JOURNAL OF MULTIVARIATE ANALYSIS 27, 447–456 (1988)

Inference Properties of a One-Parameter Curved Exponential Family of Distributions with Given Marginals

CARMEN RUIZ-RIVAS

Universidad Autónoma de Madrid, Madrid, Spain

AND

CARLES M. CUADRAS

Universitat de Barcelona, Barcelona, Spain

Communicated by the Editors

This paper introduces a one-parameter bivariate family of distributions whose marginals are arbitrary and which include Fréchet bounds as well as the distribution corresponding to independent variables. Some geometrical and statistical properties on the stochastic dependence parameter are studied, considering this family as a member of Efron's curved exponential families of distributions. © 1988 Academic Press, Inc.

1. INTRODUCTION

Let X, Y be two random variables with continuous distribution functions $F(x)$, $G(y)$. Let us consider the class $\mathscr{F}$ of all possible joint cdf's H for (X, Y).

Hoeffding [11] and Fréchet [10] stated that the following extremal cdf's

$$H^+(x, y) = \min\{F(x), G(y)\}$$

$$H^-(x, y) = \max\{F(x) + G(y) - 1, 0\}$$

define two elements of $\mathscr{F}$ with associated extreme correlations, i.e.,

Received January 29, 1988.

AMS 1980 subject classifications: 62F10, 62H05, 62H20.

Key words and phrases: Fréchet bounds, Bivariate distribution, Efron curvature, Rao distance, maximum likelihood estimation.

0047-259X/88 $3.00

$\rho^- \leqslant \rho \leqslant \rho^+$, where ρ^-, ρ, ρ^+ are the correlation coefficients for H^-, H, H^+, respectively. H^-, H^+ are called the Fréchet bounds. It is verified that

$$H^-(x, y) \leqslant H(x, y) \leqslant H^+(x, y) \qquad \forall (x, y) \in \mathbb{R}^2.$$

Furthermore, if $H = H^-$ then

$$F(X) + G(Y) = 1 \qquad \text{(a.s.)},$$

and if $H = H^+$ then

$$F(X) = G(Y) \qquad \text{(a.s.)}.$$

Many authors have been interested in constructing parametric families of cdf's with given marginals F and G. Fréchet states that every family should include H^- and H^+. Kimeldorf and Sampson [12], proposed five desirable conditions that should be satisfied by any one-parameter family $\{H_\theta: -1 \leqslant \theta \leqslant +1\}$ of cdf's with absolutely continuous marginals F and G. These conditions are:

(a) $H_1(x, y) = H^+(x, y)$;

(b) $H_0(x, y) = F(x)\, G(y)$;

(c) $H_{-1}(x, y) = H^-(x, y)$ (i.e., the family contains the Fréchet bounds as well as the stochastic independence case);

(d) H_θ is continuous in $\theta \in [-1, 1]$;

(e) H_θ is absolutely continuous for fixed $\theta \in (-1, 1)$.

The uniform representation (Kimeldorf and Sampson [13]) and the notion of copula (Schweizer and Sklar [18]) provide the natural framework in which to study certain dependence properties of bivariate distributions and non-parametric measures of correlation. The uniform representatin or copula of H_θ is

$$U_H(u, v) = H(F^{-1}(u), G^{-1}(v)) \qquad (u, v) \in [0, 1]^2,$$

the marginal distributions of U_H then being uniform on $[0, 1]$.

Fréchet, Farlie, Gumbel, Morgenstern, Plackett, Mardia., Kimeldorf, Sampson, Ruiz-Rivas, Cuadras, Auge, Algarra, Nelsen, and others, have proposed one-parameter families. One of these families (see Section 2.2) is studied here.

Some applications deal with:

(a) Variance reduction in statistical simulation (Fishman [9], Whitt [20]).

(b) The construction of non-negative quantum-mechanical distribution functions, given the marginal distribution of position and moment (Cohen and Zaparovanny [4], O'Connell and Wigner [15], Cohen [3]).

