

## *O Teorema Limite Central no Contexto da Teoria da Informação e Entropia*

**Madalena Malva**

*Escola Superior de Tecnologia de Viseu e Centro de Estatística e Aplicações da Universidade de Lisboa*

**Sandra Mendonça**

*Departamento de Matemática e Engenharias da Universidade da Madeira e Centro de Estatística e Aplicações da Universidade de Lisboa*

**Resumo:** O modelo gaussiano tem um protagonismo em Estatística que decorre de propriedades — algumas das quais características desse modelo, que portanto melhor seria apodado de “anormal” do que de normal — que permitem um tratamento matemático simples e elegante. Há uma tradição em usar o modelo gaussiano como aproximação, justificada pelo Teorema Limite Central (*TLC*). No entanto, a procura de demonstrações cada vez mais límpidas e gerais do *TLC*, exigindo cada vez menos condições sobre existência de momentos, levou a uma formulação em que a velocidade de convergência é apenas  $O\left(\frac{1}{\sqrt{n}}\right)$ . Porém, as modernas abordagens ao *TLC* usando teoria da informação levam a pensar que será possível (e mesmo natural) obter velocidades de convergência  $O\left(\frac{1}{n}\right)$ . Esta forma de equacionar o problema leva-nos naturalmente a estender a desigualdade de Cramér-Rao, e a caracterização das situações em que o limite inferior de Cramér-Rao é atingido é uma forma intuitiva de aceder à família exponencial de Fisher-Darmois-Koopman-Pitman.

**Palavras-chave:** Teoria da informação, desigualdade e limite inferior de Cramér-Rao, família exponencial de Fisher-Darmois-Koopman-Pitman.

**Abstract:** The Gaussian model has a special role in Statistics as a consequence of its properties — some of which characterize the model, that therefore would better be called “abnormal” than normal — that allow a simple and elegant mathematical treatment. There is a tradition to use the Gaussian model as an approximation, justified by the Central Limit Theorem (*CLT*). However, the continuing search for more limpid and general proofs of the *CLT*, less demanding in what regards the existence of moments, led to a formularization where the convergence speed is only  $O\left(\frac{1}{\sqrt{n}}\right)$ . However, the modern approaches to the *CLT* using information theory makes us think that it would be possible (and natural) to obtain convergence speeds  $O\left(\frac{1}{n}\right)$ . This form of equating the problem takes us to naturally extend the Cramér-Rao inequality, and the characterization of the situations where the Cramér-Rao inequality lower bound is reached is a clear form to access the exponential family of Fisher-Darmois-Koopman-Pitman.

**Keywords:** Information theory, Cramér-Rao inequality and lower bound, Fisher-Darmois-Koopman-Pitman exponential family.

## 1 Introdução e definições

Neste trabalho serão consideradas apenas variáveis aleatórias (*va's*)  $X$  puramente discretas com função massa de probabilidade  $\{p_k\}_{x_k \in S}$  ou *va's* absolutamente contínuas com função densidade de probabilidade  $f_X$  com suporte,  $S = S(X) = \{x_k : p_k = P(X = x_k) > 0\}$ , no caso discreto, e  $S = S(X) = \{x : f_X(x) > 0\}$ , no caso puramente absolutamente contínuo.

A informação  $\mathcal{I}(A)$  dada pelo acontecimento  $A$  é uma função decrescente de  $P(A)$ . Adoptamos a definição proposta por Shannon(1948):

$$\mathcal{I}(A) = -\log_2 P(A).$$

Tomando a variável aleatória definida por  $\mathcal{I}_X(x) = \mathcal{I}(X = x_k)$ ,  $x_k \in S$  no caso em que a variável aleatória  $X$  é discreta, e a generalização para o caso em que  $X$  é absolutamente contínuo,  $\mathcal{I}_X = -\log_2(f_X(X))$ , e adoptando a convenção  $0 \times \ln(0) = 0$ , define-se entropia da variável aleatória  $X$  como

$$\mathcal{H}(X) = E[\mathcal{I}_X] = \begin{cases} -\sum_{k: x_k \in S} \log_2(p_k) p_k & (X \text{ puramente discreta}) \\ -\int_S \log_2[f_X(x)] f_X(x) dx & (X \text{ absolutamente cont\u00ednua}) \end{cases}.$$

Entropia é a medida de incerteza associada a uma variável aleatória (cf. Cover e Thomas, 1991) e mede-se em *bits*.

