} \| \longrightarrow 0$ implica em $T(G_n) \longrightarrow T(G)$.

Sejam $h_n(t) = \| f_t^{1/2} - g_n^{1/2} \|$ e $h(t) = \| f_t^{1/2} - g^{1/2} \|$

$\| g_n^{1/2} - g^{1/2} \| \longrightarrow 0 \Rightarrow \sup_t | h_n(t) - h(t) | \longrightarrow 0$, pois

$$|h_n(t) - h(t)| = \big|\| f_t^{1/2} - g_n^{1/2} \| - \| f_t^{1/2} - g^{1/2} \|\big| \le \| g_n^{1/2} - g^{1/2} \|$$

Daí, $\sup_t |h_n(t) - h(t)| \longrightarrow 0 \Rightarrow |h_n(t) - h(t)| \longrightarrow 0 \Rightarrow$

$\Rightarrow h_n(t) \longrightarrow h(t) \Rightarrow h_n^2(t) \longrightarrow h^2(t) \Rightarrow |h_n^2(t) - h^2(t)| \longrightarrow 0$

$\Rightarrow \qquad |H(t,G_n) - H(t,G)| \longrightarrow 0,\ \forall\ t \in \Theta \qquad \Rightarrow$

Como $G \in \mathcal{G}$ segue que existe $\mathbb{C}$ compacto, $\mathbb{C} \subset \Theta$ tal que

$$\inf_{t\in\Theta-\mathbb{C}} H(t,G) > H(\theta^*,G),\ \theta^* \in \mathbb{C}$$

Daí, $|H(\theta^*,G_n) - H(\theta^* G)| \longrightarrow 0,\ \forall\ \theta^* \in \mathbb{C}$

Mas, $\sup_t |h_n(t) - h(t)| \longrightarrow 0 \quad \Rightarrow$

$$\Rightarrow \quad |\inf_{t\in\Theta-\mathbb{C}} H(t,G_n) - \inf_{t\in\Theta-\mathbb{C}} H(t,G)| \longrightarrow 0$$

porque:

i) Suponha que $\inf_{t\in\Theta-\mathbb{C}} h_n(t) \geq \inf_{t\in\Theta-\mathbb{C}} h(t)$, daí:

$$\forall\ \delta_o>0,\ \exists\ t_o(\delta_o) \in \Theta - \mathbb{C};\ \inf_{t\in\Theta-\mathbb{C}} h(t) \geq h(t_o)-\delta_o$$

$$h_n(t_o) \geq \inf_{t\in\Theta-\mathbb{C}} h_n(t) \geq \inf_{t\in\Theta-\mathbb{C}} h(t) \geq h(t_o)-\delta_o$$

$$\inf_{t\in\Theta-\mathbb{C}} h_n(t) - \inf_{t\in\Theta-\mathbb{C}} h(t) \leq h_n(t_o)-h(t_o)+\delta_o \leq \sup_t |h_n(t) - h(t)|+\delta_o$$

ii) Suponha que $\inf_{t\in\Theta-\mathbb{C}} h_n(t) \leq \inf_{t\in\Theta-\mathbb{C}} h(t)$, daí:

$$\forall\ \delta_1>0,\ \exists\ t_1(\delta_1) \in \Theta - \mathbb{C};\ \inf_{t\in\Theta-\mathbb{C}} h_n(t) \geq h_n(t_1)-\delta_1$$

$$h(t_1) \geq \inf_{t\in\Theta-\mathbb{C}} h(t) \geq \inf_{t\in\Theta-\mathbb{C}} h_n(t) \geq h_n(t_1)-\delta_1$$

$$\inf_{t\in\Theta-\mathbb{C}} h(t) - \inf_{t\in\Theta-\mathbb{C}} h_n(t) \leq h(t_1)-h_n(t_1)+\delta_1 \leq \sup_t |h_n(t) - h(t)|+\delta_1$$

Portanto,

$$\left|\inf_{t\in\Theta-\mathbb{C}} h_n(t) - \inf_{t\in\Theta-\mathbb{C}} h(t)\right| \le \sup_t |h_n(t) - h(t)| + \max(\delta_0, \delta_1)$$

e como δ_0 e δ_1 são quaisquer, temos:

$$\left|\inf_{t\in\Theta-\mathbb{C}} h(t) - \inf_{t\in\Theta-\mathbb{C}} h_n(t)\right| \le \sup_t |h_n(t) - h(t)|,$$

e quando $\sup_t |h_n(t) - h(t)| \longrightarrow 0$, obtemos

$$\left|\inf_{t\in\Theta-\mathbb{C}} h_n(t) - \inf_{t\in\Theta-\mathbb{C}} h(t)\right| \longrightarrow 0$$

Porem,

$$\left|\inf_{t\in\Theta-\mathbb{C}} h_n(t) - \inf_{t\in\Theta-\mathbb{C}} h(t)\right| \longrightarrow 0 \;\Rightarrow\; \left|\inf_{t\in\Theta-\mathbb{C}} h_n^2(t) - \inf_{t\in\Theta-\mathbb{C}} h^2(t)\right| \longrightarrow 0$$