(c) The construction of upper and lower bounds of the cdf's when the marginal are given, under the additional condition that $X \leqslant Y$ with probability one (Smith [19]).

2. One-Parameter System

2.1. *Definition*

Cuadras and Auge [6] defined the cdf on R^2,

$$H_\theta(x, y) = F(x)^{1-\theta} G(y) \quad \text{if} \quad F(x) \geqslant G(y),$$
$$H_\theta(x, y) = F(x)\, G(y)^{1-\theta} \quad \text{if} \quad F(x) < G(y),$$

θ being a parameter satisfying $0 \leqslant \theta \leqslant 1$. The general definition, including the negative parameter case, is:

$$\begin{aligned} H_\theta(x, y) &= [\min\{F(x), G(y)\}]^\theta \cdot [F(x)\, G(y)]^{1-\theta} \quad \text{for} \quad 0 \leqslant \theta \leqslant 1, \\ H_\theta(x, y) &= F(x) - [\min\{F(x), 1 - G(y)\}]^{-\theta} \cdot [F(x)(1 - G(y))]^{1+\theta} \\ &\qquad \text{for} \quad -1 \leqslant \theta < 0. \end{aligned} \tag{1}$$

2.2 *General properties*

The one-parameter system H_θ of cdf's has some interesting properties:

(1) If (X, Y) is distributed as $H_\theta(x, y)$, $0 \leqslant \theta \leqslant 1$, and Z verifies $G(Z) = 1 - G(Y)$ (a.s.) then (X, Z) is distributed as $H_{-\theta}(x, y)$.

(2) $H_1 = H^+$, $H_0 = FG$, $H_{-1} = H^-$, and H_θ is continuous in θ.

(3) H_θ is not absolutely continuous for $\theta \neq 0$, but can be decomposed as

$$H_\theta = H_\theta^{(1)} + H_\theta^{(2)}, \tag{2}$$

$H_\theta^{(1)}$ being its absolutely continuous part with density function (for $\theta \in [0, 1]$)

$$h_\theta(x, y) = (1 - \theta) f(x)\, g(y) \max\{F(x), G(y)\}^{-\theta} \qquad \forall (x, y) \in \mathbb{R}^2, \tag{3}$$

provided that F, G are absolutely continuous with densities f, g, and $H_\theta^{(2)}$ being its singular part corresponding to a positive mass over the curve

$F(x)=G(y)$. (The negative case $\theta\in[-1,0)$ is straightforward considering property (1).) In fact,

$$\begin{aligned} H_\theta^{(1)}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} h_\theta(u,v)\,du\,dv \\ &= -\frac{\theta}{2-\theta}\min\{F(x),G(y)\}^{2-\theta} \\ &\quad +\min\{F(x),G(y)\}[\max\{F(x),G(y)\}]^{1-\theta} \\ &= -H_\theta^{(2)}(x,y)+H_\theta(x,y) \qquad \forall(x,y)\in\mathbb{R}^2. \end{aligned}$$

(4) Let $P_\theta=P_\theta^{(1)}+P_\theta^{(2)}$ be the probability measure related to H_θ. The family $\{P_\theta\colon \theta\in[0,1]\}$ is dominated by a σ-finite measure μ, and

$$f_\theta(x,y)=h_\theta(x,y)\,I_{\bar{C}}(x,y)+\tilde{h}_\theta(x)\,I_C(x,y) \qquad \forall(x,y)\in\mathbb{R}^2,\quad \theta\in[0,1] \quad (4)$$

are the corresponding Radon–Nykodim derivatives, where $C=\{(x,y)\mid F(x)=G(y)\}$, I is the indicator function, and

$$\tilde{h}_\theta(x)=\theta f(x)\,F(x)^{1-\theta}.$$

Proof. Let λ^2, λ be the Lebesgue measures in $\mathbb{R}^2$ and $\mathbb{R}$, respectively. For any Borel set B in $\mathbb{R}^2$ let us define

$$\mu_0(B)=\lambda\{x\in\mathbb{R}\mid (x,G^{-1}F(x))\in B\},$$
$$\mu=\lambda^2+\mu_0;$$

