Efficient estimation in the bivariate normal copula model: normal margins are least favourable

CHRIS A.J. KLAASSEN[1] and JON A. WELLNER[2]*

[1]*Department of Mathematics, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands,* [2]*Department of Statistics, University of Washington, Box 354322, Seattle WA 98195-4322, USA*

Consider semi-parametric bivariate copula models in which the family of copula functions is parametrized by a Euclidean parameter θ of interest and in which the two unknown marginal distributions are the (infinite-dimensional) nuisance parameters. The efficient score for θ can be characterized in terms of the solutions of two coupled Sturm–Liouville equations. Where the family of copula functions corresponds to the normal distributions with mean 0, variance 1 and correlation θ, the solution of these equations is given, and we thereby show that the normal scores rank correlation coefficient is asymptotically efficient. We also show that the bivariate normal model with equal variances constitutes the least favourable parametric submodel. Finally, we discuss the interpretation of $|\theta|$ in the normal copula model as the maximum (monotone) correlation coefficient.

Keywords: bivariate normal; copula models; correlation; coupled differential equations; information; maximum correlation; normal scores; projection equations; rank correlation; semi-parametric model; Sturm–Liouville equations

1. Introduction: copula models

A distribution function C on the unit cube $[0, 1]^m$ in $\mathbb{R}^m$ with uniform marginal distributions is called a *copula*. A classical result of Sklar (1959) relates an arbitrary distribution F on $\mathbb{R}^m$ to a copula function C via the marginal distribution functions $F_1, \ldots, F_m$ of F:

Theorem 1.1 (Sklar 1959). *Suppose that F is a distribution function on $\mathbb{R}^m$ with one-dimensional marginal distribution functions $F_1, \ldots, F_m$. Then there is a copula C such that*

$$F(x_1, \ldots, x_m) = C(F_1(x_1), \ldots, F_m(x_m)). \tag{1.1}$$

If F is continuous, then the copula C satisfying (1.1) is unique and is given by

$$C(u_1, \ldots, u_m) = F(F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m)) \tag{1.2}$$

* To whom correspondence should be addressed.

for $u = (u_1, \ldots, u_m) \in (0,1)^m$ where $F_i^{-1}(u) = \inf\{x : F_i(x) \geqslant u\}$, $i = 1, \ldots, m$. Conversely, if C is a copula on $[0,1]^m$ and $F_1, \ldots F_m$ are distribution functions on $\mathbb{R}$, then the function F defined by (1.1) *is a distribution function on $\mathbb{R}^m$ with one-dimensional marginal distributions $F_1, \ldots, F_m$.*

See Sklar (1959) and Schweizer (1991) for some history. It seems that Hoeffding (1940) also had the basic idea of summarizing the dependence properties of a multivariate distribution by its corresponding copula, but he chose to define the corresponding function on $[-\frac{1}{2}, \frac{1}{2}]^m$ rather than on $[0,1]^m$. In particular, see the translation of the Hoeffding (1940) paper in Hoeffding (1994).

Our goal in this paper is to investigate efficient estimation for semi-parametric copula models $\mathscr{P}$ defined as follows: suppose that

$$\{C_\theta : \theta \in \Theta\} \tag{1.3}$$

is a parametric family of copula functions on $[0,1]^m$ with densities $\{c_\theta : \theta \in \Theta\}$ with respect to Lebesgue measure on $[0,1]^m$. For $\theta \in \Theta$ and arbitrary distribution functions $F_1, \ldots, F_m$ on $\mathbb{R}$, let $F_{\theta,F_1, \ldots, F_m}$ be the distribution function on $\mathbb{R}^m$ defined by

$$F_{\theta,F_1, \ldots, F_m}(x_1, \ldots, x_m) = C_\theta(F_1(x_1), \ldots, F_m(x_m)) \qquad \text{for } (x_1, \ldots, x_m) \in \mathbb{R}^m. \tag{1.4}$$

Then with $P_{\theta,F_1, \ldots, F_m}$ denoting the corresponding probability measures on $(\mathbb{R}^m, \mathscr{B}^m)$ and $\mathscr{F}$ denoting the collection of all distribution functions on $\mathbb{R}$,

$$\mathscr{P} = \{P_{\theta,F_1, \ldots, F_m} : \theta \in \Theta, F_i \in \mathscr{F}, i = 1, \ldots, m\} \tag{1.5}$$

is a *semi-parametric copula model*. Natural submodels of $\mathscr{P}$ are those with $\mathscr{F}$ replaced by $\mathscr{F}_c$, the collection of all continuous distribution functions F on $\mathbb{R}$, or by $\mathscr{F}_{ac}$, the collection of all absolutely continuous distribution functions.