É fácil provar que no caso em que  $X \sim \text{Gaussiana}(\mu, \sigma)$ ,  $\mu \in \mathcal{R}$ ,  $\sigma > 0$  a entropia é dada por

$$\mathcal{H}(X) = \frac{1}{2} \log_2(2 \pi e \sigma^2),$$

e no caso de  $X \sim \text{Exponencial}(\delta)$ ,  $\delta > 0$

$$\mathcal{H}(X) = \log_2 \delta + \frac{E(X)}{\delta} \log_2 e = \log_2(\delta e).$$

Chamamos distância da entropia relativa, ou distância de Kullback-Leibler, ou ainda divergência de informação entre duas funções densidade de probabilidade  $f_X$  e  $f_Y$  a

$$\mathcal{D}(f_X; f_Y) = \int_{S(X)} \log_2 \left[ \frac{f_X(x)}{f_Y(x)} \right] f_X(x) dx.$$

A aplicação  $\mathcal{D}$  não é uma distância no sentido topológico, uma vez que não é simétrica e não obedece à desigualdade triangular. No entanto  $\mathcal{D}(f_X; f_Y) \geq 0$ , com  $\mathcal{D}(f_X; f_Y) = 0$  se, e só se,  $X \stackrel{d}{=} Y$ .

Seja  $X$  uma variável aleatória cuja função densidade de probabilidade  $f_X(x|\theta) = f_X(x - \theta)$  tem derivada contínua. A função de pontuação (score) é a derivada logarítmica da função densidade

$$\rho_X(x) = \frac{d}{dx} \ln f_X(x) = \frac{f'_X(x)}{f_X(x)}$$

e a informação de Fisher relativamente ao parâmetro de localização  $\theta$  é

$$\mathcal{J}(X, \theta) = E \left[ \rho_X^2(X) \right].$$

Pode ainda definir-se informação estandardizada de Fisher, se  $X$  tiver função score  $\rho_X$ , valor médio  $\mu$  e variância  $\sigma^2$  como

$$\mathcal{J}_{st}(X) = \sigma^2 E \left[ \left( \rho_X(X) + \frac{X - \mu}{\sigma^2} \right)^2 \right].$$

Finalmente, a distância de informação de Fisher entre duas *va's* absolutamente contínuas é

$$\begin{aligned} \mathcal{D}_J(X, Y) &= \mathcal{D}_J(f_X, f_Y) \\ &= \int_{\mathcal{R}} f_X(x) \left[ \frac{d}{dx} \ln f_X(x) - \frac{d}{dx} \ln f_Y(x) \right]^2 dx \\ &= 4 \int_{\mathcal{R}} f_Y(x) \left[ \frac{d}{dx} \sqrt{\frac{f_X(x)}{f_Y(x)}} \right]^2 dx. \end{aligned}$$

## 2 Entropia e Informação de Fisher

As variáveis aleatórias com distribuição exponencial e gaussiana têm notáveis propriedades de maximização da entropia, respectivamente na classe das *va's* com suporte na semi-recta e de valor médio finito, e na classe das *va's* com suporte na recta e variância finita.

**Teorema 1:** Seja  $f_X$  a função densidade de probabilidade de uma *va*  $X$  com  $E(X) = \mu$  e suporte  $S = (0, \infty)$ , e seja  $Y_\mu \sim Exponencial(\mu)$ . Então  $H(X) \leq H(Y_\mu)$ , com igualdade se, e só se,  $X$  for uma *va* exponencial.