μ_0 can be characterized as a product measure on $(\mathbb{R}^2,\beta^2)$

$$\mu_0(A\times B)=\int_A \tilde{\mu}(x,B)\,d\lambda(x), \qquad A,B\in\beta,$$

where β is the Borel σ-algebra of $\mathbb{R}$, and

$$\begin{aligned} \tilde{\mu}(x,B) &= 1 \quad \text{if } G^{-1}F(x)\in B, \\ &= 0 \quad \text{if } G^{-1}F(x)\notin B. \end{aligned}$$

Applying Fubini's theorem it is easily shown that

$$H_\theta^{(2)}(x,y)=\int_{-\infty}^{x}\int_{-8}^{y}\tilde{h}_\theta(u)\,d\mu_0(u,v)$$

so that $\tilde{h}_\theta=dP_\theta^{(2)}/d\mu_0$.

Noting that $\lambda^2(C)=0$, $\mu_0(\bar{C})=0$, and $h_\theta = dP_\theta^{(1)}/d\lambda^2$, the result (4) follows.

(5) $P_\theta(F(X)>G(Y))+P_\theta(F(X)<G(Y))=2(1-\theta)/(2-\theta)$,

$$P_\theta(F(X)=G(Y))=\theta/(2-\theta).$$

(6) The relations among θ and the Pearson's ρ, Kendall's τ, and Spearman's ρ_s correlations are

$$\rho=\frac{3\theta}{4-|\theta|} \qquad \text{(for uniform marginals)},$$

$$\tau=\frac{\theta}{2-|\theta|}, \qquad \rho_s=\frac{3\theta}{4-|\theta|}$$

(Cuadras and Auge [6]; Cuadras [5]).

(7) If (X_1, Y_1), (X_2, Y_2) are i.i.d. as H_θ, then

$$\theta=2-[P_\theta((X_1-X_2)(Y_1-Y_2)>0)]^{-1}.$$

Hence θ is invariant under monotone transformations of X and Y (Cuadras [5]).

3. Some Statistical Properties

3.1. *One-Parameter Curved Exponential Family*

Let $(X_1, Y_1), ..., (X_n, Y_n)$ be a bivariate random sample from H_θ, $\theta\in[0,1]$ (the study of the negative case $\theta\in[-1,0)$ is straightforward using suitable modifications).

Let $\alpha\subset\{1,2,...,n\}$ be the set of indexes of points in the sample lying on the curve C (i.e., $i\in\alpha$ iff $(x_i, y_i)\in C$).

Using the density function (4) with respect to the measure μ, the joint density function of the sample can be expressed as

$$\begin{aligned} f_\theta(\{x_i, y_i\}) &= \left[\prod_{i\notin\alpha} h_\theta(x_i, y_i)\right]\left[\prod_{i\in\alpha} \tilde{h}_\theta(x_i)\right] \\ &= (1-\theta)^n J(\{x_i, y_i\}) \exp\{\theta T + n_c \log(\theta/(1-\theta))\}, \qquad \theta\in[0,1] \end{aligned} \tag{5}$$

where

$$J(\{x_i, y_i\}) = \left[\prod_{i=1}^{n} f(x_i)\right]\left[\prod_{i\notin\alpha} g(y_i)\right] \exp\left\{\sum_{i\in\alpha} \log F(x_i)\right\}$$

does not depend on θ,

$$n_c = \#\alpha, \qquad T = -\sum_{i=1}^{n} \log \max\{F(x_i), G(y_i)\}. \tag{6}$$

The family of densities (5) constitutes a curved exponential family as named by Efron [7], where its curvature is the geometric curvature of $\mathscr{L} = \{(\theta, \log(\theta/(1-\theta)) : \theta \in [0,1]\}$ with respect to the inner product $\sum_\theta$, being $\sum_\theta$ the covariance matrix of (T, n_c).