One simple example, which is our main focus in this paper, is provided by the family of copulas resulting from multivariate normal distributions on $\mathbb{R}^m$. Suppose that $F_{\mu,\Sigma}$ is the multivariate normal distribution with mean $\mu = (\mu_1, \ldots, \mu_m)$ and covariance Σ with elements $\rho_{ij}\sigma_i\sigma_j$, $1 \leqslant i,j \leqslant m$. Let Φ denote the one-dimensional standard normal distribution function, and Φ_θ the m-dimensional standard normal distribution function with mean 0, variances 1, and correlations ρ_{ij}, $\theta = (\rho_{12}, \ldots, \rho_{m-1,m})$. Then

$$F_i(y_i) = \Phi((y_i - \mu_i)/\sigma_i), \qquad i = 1, \ldots, m,$$

and hence $C = C_\theta$ satisfies

$$C_\theta(u_1, \ldots, u_m) = \Phi_\theta(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_m)) \tag{1.6}$$

in this case. Note that the resulting semi-parametric copula model $\mathscr{P}$ given by (1.4) and (1.5) contains the family of normal distributions: if $F_i(x_i) = \Phi((x_i - \mu_i)/\sigma_i)$ for $i = 1, \ldots, m$, then $F_{\theta,F_1, \ldots, F_m} = F_{\mu,\Sigma}$.

For other copula families of considerable interest, see Kimeldorf and Sampson (1975a; 1975b), Clayton (1978), Genest and MacKay (1986a; 1986b), Genest (1987), Joe (1993) and Genest *et al.* (1995). Copula models are also strongly connected with frailty models, typically via a reparametrization to obtain uniform marginals (Sklar's theorem): for

interesting frailty models, see, for example, Marshall and Olkin (1988). For work on related transformation models, see Clayton and Cuzick (1985; 1986) and Klaassen (1988).

For ease of exposition, we will discuss first the bivariate case with $m = 2$ in detail. As we will see in Section 5.1 by a direct argument, our results will hold for the general m-dimensional case too.

We will formulate our estimation problem as follows. Suppose that $m = 2$ and the parametric family of copula functions on $[0, 1]^2$ is given by (1.6): thus

$$C_\theta(u, v) = \Phi_\theta(\Phi^{-1}(u), \Phi^{-1}(v)), \qquad 0 \leqslant u, v \leqslant 1. \tag{1.7}$$

Then we suppose that we observe a sample from the distribution $F_{\theta, G, H}$ of $X = (Y, Z)$ given by

$$F_{\theta, G, H}(y, z) = C_\theta(G(y), H(z)), \qquad (y, z) \in \mathbb{R}^2, \tag{1.8}$$

for some θ and distribution functions G and H on $\mathbb{R}$.

Note that θ is one-dimensional here, and equals the correlation coefficient of Y and Z when X is normally distributed. In fact, in this normal copula model, $|\theta|$ equals the maximum correlation coefficient of Y and Z. We will discuss this together with related concepts and their history in more detail in Section 2. We observe n i.i.d. copies $X_1, \ldots, X_n$ of X and we want to estimate the unknown parameters θ asymptotically efficiently in the presence of the unknown, arbitrary nuisance parameters G and H.

Our main result is that the normal scores rank correlation coefficient is an efficient estimator of θ with asymptotic variance $(1 - \theta^2)^2$, just the same as the asymptotic variance of the usual sample correlation coefficient in the case of normal marginal distributions. The normal scores rank correlation coefficient is also called the Van der Waerden rank correlation coefficient; see Section III.6.1 of Hájek and Šidák (1967). A precise formulation of our main result is given in Section 3 together with a proof. The asymptotic performance of this locally regular estimator follows directly from Ruymgaart (1974). To show that this performance is optimal, we need a bound stating that this performance cannot be improved. We will obtain such a bound by a simple study of a most difficult (least favourable) parametric submodel of our normal copula model. Such a model happens to be the model with X bivariate normal with mean 0 and unknown covariance matrix with equal variances. This extremal property of the multivariate normal distribution will be discussed in Section 2, together with other extremal properties of the normal distribution.

Viewing G and H as unknown monotone transformations, we see tha the normal copula model is a transformation model in the following sense: for all distribution functions G and H,

$$P_{\theta, G, H}(G(Y) \leqslant u, \; H(Z) \leqslant v) = C_\theta(u, v) = P_{\theta, \Phi, \Phi}(\Phi(Y) \leqslant u, \Phi(Z) \leqslant v). \tag{1.9}$$

Here we might choose the class of bivariate normal distributions with mean 0, variances 1 and correlation coefficient θ as the core model. In this sense all copula models are transformation models, the discussion of which has been initiated in Sections 4.7 and 6.7 of Bickel *et al.* (1993). In particular, formula (4.7.33) (Bickel *et al.* 1993, p.162) shows that the semi-parametric paradigm of projection of the score function for θ on the nuisance parameter tangent space leads to Sturm–Liouville differential equations. For copula

models with both marginal distributions unknown, this becomes a pair of coupled Sturm–Liouville equations. This approach will be discussed in Section 4, and yields another proof of the efficiency of the normal scores rank correlation coefficient. In fact, the simple proof given in Section 3 was discovered only after doing the information calculations as in Section 4, and for copula models other than Gaussian it seems unlikely that simple proofs or computations will be possible: the Gaussian case is the only example in which we have been able to compute the efficient scores and information explicitly, even though we know that the efficient scores and information exist in a large subclass of such models.