### Demonstração

Tem-se que

$$\begin{aligned} \mathcal{D}(f_X; f_{Y_\mu}) &= \int_{\mathcal{R}} \log_2 \left[ \frac{f_X(x)}{f_{Y_\mu}(x)} \right] f_X(x) dx \\ &= \int_0^\infty [\log_2 f_X(x) - \log_2 f_{Y_\mu}(x)] f_X(x) dx \end{aligned}$$

#### 4 Malva e Mendonça/TLC no contexto da informação e entropia

$$\begin{aligned}
&= \int_0^\infty f_X(x) \log_2 f_X(x) dx + \log_2 \mu \int_0^\infty f_X(x) dx + \\
&\quad \log_2 e \int_0^\infty \frac{x}{\mu} f_X(x) dx \\
&= -\mathcal{H}(X) + \log_2 \mu + \frac{\log_2 e}{\mu} \mu \\
&= -\mathcal{H}(X) + \log_2(\mu e) \\
&= -\mathcal{H}(X) + \mathcal{H}(Y_\mu).
\end{aligned}$$

Como

$$D(f_X; f_{Y_\mu}) \geq 0 \Rightarrow -\mathcal{H}(X) + \mathcal{H}(Y_\mu) \geq 0 \Leftrightarrow \mathcal{H}(X) \leq \mathcal{H}(Y_\mu),$$

com igualdade se, e só se,  $X \stackrel{d}{=} Y_\mu$ .

**Teorema 2:** Seja  $f_X$  a função densidade de probabilidade de uma *va*  $X$  com variância  $\sigma^2$  e suporte  $S = \mathcal{R}$ , e considere-se, sem perda de generalidade, que  $E(X) = 0$ , e seja  $Y_\sigma \sim \text{Gaussiana}(0, \sigma)$ . Então  $H(X) \leq H(Y_\sigma)$ , com igualdade se, e só se,  $X$  for uma *va* gaussiana.

#### Demonstração

Tem-se que

$$\begin{aligned}
\mathcal{D}(f_X; f_{Y_\sigma}) &= \int_{\mathcal{R}} \log_2[f_X(x)] f_X(x) dx - \int_{\mathcal{R}} \log_2[f_{Y_\sigma}(x)] f_X(x) dx \\
&= -\mathcal{H}(X) + \int_{\mathcal{R}} \frac{\log_2(2\pi\sigma^2)}{2} f_X(x) dx + \frac{\log_2 e}{2\sigma^2} \int_{\mathcal{R}} x^2 f_X(x) dx \\
&= -\mathcal{H}(X) + \frac{1}{2} \log_2(2\pi e \sigma^2) \\
&= -\mathcal{H}(X) + \mathcal{H}(Y_\sigma) \geq 0,
\end{aligned}$$

e conseqüentemente  $\mathcal{H}(Y_\sigma) \geq \mathcal{H}(X)$ , com igualdade se, e só se,  $X \stackrel{d}{=} Y_\sigma$ .

De Bruijn (cf. Barron (1985)) estabeleceu uma ligação notável entre a entropia e informação de Fisher, nomeadamente que a entropia máxima e informação mínima são atingidas sob as mesmas condições.

**Teorema 3:** Na classe de todas as densidades com variância finita igual a  $\sigma^2$  satisfazendo as condições:

- i.  $f$  têm derivadas contínuas

- ii.  $\int_{\mathcal{R}} x^2 f(x) dx < \infty$
- iii.  $|x|f(x) \rightarrow 0$  quando  $|x| \rightarrow \infty$ ,

a informação mínima é atingida para distribuições gaussianas.

### Demonstração

Sem perda de generalidade vamos tomar  $\int_{\mathcal{R}} f(x) dx = 0$ . Seja  $S = \{x : f(x) > 0\}$ . Integrando por partes e tendo em conta as condições (i)-(iii) obtemos

$$\int_S x f'(x) dx = - \int_S f(x) dx$$

donde

$$\int_S x f'(x) dx = -1.$$

Pela desigualdade de Cauchy-Schwarz, tem-se

$$\begin{aligned} 1 &= \left[ \int_S x f'(x) dx \right]^2 = \left[ \int_S x \frac{f'(x)}{f(x)} f(x) dx \right]^2 \\ &\leq \int_S \left[ \frac{f'(x)}{f(x)} \right]^2 f(x) dx \cdot \int_S x^2 f(x) dx \\ &\Rightarrow \int_S \left[ \frac{f'(x)}{f(x)} \right]^2 f(x) dx \cdot \int_S x^2 f(x) dx \geq 1 \\ &\Leftrightarrow \int_S \left[ \frac{f'(x)}{f(x)} \right]^2 f(x) dx \cdot \sigma^2 \geq 1 \\ &\Leftrightarrow \int_S \{\log[f(x)]\}^2 f(x) dx \geq \frac{1}{\sigma^2} \\ &\Leftrightarrow \mathcal{J}(X) \geq \frac{1}{\sigma^2}, \end{aligned}$$

onde  $X$  é uma variável aleatória com densidade  $f$ .