It immediately follows that (T, n_c) is a minimal sufficient statistic for θ.

3.2. *Joint and Marginal Distribution of* (T, n_c)

(1) As n_c is the number of points in the sample lying on the curve C and $P_\theta(F(X) = G(Y)) = \theta/(2-\theta)$, n_c is a Binomial random variable $B(n, \theta/(2-\theta))$.

(2) T is a gamma random variable $G(2-\theta, n)$.

Proof. Let $Z = \max\{F(X), G(Y)\}$. Then $P_\theta(Z \leqslant z) = H_\theta(F^{-1}(z), G^{-1}(z)) = z^{2-\theta}$, $0 \leqslant z \leqslant 1 \Rightarrow P_\theta(-\log Z > u) = e^{-u(2-\theta)}$, $u > 0$. Thus $-\log \max\{F(X), G(Y)\} \sim G(2-\theta, 1)$ and hence $T \sim G(2-\theta, n)$.

(3) T and n_c are independent random variables.

Proof. $n_c = \sum_{i=1}^n U_i$, where $U_i = 1$ if $F(x_i) = G(y_i)$, and $U_i = 0$ if $F(x_i) \neq G(y_i)$. Then $U_i \sim B(1, \theta/(2-\theta))$, $i = 1, ..., n$, are all independent. $T = \sum_{i=1}^n V_i$ being $V_i = -\log \max\{F(X_i), G(Y_i)\} \sim G(2-\theta, 1)$. It is obvious that U_i is independent of V_j for $i \neq j$. In the case $i = j$, let $Z_1 = F(X_i)$, $Z_2 = G(Y_i)$. Then

$$P(U_i = 1, V_i > v) = P(Z_1 = Z_2, Z_1 < e^{-v}, Z_2 < e^{-v})$$

$$= \begin{cases} \theta/(2-\theta) & \text{if } v < 0, \\ \displaystyle\int_0^{e^{-v}} \theta x^{1-\theta}\, dx = \frac{\theta}{2-\theta} e^{-v(2-\theta)} & \text{if } v \geqslant 0, \end{cases}$$

and thus

$$P(U_i = 1, V_i > v) = P(U_i = 1) \cdot P(V_i > v).$$

Analogously

$$P(U_i = 0, V_i > v) = P(U_i = 0) \cdot P(V_i > v).$$

Let us remark that (T, n_c) is not a complet statistic. For instance, from (1) and (2), we have

$$E_\theta(2T - n_c) = n \qquad \forall \theta \in [0, 1].$$

3.3 *Curvature and Fisher Information Measure*

Let us denote $\eta_\theta = (\theta, \log(\theta/(1-\theta)))'$, Σ_θ the covariance matrix of (T, n_c) and

$$M_\theta = \begin{pmatrix} \dot{\eta}'_\theta \Sigma_\theta \dot{\eta}_\theta & \dot{\eta}'_\theta \Sigma_\theta \ddot{\eta}_\theta \\ \ddot{\eta}'_\theta \Sigma_\theta \dot{\eta}_\theta & \ddot{\eta}' \Sigma_\theta \ddot{\eta}_\theta \end{pmatrix}$$

the point meaning componentwise derivatives with respect to θ. If $i_\theta(X)$ represents the Fisher information measure obtained for the r.v. X, we have (Efron [7])

$$\dot{\eta}'_\theta \Sigma_\theta \dot{\eta}_\theta = \frac{n(\theta(1-\theta)+2)}{(2-\theta)^2\,\theta(1-\theta)} = i_\theta(T) + i_\theta(n_c) = i_\theta(T, n_c).$$

The curvature being

$$\gamma_\theta = \left(\frac{|M_\theta|}{i_\theta^3(T, n_c)}\right)^{1/2} = \frac{(2\theta-1)(2-\theta)}{(\theta(1-\theta)+2)^{3/2}}\sqrt{\frac{2}{n}}, \qquad \theta \in (0, 1).$$

These properties may be used to study second-order efficiency and to construct confidence intervals for the estimation of θ (Efron [8], Moolgavkar and Venzon [14]).

