A note on the Sarmanov bivariate distributions

J.S. Huang, G.D. Lin*

Department of Mathematics and Statistics, University of Guelph, Ontario, N1G 2W1, Canada
Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan, ROC

Abstract

We investigate the maximum correlation for Sarmanov bivariate distributions with fixed marginals and strengthen the existing results in the literature. The improvement in the maximum correlation is significant. A characterization of the Sarmanov distribution via chi-square divergence is also given. This extends Nelsen [13] result about the Farlie–Gumbel–Morgenstern (FGM) distribution.

1. Introduction

The problem of constructing a joint distribution with given marginals and correlation coefficient has wide ranging applications. See for instance, Lee [10], Fischer and Klein [5] and Baker [1]. A well known method is what is commonly known as the FGM distribution, due to Eyraud [3], Farlie [4], Gumbel [6] and Morgenstern [12]. It is as popular as it is simple. It admits, however, a rather limited range of correlation. Schucany et al. [15] proved that FGM distributions with absolutely continuous marginals can never have correlations higher than 1/3 and Lin [11] extended their result to the case of continuous marginals. To overcome this limitation several authors suggested various generalizations. Johnson and Kotz [8] and Lai and Xie [9] proposed different extensions of the FGM. Then there is the Sarmanov family of distributions, of which FGM is a special case. Each resulted in significant extensions of the range of extreme correlations. For an excellent account of the recent development of bivariate distributions in this regard, see Shubina and Lee [16] and Balakrishnan and Lai [2].

In the next section we shall investigate the maximum correlation for Sarmanov bivariate distributions with fixed marginals and strengthen the existing results of Shubina and Lee [16]. The improvement in the maximum correlation is significant. A characterization of the Sarmanov distribution via chi-square divergence is given in Section 3. This extends Nelsen [13] result about the FGM distribution.

2. The FGM and the Sarmanov families of bivariate density functions

Let $X$ and $Y$ be random variables with density functions $f$ and $g$, respectively. The bivariate FGM joint density with marginals $f$ and $g$ is defined by

$$h(x, y) = f(x)g(y)[1 + \alpha(1 - 2F(x))(1 - 2G(y))], \quad x, y \in \mathbb{R} \equiv (-\infty, \infty),$$

for some $\alpha \in [-1, 1]$, where $F$ and $G$ are the corresponding distribution functions.
A wider family of densities is what is called the Sarmanov density (with marginal densities $f$ and $g$), defined by

$$h(x, y) = f(x)g(y) / [1 + x\theta_1(x)\theta_2(y)], \quad x, y \in \mathbb{R},$$

(2)

where $\theta_1$ and $\theta_2$ are measurable functions satisfying

$$1 + x\theta_1(x)\theta_2(y) \geq 0 \quad \text{and} \quad E(\theta_1(X)) = E(\theta_2(Y)) = 0;$$

see Sarmanov [14, p. 508]. For applications of the Sarmanov density in electrical engineering, see Willett and Thomas [17,18]. Shubina and Lee [16] obtained the maximum (or minimum) value of correlation coefficient $\rho(X, Y)$ attainable by the density (2). They have also shown how such a density can be constructed. Take for instance the case of uniform (0,1) marginals. By setting the 'mixing functions' in (2) as $\theta_1(x) = 2x_1(x) - 1$, $i = 1, 2$, where $x_1$ is the indicator function of the set $A$, they have proved that the resulting density has $\rho = 3/4$ and that it is the absolute maximum $\rho$ attainable by any Sarmanov density with uniform marginals.

It so happens that among the FGM densities in (1) the maximum correlation of $1/3$ is achieved by the one with the uniform marginals. FGM densities with any other marginals all have lower $\rho$ (for instance, the normal has $\rho = 1/\pi = 0.318$, the logistic has $\rho = 3/\pi^2 = 0.304$ and the Laplace has $\rho = 9/32 = 0.281$; see also Guerrra [7, p. 35]). Within the Sarmanov family, however, the uniform no longer enjoys this distinction. Shubina and Lee [16] pointed out, by way of an example, that some Sarmanov densities can have $\rho$ approaching 1, which of course is higher than $\rho_{\text{max}} = 3/4$ of the uniform case (see their Example 4.3, page 1039, the symmetric beta marginals; note that the $\sigma^2$ in their (4.6) and the display on page 1041 was missing). We now offer an alternative example to the same effect. Ours is much simpler and, not having to involve the incomplete beta function, its correlation has a closed form.