A igualdade é atingida quando para alguma constante  $c$

$$\frac{f'(x)}{f(x)} = cx, \text{ quase certamente em } S \quad (1)$$

(com respeito à medida de Lebesgue). Por causa da continuidade de  $f'$  concluímos que  $\frac{f'(x)}{f(x)} = cx$ , para  $x \in S$ . Primitivando (1) obtemos

$$\int \frac{f'(x)}{f(x)} dx = \int cx dx$$

## 6 Malva e Mendonça/TLC no contexto da informação e entropia

$$\Leftrightarrow \log[f(x)] = \frac{cx^2}{2} + B,$$

ou seja,

$$f(x) = A \exp\left(\frac{cx^2}{2}\right), \quad \forall x \in S, \quad (2)$$

onde  $A = \exp B$ . Mas,  $A \exp\left(\frac{cx^2}{2}\right) \neq 0$  para  $x$  real; então, a continuidade de  $f$ , juntamente com (1) implica

$$f(x) = A \exp\left(\frac{cx^2}{2}\right), \quad \forall x$$

onde  $c = -\frac{1}{\sigma^2}$  e  $A = \frac{1}{\sigma\sqrt{2\pi}}$ .

**Teorema 4:** Na classe de todas as densidades com suporte na semi-recta positiva e com momentos  $\alpha_i = \int_0^\infty x^i f(x) dx$ ,  $i = 1, 2, 3$  satisfazendo as condições (i) e (ii) do teorema anterior e ainda as condições  $x^2 f(x) \rightarrow 0$  quando  $x \rightarrow \infty$ , e  $xf(x) \rightarrow 0$  quando  $x \rightarrow 0$ , a informação mínima é atingida pelas distribuições gama.

### Demonstração

Tem-se que

$$\int_S x f'(x) dx = -1 \quad \text{e} \quad \int_S x^2 f'(x) dx = -2\alpha_1.$$

Multiplicando a primeira relação por  $\mu$ , a segunda por  $-\lambda$  e somando obtém-se

$$\begin{aligned} \mu \int_S x f'(x) dx + \lambda \int_S x^2 f'(x) dx &= -\mu + 2\lambda\alpha_1 \\ \Leftrightarrow \int_S x(\mu - \lambda x) f'(x) dx &= (2\lambda\alpha_1 - \mu). \end{aligned}$$

Aplicando a desigualdade de Cauchy-Schwarz, obtém-se

$$\begin{aligned} (2\lambda\alpha_1 - \mu)^2 &= \left\{ \int_S \frac{f'(x)}{f(x)} x(\mu - \lambda x) f(x) dx \right\}^2 \\ &\leq \int_S \left[ \frac{f'(x)}{f(x)} \right]^2 f(x) dx \cdot \int_S [x(\mu - \lambda x)]^2 f(x) dx. \end{aligned}$$

Então, para quaisquer  $\lambda$  e  $\mu$ ,

$$\int_S \left( \frac{f'(x)}{f(x)} \right)^2 f(x) dx \geq \frac{(\mu - 2\lambda\alpha_1)^2}{\int_S [x(\mu - \lambda x)]^2 f(x) dx} = \frac{(\mu\alpha - 2\lambda\alpha_1)^2}{\mu^2\alpha_1 - 2\mu\lambda\alpha_2 + \lambda^2\alpha_3}$$

$$\Leftrightarrow \mathcal{I}(X) \geq \frac{(\mu - 2\lambda\alpha_1)^2}{\mu^2\alpha_1 - 2\mu\lambda\alpha_2 + \lambda^2\alpha_3},$$

onde para  $\lambda$  e  $\mu$  dados a igualdade é verificada quando, para alguma constante  $c$