3.4. *Rao Distance*

Let $\psi = \psi(\theta)$ be an admissible transformation of the parameter θ. The Fisher information measure on ψ contained in (T, n_c) satisfies

$$i_\psi(T, n_c) = \left(\frac{d\psi}{d\theta}\right)^2 i_\theta(T, n_c).$$

Thus, $i_\theta(T, n_c)$ can be considered as a covariant tensor of the second order for all $\theta \in (0, 1)$ and we can obtain the Rao distance [17] for the family H_θ (see Burbea and Rao [2]; Burbea [1]; Oller and Cuadras [16]). The Rao distance between θ_1 and θ_2 is given by

$$S(\theta_1, \theta_2) = \int_{\theta_1}^{\theta_2} \frac{\sqrt{n[\theta(1-\theta)+2]}}{(2-\theta)\sqrt{\theta(1-\theta)}}\, d\theta.$$

This distance is invariant under any admissible transformation of the parameter θ and the random vector (X, Y).

Using the function

$$\Phi(\varphi)=\sum_{i=0}^{\infty}\left(-\frac{1}{2}\right)^{i}\beta_{2i}(\varphi)\sum_{j=0}^{i}\binom{-\frac{1}{2}}{j}(9/8)^{j}$$

where

$$\beta_{2i}(\varphi)=\frac{(1;2;i)}{2^{i}i!}\varphi-\frac{1}{2}\sin\varphi\cos\varphi\sum_{k=1}^{i}\frac{(2i-1;-2;k-1)}{2^{k-1}(i;-1;k)}\sin^{2(i-k)}(\varphi)$$

with

$$(a;b;c)=a(a+b)(a+2b)\cdots(a+(c-1)b)$$

for real numbers a, b, and integer number c, we obtain

$$S(\theta_1,\theta_2)=\sqrt{n}\left[\Phi\left(\sin^{-1}\sqrt{\frac{2\theta_2}{1+\theta_2}}\right)-\Phi(\sin^{-1}\sqrt{\frac{2\theta_1}{1+\theta_1}})\right].$$

If $\theta_1\simeq\theta_2$, it is easy to check that

$$\frac{1}{\sqrt{n}}S(\theta_1,\theta_2)=\frac{\sqrt{2+\theta_1(1-\theta_1)}}{(2-\theta_1)}[\sin^{-1}(2\theta_2-1)-\sin^{-1}(2\theta_1-1)]$$
$$+O((\theta_2-\theta_1)(\sqrt{\theta_2}-\sqrt{\theta_1})),$$

which provides a useful approximation for $S(\theta_1,\theta_2)$.

4. Maximum Likelihood Estimation of θ

From expression (5) we obtain the log-likelihood function

$$\log L(\{x_i,y_i\};\theta)=(n-n_c)\log(1-\theta)+n_c\log\theta+\theta T$$

and by solving the equation

$$\frac{\partial}{\partial\theta}\log L(\{x_i,y_i\};\theta)=0$$

we get the maximum likelihood estimation of θ

$$\hat{\theta}=\frac{T-n+\sqrt{(n-T)^2+4n_cT}}{2T}.$$

Let $a = \sqrt{(n-T)^2 + 4n_c T}$. Since $|n-T| \leqslant a \leqslant n+T$, we see that

$$0 \leqslant \frac{T-n+|n-T|}{2T} \leqslant \frac{T-n+a}{2T} \leqslant \frac{n+T+T-n}{2T} = 1.$$

Thus, we check that $0 \leqslant \hat{\theta} \leqslant 1$.