Since $|\theta|$ equals the maximum correlation coefficient in the normal copula model, one wonders if the maximum correlation coefficient can be estimated by a locally regular estimator in the nonparametric model of all bivariate distributions (or even some appropriate subset thereof). This is not the case. The maximum correlation coefficient cannot even be estimated locally consistently, since it is not a continuous parameter on any appropriately large class of bivariate distributions, as will be shown in Section 5.2.

2. Correlation and extremal properties of normal distributions

Many measures of dependence in bivariate distributions have been proposed. The first and most important of these is still the correlation coefficient, which may be ascribed to Galton (1888); see, for example, Stigler (1986). However, it has the unpleasant property that it can vanish for dependent variables. The maximum correlation coefficient, as proposed by Gebelein (1941), does not have this drawback. If (Y, Z) is a random vector one may consider the correlation coefficient $\rho(a(Y), b(Z))$ of $a(Y)$ and $b(Z)$ for transformations a and b from $\mathbb{R}$ to $\mathbb{R}$. Taking the supremum over all a and b such that $\mathrm{var}(a(Y))$ and $\mathrm{var}(b(Z))$ are positive and finite, we arrive at the maximum correlation coefficient

$$\rho_{\mathrm{M}}(Y, Z) = \sup_{a,b} \rho(a(Y), b(Z)). \tag{2.10}$$

Clearly, $\rho_{\mathrm{M}}(Y, Z) = 0$ if and only if Y and Z are independent; take a and b to be indicator functions (cf. Feller 1971, p. 136). If (Y, Z) is normal, then the maximum correlation coefficient equals the absolute value of the correlation coefficient, that is,

$$\rho_{\mathrm{M}}(Y, Z) = |\rho(Y, Z)|. \tag{2.11}$$

We will give a short proof of this equality and discuss its history in Section 6.

Within the normal copula model (1.7), it is straightforward to check that (Y, Z) and $(G^{-1}(G(Y)), H^{-1}(H(Z)))$ have the same distribution function $\Phi_\theta(\Phi^{-1}(G(y)), \Phi^{-1}(H(z)))$, and that $(\Phi^{-1}(G(Y)), \Phi^{-1}(H(Z)))$ has the standard normal distribution with correlation coefficient θ. Together with (2.11), this yields

$$|\theta| = |\rho(\Phi^{-1}(G(Y)), \Phi^{-1}(H(Z)))| \leqslant \rho_{\mathrm{M}}(Y, Z)$$
$$= \rho_{\mathrm{M}}(G^{-1}(G(Y)), H^{-1}(H(Z))) \tag{2.12}$$

$$= \rho_{\mathrm{M}}(G^{-1} \circ \Phi \circ \Phi^{-1} \circ G(Y), H^{-1} \circ \Phi \circ \Phi^{-1} \circ H(Z))$$

$$\leqslant \rho_{\mathrm{M}}(\Phi^{-1}(G(Y)), \Phi^{-1}(H(Z))) = |\theta|$$

and in the normal copula model the maximum correlation coefficient of Y and Z equals $|\theta|$,

$$\rho_{\mathrm{M}}(Y, Z) = |\theta|. \tag{2.13}$$

Since the copula model is a transformation model with monotone transformations, it is natural to restrict a and b in (2.10) to monotone functions. This leads to the (maximum) monotone correlation coefficient $\rho_{\mathrm{m}}(Y, Z)$ as defined in Section 4 of Kimeldorf and Sampson (1978). Again by (2.12), we see that

$$\rho_{\mathrm{m}}(Y, Z) = |\theta|. \tag{2.14}$$

Note that (2.13) and (2.14) imply that we are essentially estimating the maximum correlation coefficient and the maximum monotone correlation coefficient in our normal copula model. However, the parameter θ itself is the correlation coefficient proper of the normal core model, that is, after transformation of the marginals to normal distributions. Therefore we will call θ the *normal correlation coefficient*. As indicated above, we will show in Sections 3 and 4 that a least favourable parametric submodel of our copula model in estimating the normal correlation coefficient is the symmetric normal scale model. As a matter of fact, this shows that the information $I(P_0 | \theta, \mathscr{P})$ (cf. Bickel *et al.* (1993, (3.1.2) and (3.3.24), pp. 46 and 63) about θ at any distribution $P_0 = P_{\theta_0, G_0, H_0}$ within our normal copula model

$$\mathscr{P} = \{P_{\theta, G, H} \colon \theta \in (-1, 1), G, H \text{ continuous d.f.s}\} \tag{2.15}$$

equals the information $I(P_{\theta_0, \Phi, \Phi} | \theta, \mathscr{P}_{\mathrm{s}})$ about θ at $P_{\theta_0, \Phi, \Phi}$ within the symmetric normal scale model

$$\mathscr{P}_{\mathrm{s}} = \{P_{\theta, \Phi(\cdot/\sigma), \Phi(\cdot/\sigma)} \colon \theta \in (-1, 1), \sigma > 0\}. \tag{2.16}$$