$$\frac{x f'(x)}{f(x)} = c(\mu - \lambda x)$$

se verifica quase certamente em  $S$ . Por causa da continuidade de  $f$  a relação anterior é satisfeita para todo o  $x$  em  $S$ . Absorvendo a constante  $c$  em  $\mu$  e  $\lambda$  e integrando a igualdade anterior obtemos

$$f(x) = k \exp(-\lambda x) x^\mu, \quad x \in S. \quad (3)$$

Como  $\exp(-\lambda x) x^\mu \neq 0, \forall x \in \mathcal{R}^+$ , a continuidade de  $f$  garante a validade de (3) para todo o  $x$  em  $\mathcal{R}^+$ . Fazendo  $\mu = \nu - 1$  e determinando  $k$  pelas condições de normalidade chegamos à forma usual da densidade Gama

$$f(x) = \frac{\lambda^\nu}{\Gamma(\nu)} \exp(-\lambda x) x^{\nu-1}, \quad x > 0.$$

Se tomarmos  $\lambda = \frac{\alpha_1}{\alpha_2 - \alpha_1^2}$  e  $\nu = \frac{\alpha_1^2}{\alpha_2 - \alpha_1^2}$  os momentos da distribuição são  $\alpha_1$  e  $\alpha_2$ .

### 3 Desigualdade de Cramér-Rao generalizada

A famosa desigualdade de Cramér-Rao pode ser demonstrada de uma forma simples usando as propriedades da função score. Por simples integração por partes, e supondo que uma dada função  $g$  satisfaz as condições de regularidade necessárias — existência de derivada e  $\lim_{x \rightarrow \pm\infty} f_X(x)g(x) = 0$  — resulta que

$$E[\rho_X(X)g(X)] = \int_{\mathcal{R}} \rho_X(x)g(x)f_X(x)dx = -E[g'(X)].$$

Supondo que  $X$  tem valor médio  $\mu$ , as funções lineares  $g(x) = ax + b$  satisfazem as condições de regularidade necessárias

$$E[\rho_X(X)(aX + b)] = -E[(aX + b)'] = -a.$$

## 8 Malva e Mendonça/TLC no contexto da informação e entropia

Conclui-se assim que fazendo  $a = \frac{1}{\sigma^2}$ ,

$$0 \leq E \left[ \left( \rho_X(X) + \frac{X - \mu}{\sigma^2} \right)^2 \right] \Rightarrow \mathcal{J}(X) \geq \frac{1}{\sigma^2}.$$

Pretende-se através da informação de Fisher chegar a uma das formas da desigualdade de Cramér-Rao. Considere-se duas *va's*  $X$  e  $Y$ , ambas com suporte  $S = S(X) = \{x : f_X(x) > 0\}$  e respectivas funções densidade de probabilidade,  $f_X$  e  $f_Y$ . Considere-se ainda  $K$  uma função regular e  $\alpha$  um número real. Temos

$$\int_S \left\{ \frac{d}{dx} \ln \left[ \frac{f_X(x)}{f_Y(x)} \right] + \alpha K(x) \right\}^2 f_X(x) dx \geq 0.$$

Recordando a definição da distância da informação de informação de Fisher

$$\mathcal{D}_J(f_X, f_Y) = \int_S \left\{ \frac{d}{dx} \ln \left[ \frac{f_X(x)}{f_Y(x)} \right] \right\}^2 f_X(x) dx$$

obtém-se

$$\begin{aligned} & \int_S \left\{ \frac{d}{dx} \ln \left[ \frac{f_X(x)}{f_Y(x)} \right] + \alpha K(x) \right\}^2 f_X(x) dx = \\ &= \int_S \left\{ \frac{d}{dx} \ln \left[ \frac{f_X(x)}{f_Y(x)} \right] \right\}^2 f_X(x) dx + \int_S 2 \frac{d}{dx} \ln \left[ \frac{f_X(x)}{f_Y(x)} \right] \alpha K(x) f_X(x) dx + \\ & \quad \int_S [\alpha K(x)]^2 f_X(x) dx \\ &= \mathcal{D}_J(f_X, g_Y) + 2\alpha \int_S \frac{d}{dx} \ln \left[ \frac{f_X(x)}{f_Y(x)} \right] K(x) f_X(x) dx + \alpha^2 E \{ [K(X)]^2 \}. \end{aligned} \tag{4}$$