**Example.** Sarmanov density with identical $U$-shaped marginals.

Let

$$f(x) = g(x) = \begin{cases} (n+1)(1-2x)^n, & 0 < x < \frac{1}{2}, \\ (n+1)(2x-1)^n, & \frac{1}{2} < x < 1 \end{cases}$$

and

$$h(x, y) = \begin{cases} 2f(x)g(y) = 2(n+1)^2(1-2x)^n(1-2y)^n, & 0 < x, y < \frac{1}{2}, \\ 2\theta(x)g(y) = 2(n+1)^2(2x-1)^n(2y-1)^n, & \frac{1}{2} < x, y < 1, \\ 0, & \text{elsewhere}. \end{cases}$$

Then $\rho(X, Y) = \frac{(n+1)(n+3)}{(n+2)^2} \to 1$ as $n \to \infty$. To achieve the utmost negative $\rho$, the modification of $h(x, y)$ is obvious.

Our example shares a characteristic feature with Example 4.3 of Shubina and Lee [16], who showed that maximization of correlation necessitates purging of all the mass from two quadrants: $h(x, y) = 0$ on $(x-x_0)(y-y_0) < 0$ for some suitable $x_0$, $y_0$. Should we insist on a joint density with positive density on the entire region $(0, 1)^2$, it is of course easily effected through a convex combination: $h_q(x, y) = qh(x, y) + (1-q)f(x)g(y)$. For large $q$ the new density enjoys a near maximum correlation.

For the Sarmanov densities with identical normal (0,1) marginals, Shubina and Lee [16] showed that the maximum correlation is $\rho_{\text{max}} = 2/\pi$ (its proof is very complicated), which is twice as high as that for the FGM density of the same marginals. Our next theorem gives a general result for the Sarmanov with identical symmetric marginals. The specific results about the logistic and Laplace distributions are also presented as corollaries. It is seen that in each case, the maximum correlation is much higher than that for the FGM density of the same marginals. Therefore, the improvement in the maximum correlation is significant.

**Theorem 1.** Let $X$ and $Y$ be random variables identically distributed by $F$, symmetric about 0, with finite variance $\sigma^2 = E(X^2) = E(Y^2)$ and the common density $f = F$. Then the bivariate density

$$h(x, y) = \begin{cases} 2f(x)f(y), & -\infty < x, y < 0 \text{ or } 0 < x, y < \infty, \\ 0, & \text{elsewhere}, \end{cases}$$

belongs to the Sarmanov family with identical marginal distributions $F$. Its correlation $\rho(X, Y)$ is $\sigma^2 E(|X|)^2$. Consequently, $\rho_{\text{max}}$ for the Sarmanov densities with marginals $F$ has a lower bound

$$\rho_{\text{max}} \geq \sigma^2 E(|X|)^2 / E(X^2).$$

Moreover, if

$$f(p)/l(p) \geq \frac{1}{2} \left( p^{-1} - (1-p)^{-1} \right), \quad 0 < p \leq \frac{1}{2}$$

(4)

where $l(p) = \sigma^{-1} \int_0^p F^{-1}(t)dt$, $0 < p \leq \frac{1}{2}$ and $F^{-1}$ is the quantile function of $F$, $F^{-1}(t) = \inf \{ x : F(x) \geq t \}$, $t \in (0, 1)$, then the equality in (3) attains.
Proof. Note that for the Sarmanov density function \( h \),
\[
E(XY) = \int_{-\infty}^{0} \int_{-\infty}^{0} 2xyf(x)f(y)dx
dy + \int_{0}^{\infty} \int_{0}^{\infty} 2xyf(x)f(y)dx
dy = \left( \int_{0}^{\infty} 2xf(x)dx \right)^2 = (E[X])^2,
\]
from which (3) follows immediately. To prove the final conclusion, we recall that by Theorem 3 of Shubina and Lee [16], \( \rho_{\text{max}} \) for the Sarmanov densities with identical marginals \( F \) symmetric about 0 has an upper bound:
\[
\rho_{\text{max}} \leq \max_{p \in (0,1/2]} \frac{\rho^2(p)}{p(1-p)} \equiv \max_{p \in (0,1/2]} \rho(p),
\]
where \( \rho(p) = \rho^2(p)/|p(1-p)|. \) Under the condition (4), the upper bound is equal to \( \rho(1/2) = 4\rho^2(1/2) = (E[X])^2/E(X^2) \) because (4) implies that \( \rho(p) \) is monotonically increasing on \( (0,1/2] \). This completes the proof. \( \square \)