We formulate this more precisely as follows: for $P_0 = P_{\theta_0, G_0, H_0} \in \mathscr{P}$,

$$I^{-1}(P_0 | \theta, \mathscr{P}) = \sup\{I^{-1}(P_0 | \theta, \mathscr{Q}) \colon \mathscr{Q} \subset \mathscr{P}, \mathscr{Q} \text{ regular parametric}\}$$

$$= I^{-1}(P_{\theta_0, \Phi, \Phi} | \theta, \mathscr{P}_{\mathrm{s}}), \tag{2.17}$$

or equivalently,

$$I(P_0 | \theta, \mathscr{P}) = \inf\{I(P_0 | \theta, \mathscr{Q}) \colon \mathscr{Q} \subset \mathscr{P}, \mathscr{Q} \text{ regular parametric}\}$$

$$= I(P_{\theta_0, \Phi, \Phi} | \theta, \mathscr{P}_{\mathrm{s}}). \tag{2.18}$$

Thus the regular parametric submodel $\mathscr{P}_{\mathrm{s}} \subset \mathscr{P}$ is least favourable. This is a surprising extremal property of the bivariate normal distribution which is similar in nature to the well-known fact that information for location, given the variance, is minimal at the normal distribution; cf. Huber (1981, p. 83), and Barron (1986, p. 337). An extension of this extremal property runs as follows. Fix the natural number n. Let T_n be a translation equivariant estimator of the location parameter ν of n i.i.d. random variables with

symmetric density $f(\cdot - \nu)$ and Fisher information for location $I(f)$. Then Theorem 2.3.2 of Klaassen (1981, p. 25), presents a sharpening of the Fréchet–Cramér–Rao inequality

$$\operatorname{var}_f(T_n) \geqslant (nI(f))^{-1} \tag{2.19}$$

and proves that equality can hold here if and only if f is a normal density. Inequality (2.19) itself has been given by Fréchet (1943, p. 191), without explicit mention of regularity conditions. Note that for $n = 1$ this reduces to the earlier result.

A related extremal property of the normal density is that it maximizes the Shannon entropy $-\int f \log f$ for a given variance; again cf. Barron (1986, p. 337). Finally, we mention another extremal property of the normal density. Let, for the moment, X be a random variable with density f. For f normal and g an absolutely continuous function with derivative g', Chernoff (1981) proved

$$\operatorname{var}_f g(X) \leqslant \operatorname{var}_f(X) \mathrm{E}_f(g'(X))^2. \tag{2.20}$$

This inequality has been generalized by Klaassen (1985) to general f. Borovkov and Utev (1983) defined

$$U_f = \sup_g \frac{\operatorname{var}_f g(X)}{\operatorname{var}_f(X) \mathrm{E}_f(g'(X))^2}$$

and showed

$$U_f \leqslant 1, \tag{2.21}$$

with equality if and only if f is normal.

3. Approach 1: estimation by rank correlation

Let $X = (Y, Z)$ and suppose there exist *known* transformations σ and τ such that

$$\begin{pmatrix} \sigma(Y) \\ \tau(Z) \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \eta_1^2 & \theta \eta_1 \eta_2 \\ \theta \eta_1 \eta_2 & \eta_2^2 \end{pmatrix} \right). \tag{3.22}$$

For example, we might have $\sigma(x) = \tau(x) = \log(x)$ and hence X log-normal. Suppose that we observe $X_1 = (Y_1, Z_1), \ldots, X_n = (Y_n, Z_n)$ i.i.d. as $X = (Y, Z)$. By applying σ and τ to the Ys and Zs, respectively, we arrive at the well-known situation of data with bivariate normal distribution with unknown covariance matrix and mean zero; see, for example, Example 2.4.6 of Bickel *et al.* (1993, pp. 36–38). The parameter of interest is θ, which can be estimated efficiently in the presence of the nuisance parameters η_1, η_2 by

$$\hat{\theta}_n = \frac{\frac{1}{n} \sum_{i=1}^n \sigma(Y_i) \tau(Z_i)}{\{\frac{1}{n} \sum_{i=1}^n \sigma^2(Y_i) \frac{1}{n} \sum_{i=1}^n \tau^2(Z_i)\}^{1/2}}$$

attaining the information lower bound $(1 - \theta^2)^2$; cf. Bickel *et al.* (1993, p. 38). In fact, this estimator is asymptotically linear with efficient influence function (at $\eta_1 = \eta_2 = 1$)