Mas

$$\begin{aligned} & \int_S \frac{d}{dx} \ln \left[ \frac{f_X(x)}{f_Y(x)} \right] K(x) f_X(x) dx = \\ &= \int_S \frac{f'_X(x)}{f_X(x)} K(x) f_X(x) dx - \int_S \frac{f'_Y(x)}{f_Y(x)} K(x) f_X(x) dx \\ &= \int_S f'_X(x) K(x) dx - \int_S \rho_Y(x) K(x) f_X(x) dx, \end{aligned}$$

onde  $\rho_Y = \frac{f'_Y}{f_Y}$ , recorde-se, é a função score de  $Y$ . Integrando o primeiro integral por partes vem (da regularidade de  $K$ )

$$\int_S \frac{d}{dx} \ln \left[ \frac{f_X(x)}{f_Y(x)} \right] K(x) f_X(x) dx =$$

$$\begin{aligned}
&= [f_X(x)K(x)]_S - \int_S f_X(x)K'(x)dx - E[\rho_Y(X)K(X)] \\
&= - \int_S f_X(x)K'(x)dx - E[\rho_Y(X)K(X)] \\
&= -E[K'(X)] - E[\rho_Y(X)K(X)].
\end{aligned}$$

Substituindo em (4) obtém-se

$$\begin{aligned}
&\int_S \left\{ \frac{d}{dx} \ln \left[ \frac{f_X(x)}{f_Y(x)} \right] + \alpha K(x) \right\}^2 f_X(x)dx = \\
&= \mathcal{D}_J(f_X, f_Y) - 2\alpha E[K'(X)] - 2\alpha E[\rho_Y(X)K(X)] + \alpha^2 E\{[K(X)]^2\},
\end{aligned}$$

de onde resulta

$$\mathcal{D}_J(f_X, f_Y) \geq 2\alpha \left\{ E[K'(X)] + E[\rho_Y(X)K(X)] \right\} - \alpha^2 E\{[K(X)]^2\}, \quad \forall \alpha \in \mathcal{R}.$$

O segundo membro é máximo para

$$\alpha = \frac{E[K'(X)] + E[\rho_Y(X)K(X)]}{E[K^2(X)]}.$$

De facto,

$$\begin{aligned}
\frac{d}{d\alpha} (2\alpha \{E[K'(X)] + E[\rho_Y(X)K(X)]\} - \alpha^2 E[K^2(X)]) &= 0 \Leftrightarrow \\
2E[K'(X)] + 2E[\rho_Y(X)K(X)] - 2\alpha E[K^2(X)] &= 0 \Leftrightarrow \\
\frac{E[K'(X)] + E[\rho_Y(X)K(X)]}{E[K^2(X)]} &= \alpha
\end{aligned}$$

e

$$\frac{d^2}{d\alpha^2} (2\{E[K'(X)] + E[\rho_Y(X)K(X)]\} - \alpha^2 E\{[K(X)]^2\}) = -2E[K^2(X)] < 0.$$

Consequentemente,

$$\begin{aligned}
\mathcal{D}_J(f_X, f_Y) &\geq 2 \frac{\left\{ E[K'(X)] + E[\rho_Y(X)K(X)] \right\}^2}{E[K^2(X)]} - \\
&\quad \left\{ \frac{E[K'(X)] + E[\rho_Y(X)K(X)]}{E[K^2(X)]} \right\}^2 E[K^2(X)] \\
&= \frac{\left\{ E[K'(X)] + E[\rho_Y(X)K(X)] \right\}^2}{E[K^2(X)]}.
\end{aligned}$$

## 10 Malva e Mendonça/TLC no contexto da informação e entropia

Seja  $X$  uma va com  $E(X) = 0$ ,  $Var(X) = \sigma^2$  e  $Z_\sigma \sim \text{Gaussiana}(0, \sigma)$ . Mostra-se que

$$\mathcal{J}_{st}(X) = \sigma^2 \mathcal{D}_J(X, Z_\sigma) \geq \frac{\sigma^2 \{E[K'(X)] - \frac{X}{\sigma^2} K(X)\}^2}{E\{[K(X)]^2\}}. \quad (5)$$