Corollary 1. Maximum correlation \( \rho_{\text{max}} \) for the Sarmanov densities with identical logistic marginals \( F(x) = (1 + e^{-x})^{-1}, x \in \mathbb{R} \), is \( \rho_{\text{max}} = \sigma^2(E[X])^2 = 12(\ln 2)^2/\pi^2 = 0.5842. \)

Proof. Recall that \( E|X| = 2 \ln 2 \) and \( E(X^2) = \pi^2/3. \) Since \( F^{-1}(t) = \ln(t/(1-t)), \ t \in (0,1), \) and
\[
I(p) = \frac{\sqrt{3}}{\pi} \left[ p \ln p + (1-p) \ln(1-p) \right], \ p \in (0,1/2],
\]
(4) is equivalent to
\[
g(p) = p \ln p - (1-p) \ln(1-p) \leq 0, \ p \in (0,1/2].
\]
To verify the last inequality, we use the fact that \( g(0^+) = g(1/2) = 0 \) and \( g''(p) = (1-2p)/|p(1-p)| > 0 \) on \( (0,1/2). \) The proof is complete. \( \square \)

Corollary 2. Maximum correlation \( \rho_{\text{max}} \) for the Sarmanov densities with identical Laplace marginals is \( \rho_{\text{max}} = 1/2. \)

Proof. For Laplace density \( f(x) = \frac{1}{2} e^{-|x|}, x \in \mathbb{R}, \) we have \( E|X| = 1, \ E(X^2) = 2, \ F(x) = \frac{1}{2} e^x, x \in (-\infty,0], \) and \( F^{-1}(t) = \ln(2t), \ t \in (0,1/2]. \) Thus,
\[
I(p) = \frac{1}{\sqrt{2}} p \ln(2p) - 1, \ p \in (0,1/2].
\]
In this case, (4) is equivalent to
\[
g(p) = -\ln(2p) + 2p - 1 \geq 0, \ p \in (0,1/2],
\]
which is true because \( g(0^+) = \infty, \ g(1/2) = 0 \) and \( g''(p) = -p^{-1} + 2 < 0 \) on \( (0,1/2). \) \( \square \)

3. Characterization of the Sarmanov distribution via \( \chi^2 \) divergence

Let \( h \) be the joint density of \( X \) and \( Y \) with marginal densities \( f = F' \) and \( g = G' \). Define the \( \chi^2 \) divergence (distance) between the joint density \( h \) and the product density \( fg \) of independent random variables by
\[
\chi^2(h,f,g) = \int_{S_F \times S_G} \left[ \frac{h(x,y)}{f(x)g(y)} - 1 \right]^2 f(x)g(y)dx
dy,
\]
(5)
where \( S_F \) and \( S_G \) are the supports of \( F \) and \( G \), respectively. Recall that the Spearman correlation of \( X \) and \( Y \) has the form of
\[
\rho^*(X,Y) = 12E \left[ \left( F(X) - \frac{1}{2} \right) \left( G(Y) - \frac{1}{2} \right) \right] = 3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1-2F(x))(1-2G(y))h(x,y)dx
dy.
\]
Nelsen [13] obtained the following interesting characterization of the FGM distributions.

Theorem N. Among all absolutely continuous bivariate distributions with marginal densities \( f \) and \( g \) and a given Spearman correlation \( \rho_{\text{max}} \), where \( |\rho_{\text{max}}| \leq 1/2 \), the one whose joint density is closest to the product density of independent random variables (in the sense of minimizing the \( \chi^2 \) divergence) is the FGM distribution with parameter \( \alpha = 3\rho_{\text{max}} \).