$$\sigma(Y)\tau(Z) - \frac{\theta}{2}(\sigma^2(Y) + \tau^2(Z)). \tag{3.24}$$

Furthermore, this information lower bound for θ is valid also in the submodel of (3.22) with the one-dimensional nuisance parameter $\eta^2 = \eta_1^2 = \eta_2^2$. In the case of the normal copula model with known marginal distributions G and H, the estimator (3.23) becomes

$$\hat{\theta}_n = \frac{\frac{1}{n}\sum_{i=1}^{n}\Phi^{-1}(G(Y_i))\Phi^{-1}(H(Z_i))}{\left\{\frac{1}{n}\sum_{i=1}^{n}[\Phi^{-1}(G(Y_i))]^2\frac{1}{n}\sum_{i=1}^{n}[\Phi^{-1}(H(Z_i))]^2\right\}^{1/2}}. \tag{3.25}$$

If σ and τ are *unknown*, then we have a semi-parametric model for which the parametric information lower bound $(1 - \theta^2)^2$ is still valid. In fact, we have the following model now: there exist monotone transformations $\tilde{\sigma}$ and $\tilde{\tau}$ such that $(\tilde{\sigma}(Y), \tilde{\tau}(Z)) \sim \Phi_\theta$, the bivariate normal distribution with standard normal marginal distributions and correlation θ. In agreement with Section 1 above and with Example 4.7.4.III of Bickel *et al.* (1993, p. 157) we call this model the normal copula model. In this normal copula model, it is natural to consider properties of the normal scores rank correlation coefficient obtained from $\hat{\theta}_n$ given in (3.25) by estimating G and H by the corresponding marginal empirical distributions $\mathbb{G}_n$ and $\mathbb{H}_n$ rescaled by $n/(n+1)$: with

$$\mathbb{G}_n^* \equiv \frac{n}{n+1}\mathbb{G}_n \qquad \mathbb{H}_n^* \equiv \frac{n}{n+1}\mathbb{H}_n, \tag{3.26}$$

the normal scores rank correlation coefficient $\hat{\rho}_n$ is

$$\hat{\rho}_n = \frac{\frac{1}{n}\sum_{i=1}^{n}\Phi^{-1}(\mathbb{G}_n^*(Y_i))\Phi^{-1}(\mathbb{H}_n^*(Z_i))}{\frac{1}{n}\sum_{i=1}^{n}\left[\Phi^{-1}\left(\frac{i}{n+1}\right)\right]^2} \tag{3.27}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}\Phi^{-1}\left(\frac{i}{n+1}\right)\Phi^{-1}\left(\frac{R_{ni}}{n+1}\right)}{\frac{1}{n}\sum_{i=1}^{n}\left[\Phi^{-1}\left(\frac{i}{n+1}\right)\right]^2}; \tag{3.28}$$

where $R_{ni} \equiv n\mathbb{H}_n(\mathbb{G}_n^{-1}(\frac{i}{n})), i = 1, \ldots, n$; see Hájek and Šidák (1967, p. 113).

In our present context, however, we need to consider the large-sample behaviour of $\hat{\rho}_n$ not only under the usual (independence) null hypothesis, but also under the normal copula model as specified above with $\theta \neq 0$. Fortunately, a very general study of the large-sample theory of rank correlation statistics under fixed alternatives has already been done by Ruymgaart *et al.* (1972) and Ruymgaart (1974). In particular, we will use Theorem 2.2 of Ruymgaart (1974) which treats the case of local sequences converging to a fixed alternative. When specialized to our normal copula model and $\hat{\rho}_n$, we obtain the following theorem.

Theorem 3.1 (Asymptotic linearity and efficiency of the normal scores rank correlation coefficient estimator). *If* $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ *are i.i.d.* $P \equiv P_{\theta_0, G, H} \in \mathscr{P}$, *then* $\hat{\rho}_n$ *is a locally asymptotically linear estimator of* θ *with (efficient) influence function*

$$\tilde{l}_\theta(y, z) \equiv \tilde{l}_\theta(y, z; \theta, G, H)$$

$$= \Phi^{-1}(G(y))\Phi^{-1}(H(z)) - \frac{\theta}{2}\{[\Phi^{-1}(G(y))]^2 + [\Phi^{-1}(H(z))]^2\}. \quad (3.29)$$