Seja  $S_n = \sum_{k=1}^n X_k$ , e denote-se  $Y_n = \frac{S_n}{\sqrt{n}}$ . Sem perda de generalidade, suponha-se que  $E(X_k) = 0$  e que  $Var(X_k) = \sigma^2$ , e em (5) considere-se  $K(Y_n) = Y_n^2 - \sigma^2$ . Admitindo que  $E(X_k^4) < \infty$  tem-se

$$\mathcal{J}_{st}(Y_n) \geq \frac{\gamma_1^2(X)}{\gamma_2(X)(Y_n) + 2} \quad \text{i.e.,} \quad \mathcal{J}_{st}(Y_n) \geq \frac{\gamma_1^2(X)}{2 + \frac{\gamma_2(X)}{n}},$$

onde  $\gamma_1 = \rho_3 = \frac{\mu_3}{\sigma^3}$  e  $\gamma_2 = \rho_4 = \frac{\mu_4}{\sigma^4} - 3$ .

Como  $\liminf_{n \rightarrow \infty} n \mathcal{J}_{st}(Y_n) \geq \frac{\gamma_1^2(X)}{2}$  o resultado acima mostra que a menos que  $\gamma_1^2(X) = 0$  (simetria), o melhor que há a esperar é uma velocidade de convergência  $O\left(\frac{1}{n}\right)$ .

## 4 A família exponencial de Fisher-Darmois-Koopman-Pitman

As famílias exponenciais surgem naturalmente quando se investiga as condições sob as quais é atingido o limite inferior na desigualdade de Cramér-Rao. Consideremos uma função densidade de probabilidade  $f$  que depende de modo regular do parâmetro  $\theta$  ( $\theta \in \Theta$ ) e que verifica a condição:

$$\int_S \frac{\partial}{\partial \theta} f(x, \theta) dx = 0.$$

Considere-se variáveis aleatórias,  $X_1, \dots, X_n$ , independentes e identicamente distribuídas com função densidade de probabilidade comum dada por  $f$ . A

função densidade de probabilidade conjunta é dada por  $\mathbf{f}(\mathbf{x}, \theta) = \prod_{i=1}^n f(x_i, \theta)$ ,

para  $\mathbf{x} = (x_1, \dots, x_n) \in S^n$ . Seja  $\theta^*$  um estimador não enviesado de  $\theta$  satisfazendo a condição

$$\int_{S^n} \theta^*(\mathbf{x}) \frac{\partial}{\partial \theta} \mathbf{f}(\mathbf{x}, \theta) d\mathbf{x} = 1.$$

Não é difícil mostrar que, sob esta condição,  $Var(\theta^*) \geq \frac{1}{nA(\theta)}$ , onde  $A(\theta) = \int_{S^n} \left[ \frac{\partial \mathbf{f}(\mathbf{x}, \theta)}{\partial \theta} \right]^2 \frac{1}{\mathbf{f}(\mathbf{x}, \theta)} d\mathbf{x}$  é a informação de Fisher.

Seja  $D(x, \theta) = \frac{\partial}{\partial \theta} \ln f(x, \theta)$ . Prova-se que

$$\int_{S^n} (\theta^*(\mathbf{x}) - \theta) \left( \sum_{i=1}^n D(x_i, \theta) \right) \mathbf{f}(\mathbf{x}, \theta) d\mathbf{x} = 1, \theta \in \Theta. \quad (6)$$

Tomando

$$\langle g_1, g_2 \rangle = \int_{S^n} g_1(\mathbf{x}) g_2(\mathbf{x}) \eta(\mathbf{x}) d\mathbf{x}$$

podemos reescrever a equação (6) na forma

$$\langle g_1, g_2 \rangle = 1,$$

com  $g_1(\mathbf{x}) = \theta^*(\mathbf{x}) - \theta$ ,  $g_2(\mathbf{x}) = \sum_{i=1}^n D(x_i, \theta)$  e  $\eta(\mathbf{x}) = \mathbf{f}(\mathbf{x}, \theta) \geq 0$ .