We now extend Nelsen [13] result to the case of Sarmanov distributions. For \( i = 1,2 \), consider the functions \( \theta_i : [0,1] \rightarrow [-1,1] \) satisfying
\[
\begin{align*}
\int_0^1 \theta_i'(u) du &= 0, \\
\sup_{u \in [0, 1]} \theta_i'(u) &= 1, \\
\inf_{u \in [0, 1]} \theta_i'(u) &= -h_i,
\end{align*}
\]

(6)

where \( h_i \in (0, 1] \). Then we have the following.

**Theorem 2.** Among all absolutely continuous bivariate distributions with marginal densities \( f = F \) and \( g = G \), the one whose joint density is closest to the product density of independent random variables (in the sense of minimizing the \( \chi^2 \) divergence) subject to the constraint

\[
E[\theta_1'(F(X))\theta_2'(G(Y))] = c_0,
\]

(7)

where \(-1 \leq \frac{c_0}{c_1c_2} \leq (\max(h_1, h_2))^{-1} \) and \( \theta_1', \theta_2' \) are given in (6), is the Sarmanov distribution having joint density

\[
h(x, y) = f(x)g(y)\left[1 + \frac{c_0}{c_1c_2}\theta_1'(F(x))\theta_2'(G(y))\right], \quad x, y \in \mathbb{R}.
\]

(8)

**Proof.** Set

\[
\varepsilon(x, y) = \frac{h(x, y)}{f(x)g(y)} - 1, \quad (x, y) \in S_F \times S_G.
\]

Then our problem is to find the joint density \( h \) which minimizes the \( \chi^2 \) divergence

\[
\chi^2(h; f, g) = \int_{S_F \times S_G} [\varepsilon(x, y)]^2 f(x)g(y) dx dy,
\]

(see (5)) subject to the constraint (7).

Consider the non-negative function

\[
A(a) = \int_{S_F \times S_G} [\varepsilon(x, y) - a\theta_1'(F(x))\theta_2'(G(y))]^2 f(x)g(y) dx dy = \chi^2(h; f, g) - (2c_0a - c_1c_2a^2) \equiv \chi^2(h; f, g) - A(a) \geq 0, \quad a \in \mathbb{R}.
\]

Therefore, for any \( h \),

\[
\chi^2(h; f, g) \geq \max_{a \in \mathbb{R}} A(a) = A\left(\frac{c_0}{c_1c_2}\right) = \frac{c_0^2}{c_1c_2}.
\]

A simple calculation shows that the \( \chi^2 \) divergence achieves the minimal value \( \frac{c_0^2}{c_1c_2} \) if we take the joint density \( h \) to be that in (8) which satisfies the constraint (7). On the other hand, for any joint density \( h \) satisfying the constraint (7) and

\[
\chi^2(h; f, g) = \frac{c_0^2}{c_1c_2},
\]

we have

\[
\int_{S_F \times S_G} \left[\varepsilon(x, y) - \frac{c_0}{c_1c_2}\theta_1'(F(x))\theta_2'(G(y))\right]^2 f(x)g(y) dx dy = \chi^2(h; f, g) - \frac{c_0^2}{c_1c_2} = 0.
\]

Hence \( \varepsilon(x, y) = \frac{c_0}{c_1c_2}\theta_1'(F(x))\theta_2'(G(y)) \) for \( (x, y) \in S_F \times S_G \). This in turn implies that \( h \) should be the joint density in (8). The proof is complete.

**Remark 1.** Let \( \theta'_i(u) = 1 - 2u, u \in [0, 1] \), then \( h_i = 1 \) and \( c_i = \frac{1}{2} \). When \( c_0 = \rho_0^2/3 \), Theorem 2 reduces to **Theorem N**.

**Acknowledgements**

The authors thank the Editor-in-Chief and the reviewer for helpful comments. The work of the second author is partly supported by the National Science Council of Taiwan, Republic of China, under Grant NSC 99–2118-M-001-003-MY3.

**References**