Thus, for $\theta_n = \theta_0 + O(1/\sqrt{n})$, *we have under* $P_{\theta_n, G, H}$

$$\sqrt{n}(\hat{\rho}_n - \theta_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{l}_\theta(Y_i, Z_i; \theta_n, G, H) + o_p(1) \quad (3.30)$$

$$\to_d N(0, (1 - \theta_0^2)^2). \quad (3.31)$$

Before proving the theorem, we note that the normal scores rank correlation coefficient $\hat{\rho}_n$ is, in fact the *efficient score equation estimator* of θ with G, H estimated by $\mathbb{G}_n^*$, $\mathbb{H}_n^*$: the solution θ of

$$0 = \frac{1}{n}\sum_{i=1}^n l_\theta^*(Y_i, Z_i; \theta, \mathbb{G}_n^*, \mathbb{H}_n^*)$$

$$= \frac{1}{n}\sum_{i=1}^n \frac{1}{(1 - \theta^2)^2}\left\{\Phi^{-1}(\mathbb{G}_n^*(Y_i))\Phi^{-1}(\mathbb{H}_n^*(Z_i)) - \frac{\theta}{2}\{[\Phi^{-1}(\mathbb{G}_n^*(Y_i))]^2 + [\Phi^{-1}(\mathbb{H}_n^*(Z_i))]^2\}\right\}$$

is just $\theta = \hat{\rho}_n$.

Proof. We first give a heuristic development showing why the result is true, followed by a formal proof based on Ruymgaart's (1974) theorem.

First note that

$$(\Phi^{-1}(G(Y_i)), \Phi^{-1}(H(Z_i))) \sim \Phi_\theta, \quad (3.32)$$

the bivariate normal distribution with mean 0, variances 1 and correlation θ. Since $n^{-1}\sum_{i=1}^n [\Phi^{-1}(i/(n+1))]^2 - 1 = O(n^{-1}\log n)$, we can rewrite $\sqrt{n}(\hat{\rho}_n - \theta)$ as

$$\sqrt{n}(\hat{\rho}_n - \theta) = \frac{1 + O(n^{-1}\log n)}{\sqrt{n}}\sum_{i=1}^n \{\Phi^{-1}(\mathbb{G}_n^*(Y_i))\Phi^{-1}(\mathbb{H}_n^*(Z_i)) - \theta\} + O(n^{-1/2}\log n)$$

$$= \frac{1 + o(1)}{\sqrt{n}}\sum_{i=1}^n \{\Phi^{-1}(G(Y_i))\Phi^{-1}(H(Z_i)) - \theta\} \quad (3.33)$$

$$+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \{\Phi^{-1}(\mathbb{G}_n^*(Y_i)) - \Phi^{-1}(G(Y_i))\}\Phi^{-1}(H(Z_i))$$

$$+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \{\Phi^{-1}(\mathbb{H}_n^*(Z_i)) - \Phi^{-1}(H(Z_i))\}\Phi^{-1}(G(Y_i)) \quad + o_p(1),$$

where the $o_{\mathrm{p}}(1)$ comes from replacement of $\Phi^{-1}(\mathbb{H}_n^*(Z_i))$ by $\Phi^{-1}(H(Z_i))$ in the second term.

Now we rewrite the second term, using $(\mathrm{d}/\mathrm{d}u)\Phi^{-1}(u) = 1/\phi(\Phi^{-1}(u))$ and Taylor expansion:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\phi(\Phi^{-1}(\tilde{\mathbb{G}}_n^*(Y_i)))}\,(\mathbb{G}_n^* - G)(Y_i)\cdot\Phi^{-1}(H(Z_i)) \tag{3.34}$$

$$= \int\frac{\Phi^{-1}(H(z))}{\phi(\Phi^{-1}(G(y)))}\,\sqrt{n}\,(\mathbb{G}_n^* - G)(y)\mathrm{d}\mathbb{P}_n(y,z) + o_{\mathrm{p}}(1)$$

$$= \int\frac{\Phi^{-1}(H(z))}{\phi(\Phi^{-1}(G(y)))}\,\sqrt{n}\,(\mathbb{G}_n^* - G)(y)\mathrm{d}P_{\theta,G,H}(y,z) + o_{\mathrm{p}}(1)$$

$$= \int\theta\Phi^{-1}(G(y))\sqrt{n}\,(\mathbb{G}_n^* - G)(y)\mathrm{d}\Phi^{-1}(G(y)) + o_{\mathrm{p}}(1)$$

$$= -\frac{\theta}{2}\int\{\Phi^{-1}(G(y))\}^2\mathrm{d}\{\sqrt{n}\,(\mathbb{G}_n^* - G)(y)\} + o_{\mathrm{p}}(1)$$

$$= -\frac{\theta}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\Phi^{-1}(G(Y_i))^2 - 1\} + o_{\mathrm{p}}(1);$$

here the third equality comes from computing conditionally on $Y = y$ and noting (3.32). An analogous development for the third term on the right-hand side of (3.33) shows that it can be rewritten as

$$-\frac{\theta}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\Phi^{-1}(H(Z_i))^2 - 1\} + o_{\mathrm{p}}(1). \tag{3.35}$$

Combining (3.33), (3.34) and (3.35) yields the conclusion – with the understanding that the arguments for the $o_{\mathrm{p}}(1)$ terms have been only heuristic.