Let $(x_1, y_1), ..., (x_n, y_n)$ be a bivariate random sample from H_θ, $-1 \leqslant \theta < 0$. Let Z_i be such that $G(Z_i) = 1 - F(X_i)$ (a.s.), $i = 1, ..., n$, so $(x_1, z_1), ..., (x_n, z_n)$ is a sample from $H_{-\theta}$ and we obtain the maximum likelihood estimate for θ,

$$\hat{\theta} = \frac{n - T - \sqrt{(n-T)^2 + 4n_c T}}{2T},$$

where now n_c is the number of pairs (x_i, y_i) satisfying $F(x_i) + G(y_i) = 1$ and $T = -\sum_{i=1}^{n} \log \max\{F(x_i), 1 - G(y_i)\}$.

References

[1] Burbea, J. (1986). Informative geometry of probability spaces. *Exposition Math.* **4**, 347–378.

[2] Burbea, J., and Rao, C. R. (1982). Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *J. Multivariate Anal.* **12** 575–596.

[3] Cohen, L. (1984). Probability distributions with given marginals *J. Math. Phys.* **25** 2402–2403.

[4] Cohen, L., and Zaparovanny, Y. I. (1980). Positive quantum joint distributions. *J. Math. Phys.* **21** 794–796.

[5] Cuadras, C. M. (1985). Sobre medidas de dependencia estocástica invariantes. *Homenatge a F. d'A. Sales. Contribucions Científiques*, pp. 28–47, Faculty of Mathematics, Univ. of Barcelona.

[6] Cuadras, C. M., and Auge, J. (1981). A continuous general multivariate distribution and its properties. *Comm. Statist. A—Theory Methods* **10** 339–353.

[7] Efron, B. (1975). Defining the curvature of a statistical problem with applications to second order efficiency (with discussion). *Ann. Statist.* **3** 1189–1242.

[8] Efron, B. (1978). The geometry of exponential families. *Ann. Statist.* **6** 362–376.

[9] Fishman, G. S. (1972). Variance reduction in simulation studies. *J. Statist. Comput. Simulation* **1** 173–182.

[10] Fréchet, M. (1951). Sur les tableaux de correlations dont les marges sont données. *Ann. Univ. Lyon Sect. A Ser. 3* **14** 53–77.

[11] Hoeffding, W. (1940). Masstabinvariante Korrelations theorie. *Schriftenreihe Math. Inst. Angew. Math. Univ. Berlin* **5** 179–233.

[12] Kimeldorf, G., and Sampson, A. (1975). One-parameter families of bivariate distributions with fixed marginals. *Comm. Statist.* **4** 293–301.

[13] Kimeldorf, G., and Sampson, A. (1975). Uniform representations of bivariate distributions. *Comm. Statist.* **4** 617–627.

[14] Moolgavkar, S. H., and Venzon, D. J. (1987). Confidence regions in curved exponential families: Application to matched case-control and survival studies with general relative risk function. *Ann. Statist.* **15** 346–359.

[15] O'CONNELL, R. F., AND WIGNER, E. P. (1981). Some properties of a non-negative quantum-mechanical distribution function. *Phys. Lett. A* **85** 121–126.

[16] OLLER,, J. M., AND CUADRAS, C. M. (1985). Rao's distance for negative multinomial distributions. *Sankhyā A* **47** 75–83.

[17] RAO, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37** 81–91.

[18] SCHWEIZER, R., AND SKLAR, E. F. (1981). *Probabilistic Metric Spaces.* Elsevier Science, New York.

[19] SMITH, W. (1983). Inequalities for bivariate distributions with $X \leqslant Y$ and marginals given. *Comm. Statist. A—Theory Methods* **12** 1371–1379.

[20] WHITT, W. (1976). Bivariate distributions with given marginals. *Ann. Statist.* **4** 1280–1289.