Logo,

$$\left|\inf_{t\in\Theta-\mathbb{C}} H(t,G_n) - \inf_{t\in\Theta-\mathbb{C}} H(t,G)\right| \longrightarrow 0$$

Entao temos,

$$\forall\, \varepsilon>0,\ \exists\, n_0;\ \forall\, n>n_0 \;\Rightarrow\; \left|\inf_{t\in\Theta-\mathbb{C}} H(t,G_n) - \inf_{t\in\Theta-\mathbb{C}} H(t,G)\right| < \varepsilon/2 \text{ e}$$

$$\exists\, n_1;\ \forall\, n>n_1 \;\Rightarrow\; |H(\theta^*,G_n) - H(\theta^*,G)| < \varepsilon/2$$

Tomando $\varepsilon = \inf_{t\in\Theta-\mathbb{C}} H(t,G) - H(\theta^*,G)|$, $\theta^*\in\mathbb{C}$ e $\max(n_0,n_1) = n_2$ temos,

$$\forall\, n>n_2 \qquad H(\theta^*,G) - \varepsilon/2 < H(\theta^*,G_n) < H(\theta^*,G) + \varepsilon/2$$

$$\inf_{t\in\Theta-\mathbb{C}} H(t,G) - \varepsilon/2 < \inf_{t\in\Theta-\mathbb{C}} H(t,G_n) < \inf_{t\in\Theta-\mathbb{C}} H(t,G) + \varepsilon/2$$

Consequentemente,

$$\inf_{t\in\Theta-\mathbb{C}} H(t,G_n) - H(\theta^*,G_n) > \inf_{t\in\Theta-\mathbb{C}} H(t,G) - \varepsilon/2 - H(\theta^*,G) - \varepsilon/2 = 0$$

e temos

$$\inf_{t\in\Theta-\mathbb{C}} H(t,G_n) > H(\theta^*,G_n)$$

para todo $n > n_2$.

Portanto, $G_n \in \mathfrak{F}$ eventualmente, então existe $T(G_n)$ ε $\mathbb{C}$ tal que:

$$H(T(G_n),G_n) = \min_{\theta^* \in \mathbb{C}} H(\theta^*,G_n)$$

Seja $T(G) = \theta$, único e $T(G_n) = \theta_n$ tal que

$$H(T(G_n),G_n) = \min_{t\in\mathbb{C}} H(t,G_n) \quad e \quad H(T(G),G) = \min_{t\in\mathbb{C}} H(t,G)$$

De maneira totalmente análoga ao feito anteriormente temos que

$$\| g_n^{1/2} - g^{1/2} \| \longrightarrow 0 \;\Rightarrow\; \sup_{t\in\mathbb{C}} |h_n(t) - h(t)| \longrightarrow 0 \;\Rightarrow$$

$$| \inf_{t\in\mathbb{C}} h_n(t) - \inf_{t\in\mathbb{C}} h(t) | \longrightarrow 0 \;\Rightarrow\; h_n(\theta_n) \longrightarrow h(\theta)$$

Desde que $\sup_{t\in\mathbb{C}} |h_n(t) - h(t)| \longrightarrow 0 \;\Rightarrow$

$$|h_n(\theta_n) - h(\theta_n)| \longrightarrow 0.$$

$$|h(\theta_n) - h(\theta)| \le |h(\theta_n) - h_n(\theta_n)| + |h_n(\theta_n) - h(\theta)|$$

Portanto, $h(\theta_n) \longrightarrow h(\theta)$. Agora, se $\theta_n \not\longrightarrow \theta$, pela compacidade de $\mathbb{C}$, existiria uma subsequencia $(\theta_m) \subset (\theta_n)$ tal que $\theta_m \longrightarrow \theta_1 \neq \theta$ e como h é contínua $h(\theta_m) \longrightarrow h(\theta_1)$, daí temos que $h(\theta_1) = h(\theta)$, o que contradiz a unicidade de $T(G)=\theta$.

Note que se $T(G)$ é único então a consistência de (T_n) segue diretamente do teorema I.5.

<u>Corolário:</u> Suponha

(i) Γ identificável, i.e., $\theta_1 \neq \theta_2 \Rightarrow \| f_{\theta_1}^{1/2} - f_{\theta_2}^{1/2} \| > 0$;

(ii) $\inf_{t\in\Theta-\mathbb{C}} \| f_\theta^{1/2} - f_t^{1/2} \| > 0$, para algum compacto $\mathbb{C} \subset \Theta$ e $f_\theta(x)$ contínua em $\theta \in \mathbb{C}$ para cada x.

Então quando $n \longrightarrow \infty$, $\| f_\theta^{1/2} - g_n^{1/2} \| \longrightarrow 0$ implica $T(G_n) \longrightarrow \theta$.