We proceed with the formal proof by verifying Assumptions 2.1–2.3 and 2.5 of Ruymgaart (1974) with $\mathscr{H}' = \{P_{\theta_n,G,H}\colon n = 0,1,2,\ldots\}$ and $\varepsilon = 1/4$. (Note that regularity of the estimators is automatic as far as the nuisance parameters go.) First, note that $J = K = \Phi^{-1}$.

Assumption 2.1 holds easily, since $B_{0n}^* = 0$ if $a_n(i) = b_n(i) = \Phi^{-1}(i/(n+1))$.

As for Assumption 2.2, $J_{\mathrm{d}} = K_{\mathrm{d}} = 0$, $J_{\mathrm{c}} = J = K_{\mathrm{c}} = K = \Phi^{-1}$ is continuously differentiable, $r_2 = r_1 = |J| = |\Phi^{-1}|$, $\bar{r}_2 = \bar{r}_1 = |J'| = 1/\phi(\Phi^{-1})$.

Moving on to Assumption 2.3, the first supremum over $\mathscr{H}'$ is finite since, for all θ,

$$\int\int|\Phi^{-1}(G(y))\Phi^{-1}(H(z))|^{2+\varepsilon}\mathrm{d}P_{\theta,G,H}(y,z) = \mathrm{E}_{\Phi_\theta}(YZ)^{2+\varepsilon} < \infty. \tag{3.36}$$

Let $q_1(s) = \phi^{1/3}(\Phi^{-1}(s))$. Then

$$\int q_1^{-2-\varepsilon}(s)\mathrm{d}s = \int\phi^{1/4}(y)\mathrm{d}y < \infty, \tag{3.37}$$

and, by Hölder's inequality, the second supremum is finite because

$$E_{\Phi_\theta} |\phi^{1/3-1}(Y)Z|^{1+1/4} < \infty \tag{3.38}$$

since

$$E\phi^{-\frac{257}{346}}(Y) = \int \phi^{1/36}(y)\mathrm{d}y < \infty. \tag{3.39}$$

By symmetry the last supremum is the same and Assumption 2.3 is satisfied. Finally, we verify Assumption 2.5: (a) is satisfied in view of $J_{\mathrm{d}} = K_{\mathrm{d}} = 0$. Thus Theorem 2.2 of Ruymgaart (1974) applies and yields the asymptotic normality claimed, since the random variable in Ruymgaart's (3.5) equals $\theta + \tilde{l}_\theta(Y, Z)$. Asymptotic linearity follows from efficiency and the convolution theorem; see, for example, Theorem 3.3.2 of Bickel *et al.* (1993, p. 63). $\qquad\square$

As noted immediately following the theorem, $\hat{\rho}_n$ is the efficient score estimator of θ. It is interesting to note that it is also asymptotically equivalent to the 'pseudo-maximum likelihood' estimator obtained by estimating the unknown marginal distribution G and H by the marginal empiricals $\mathbb{G}_n^*$ and $\mathbb{H}_n^*$ and then maximizing the resulting 'pseudo-likelihood' as a function of θ, or by solving the (ordinary) score equation with G and H estimated away by the marginal empiricals $\mathbb{G}_n^*$ and $\mathbb{H}_n^*$: from the score for θ (see, for example, Bickel *et al.* (1993, pp. 36–37) we find that this estimator $\hat{\theta}_n^{\mathrm{pml}}$ is the solution θ of

$$0 = \theta(1 - \theta^2) - \frac{1}{n}\sum_{i=1}^{n}\{\theta[\Phi^{-1}(\mathbb{G}_n^*(Y_i))^2 + \Phi^{-1}(\mathbb{H}_n^*(Z_i))^2]$$

$$+ (1 + \theta^2)\Phi^{-1}(\mathbb{G}_n^*(Y_i))\Phi^{-1}(\mathbb{H}_n^*(Z_i))\}.$$

In view of $\sqrt{n}\{n^{-1}\Sigma_{i=1}^{n}[\Phi^{-1}(i/(n+1)]^2 - 1\} = o(1)$, the above equation may be rewritten as

$$0 = \theta(1 - \theta^2) - 2\theta + o(n^{-1/2}) + (1 + \theta^2)\frac{1}{n}\sum_{i=1}^{n}\Phi^{-1}(\mathbb{G}_n^*(Y_i))\Phi^{-1}(\mathbb{H}_n^*(Z_i))$$

$$= -\theta(1 + \theta^2) + o(n^{-1/2}) + (1 + \theta^2)\frac{1}{n}\sum_{i=1}^{n}\Phi^{-1}(\mathbb{G}_n^*(Y_i))\Phi^{-1}(\mathbb{H}_n^*(Z_i)), \tag{3.40}$$

which shows that

$$\hat{\theta}_n^{\mathrm{pml}} = o(n^{-1/2}) + \frac{1}{n}\sum_{i=1}^{n}\Phi^{-1}(\mathbb{G}_n^*(Y_i))\Phi^{-1}(\mathbb{H}_n^*(Z_i)) = \hat{\rho}_n + o(n^{-1/2}). \tag{3.41}$$