Para obter a igualdade na desigualdade de Cramér-Rao é necessário (e suficiente) que as funções  $g_1$  e  $g_2$  sejam linearmente dependentes; assim considere-se  $g_1(\mathbf{x}) = \lambda g_2(\mathbf{x})$ . Então

$$\int_{S^n} g_1(\mathbf{x}) g_2(\mathbf{x}) \eta(\mathbf{x}) d\mathbf{x} = \lambda \int_{S^n} g_2^2(\mathbf{x}) \eta(\mathbf{x}) d\mathbf{x}.$$

Mas, como neste caso  $\int_{S^n} g_1(\mathbf{x}) g_2(\mathbf{x}) \eta(\mathbf{x}) d\mathbf{x} = 1$ , tem-se que  $\lambda = \frac{1}{nA(\theta)}$ . Obtém-se assim a relação

$$\theta^*(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \left[ \theta + \frac{D(X_i, \theta)}{A(\theta)} \right].$$

Se se tomar  $c_i(\mathbf{x}) = \theta + \frac{D(x_i, \theta)}{A(\theta)}$ , tem-se  $D(x_i, \theta) = [c_i(\mathbf{x}) - \theta] A(\theta)$  e o estimador  $\theta^*$  tem a forma  $\theta^*(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n c_i(\mathbf{X})$ . Se tomarmos  $\sum_{i=1}^n c_i(\mathbf{x}) = c(\mathbf{x})$  obtém-se

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln [\mathbf{f}(\mathbf{x}, \theta)] &= \frac{\partial}{\partial \theta} \ln \left[ \prod_{i=1}^n f(x_i, \theta) \right] = \sum_{i=1}^n \frac{\partial}{\partial \theta} \ln f(x_i, \theta) = \sum_{i=1}^n D(x_i, \theta) = \\ &= \sum_{i=1}^n [c_i(\mathbf{x}) - \theta] A(\theta) = [c(\mathbf{x}) - n\theta] A(\theta). \end{aligned}$$

Resolvendo a equação

$$\frac{\partial}{\partial \theta} \ln [\mathbf{f}(\mathbf{x}, \theta)] = [c(\mathbf{x}) - n\theta] A(\theta),$$

## 12 Malva e Mendonça/TLC no contexto da informação e entropia

onde  $A(\theta) = B''(\theta)$ , vem

$$\frac{\partial}{\partial \theta} \ln [\mathbf{f}(\mathbf{x}, \theta)] = [c(\mathbf{x}) - n\theta] A(\theta) \Rightarrow$$

$$\ln [\mathbf{f}(\mathbf{x}, \theta)] = B'(\theta) [c(\mathbf{x}) - n\theta] + nB(\theta) + H(\mathbf{x}) \Rightarrow$$

$$\Rightarrow f(\mathbf{x}, \theta) = \exp \left[ B'(\theta) \left( n \frac{c(\mathbf{x})}{n} - n\theta \right) + nB(\theta) + H(\mathbf{x}) \right] \quad (7)$$

$$= \exp [B'(\theta) n (\theta^*(\mathbf{x}) - \theta) + nB(\theta) + H(\mathbf{x})]. \quad (8)$$

Ou seja, a família anterior pertence à família exponencial de Fisher-Darmois-Koopman-Pitman. Assim,  $\{f(x, \theta), \text{ com } \theta \in \Theta\}$  é exponencial, isto é, dada pela equação (7), onde  $B''(\theta) > 0$ .

### Agradecimentos

Ao professor Dinis Pestana pelo constante apoio prestado.

### Referências

- [1] Barron, R. A. (1985). Entropy and the central limit theorem. *Ann. Probab.*, Vol. 1, p. 336-342.
- [2] Cover, T. e Thomas, J. (1991). *Elements of Information Theory*. Wiley.
- [3] Feller, W. (1943). Generalization of a probability limit theorem of Cramér. *Trans. Amer. Soc.*, Vol. 54, p. 361-372.
- [4] Feller, W. (1966). *An Introduction to Probability Theory and its Applications*, Vol. II. Wiley.
- [5] Kotz, S. (1966). Recent results in information theory. *J. Appl. Probab.*, Vol. 39, p. 399-409.
- [6] Pestana, D. e Velosa, S. (2006). *Introdução à Probabilidade e à Estatística*, Vol. I. Fundação Calouste Gulbenkian.
- [7] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.*, Vol. 27, p. 379-423 e 623-656.