Prova:

$\inf_{t \in \Theta - \mathbb{C}} \| f_\theta^{1/2} - f_t^{1/2} \| > 0$ implica que θ é ponto interior de $\mathbb{C}$. Daí, existe $T(F_\theta) = \theta$ e $T(F_\theta)$ é único, já que Γ é identificável. Da demonstração do Teorema I.8 (ver Simpson(1987)), pode-se ver que se $\| g_n^{1/2} - g^{1/2} \| \longrightarrow 0$, quando $n \to \infty$, então $G_n \in \mathcal{G}$ eventualmente, daí $T(G_n)$ existe e pertence a $\mathbb{C}$ eventualmente.

Agora, quando $n \to \infty$

$$\| f_{T(G_n)}^{1/2} - g_n^{1/2} \| \longrightarrow 0$$

pois $\| f_{T(G_n)}^{1/2} - g_n^{1/2} \| \leq \| f_\theta^{1/2} - g_n^{1/2} \|$, $\forall\, \theta \in \Theta$

desde que

$$\| f_{T(G_n)}^{1/2} - f_\theta^{1/2} \| \leq \| g_n^{1/2} - f_{T(G_n)}^{1/2} \| + \| f_\theta^{1/2} - g_n^{1/2} \| .$$

Portanto, quando $n \to \infty$

$$\| f_\theta^{1/2} - f_{T(G_n)}^{1/2} \| \longrightarrow 0 \quad \Rightarrow \quad f_{T(G_n)} \longrightarrow f_\theta$$

e como f_θ é contínua em θ temos

$$T(G_n) \longrightarrow \theta .$$

Ainda no artigo de Simpson(1987) encontramos que $H(t,G)$ é duas vezes diferenciável em $t \in \Theta$ e que o estimador MDH é um zero de $\frac{\partial}{\partial t} H(t,G_n)$, onde G_n é a função distribuição empírica. O que torna o problema de minimização de $H(t,G)$ num problema de resolver a seguinte equação: $\frac{\partial}{\partial t} H(t,G) = 0$. Mais ainda, em seu Teorema 2 Simpson mostra que

$$n^{1/2}(T_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0,V_\theta) \text{ quando } T_n \xrightarrow{P} \theta$$

onde

$$T_n \text{ é um zero de } \frac{\partial}{\partial t} H(t,G)\ ,$$

$$V_\theta = (1/4)\left[\frac{\partial^2}{\partial\theta^2} H(\theta,G)\right]^{-1} i(\theta) \left[\frac{\partial^2}{\partial\theta^2} H(\theta,G)\right]^{-1} \text{e}$$

$$i(\theta) = \int l_\theta . l'_\theta . f_\theta . d\mu,\ \mu \text{ medida dominante em } \Gamma$$

e que quando $G \equiv F_\theta$ então o estimador MDH é equivalente ao EMV.

Sob o ponto de vista da robustez qualitativa, o estimador MDH possui uma cota inferior para seu ponto de ruptura numa distribuição G. Considere o modelo de contaminação

$$\forall\, \varepsilon;\ 0 < \varepsilon < 1,\ \ H_n = (1-\varepsilon)G + \varepsilon K_n \qquad (1.3.7)$$

onde $\langle K_n \rangle_{n\geq 1}$ é uma sequência de distribuições.

Suponha que h_n, g e k_n sejam as densidades de H_n, G e K_n, respectivamente.

<u>Teorema I.8:</u> Seja $\hat{\rho} = \max_{t\in\Theta} \rho(G,F_t)$ e suponha que o máximo ocorra no interior de Θ. Seja $\rho^* = \lim_{M\to\infty} \sup_{|t|>M} \rho(G,F_t)$.
Se $\varepsilon < \dfrac{(\hat{\rho}-\rho^*)^2}{1+(\hat{\rho}-\rho^*)^2}$, então não existe nenhuma sequência da forma (1.3.7) para a qual $|T(H_n) - T(G)| \longrightarrow +\infty$, quando $n \to \infty$.

Prova: Suponha que $|T(H_n) - T(G)| \longrightarrow +\infty$, quando $n \rightarrow \infty$, onde H_n é da forma (1.3.7). Mostraremos que isto implica em $\varepsilon > \dfrac{(\hat{\rho}-\rho^*)^2}{1 + (\hat{\rho}-\rho^*)^2}$.

Sejam $T(G)=\theta$ e $T(H_n)=\theta_n$ os valores que maximizam $\rho(G,F_t)$ e $\rho(H_n,F_t)$, respectivamente.

Deve existir uma sequência $(\theta_n)_{n\geq 1}$ com $|\theta_n| \rightarrow \infty$, para a qual $\rho(H_n,F_{\theta_n}) > \rho(H_n,F_\theta)$ infinitas vezes. Admitamos que somente para um número finito de vezes tenhamos $\rho(H_n,F_{\theta_n}) > \rho(H_n,F_\theta)$, isto é:

$\exists\ n_1,n_2,\ldots,n_k;\ \rho(H_{n_i},F_{\theta_{n_i}}) > \rho(H_{n_i},F_\theta)$ para todo $i=1,\ldots,k$