Hence even the pseudo-maximum likelihood estimator $\hat{\theta}_n^{\mathrm{pml}}$ is asymptotically efficient in our present normal copula model. The pseudo-maximum likelihood method has been studied more generally in the context of copula models by Genest *et al.* (1995), who prove

asymptotic normality of $\hat{\theta}_n^{\mathrm{pml}}$. When specialized to the normal copula model, their asymptotic variance formula yields $(1 - \theta^2)^2$ in agreement with the preceding argument, as has been shown by Hu (1995).

4. Approach 2: information calculations for copula models

In the framework of Section 1, suppose that the copula distributions C_θ have densities c_θ with respect to Lebesgue measure on $[0, 1]^2$. We assume that $c_\theta^{1/2}$ is Fréchet differentiable in θ in the Hilbert space of square Lebesgue-integrable functions on the unit square. The resulting Fréchet derivative multiplied by $2c_\theta^{-1/2} 1_{[c_\theta > 0]}$ is called the *score function* for θ and denoted by $\dot{l}_\theta$.

We will follow the development in Section 4.7 of Bickel *et al.* (1993), especially Propositions 4–7 (pp. 166–169), together with Proposition A.4.1 (p. 439). For copula models with two unknown marginal distributions, the equations determining the projection of the score function for θ onto the nuisance parameter tangent space given in general by (A.4.11)–(A.4.13) can be written as

$$\dot{l}_g \mathrm{a} = \Pi_g(\dot{l}_\theta - \dot{l}_h b) \tag{4.42}$$

and

$$\dot{l}_h b = \Pi_h(\dot{l}_\theta - \dot{l}_g a) \tag{4.43}$$

where $\Pi_g = \dot{l}_g (\dot{l}_g^{\mathrm{T}} \dot{l}_g)^{-1} \dot{l}_g^{\mathrm{T}}$ is the projection operator onto $\dot{\mathscr{P}}_g$, the tangent space of score functions for g, and $\Pi_h = \dot{l}_h (\dot{l}_h^{\mathrm{T}} \dot{l}_h)^{-1} \dot{l}_h^{\mathrm{T}}$ is the projection operator onto $\dot{\mathscr{P}}_h$; here

$$\dot{l}_g a(u, v) = a(u) + \dot{l}_u(u, v) \int_0^u a(s) \mathrm{d}s, \tag{4.44}$$

$$\dot{l}_h b(u, v) = b(v) + \dot{l}_v(u, v) \int_0^v b(s) \mathrm{d}s \tag{4.45}$$

for $a, b \in L_2^0([0, 1], \text{Lebesgue})$ with

$$\dot{l}_u(u, v) = \frac{\partial}{\partial u} \log\ c_\theta(u, v), \qquad \dot{l}_v(u, v) = \frac{\partial}{\partial v} \log\ c_\theta(u, v),$$

while

$$\dot{l}_g^{\mathrm{T}} a(u) = \int_0^1 a(u, v) c_\theta(u, v) \mathrm{d}v + \int_0^1 \int_0^1 (1_{[u \leqslant s]} - s) \dot{l}_u(s, v) a(s, v) c_\theta(s, v) \mathrm{d}s \mathrm{d}v, \tag{4.46}$$

$$\dot{l}_h^{\mathrm{T}} b(u) = \int_0^1 b(u, v) c_\theta(u, v) \mathrm{d}u + \int_0^1 \int_0^1 (1_{[v \leqslant s]} - s) \dot{l}_v(u, s) b(u, s) c_\theta(u, s) \mathrm{d}u \mathrm{d}s \tag{4.47}$$

for $a, b \in L_2^0(C_\theta)$. Actually, we only know that the nuisance tangent space contains $\dot{\mathscr{P}}_g + \dot{\mathscr{P}}_h$ and that

$$\dot{\mathscr{P}}_g \supset [\dot{l}_g a\colon a \in L_2^0([0,1], \text{Lebesgue})] = \mathscr{R}(\dot{l}_g)$$

and

$$\dot{\mathscr{P}}_h \supset [\dot{l}_h b\colon b \in L_2^0([0,1], \text{Lebesgue})] = \mathscr{R}(\dot{l}_h).$$

Consequently, in (4.42) and (4.43), Π_g and Π_h describe projections onto possibly proper subspaces of the nuisance tangent spaces $\dot{\mathscr{P}}_g$ and $\dot{\mathscr{P}}_h$, respectively. However this may be, the resulting projection onto a subspace of the nuisance tangent space will yield a valid information bound for our semi-parametric model, and, in fact, it will yield the efficient information bound and corresponding efficient score function (4.70) (cf. (3.29)); see the discussion in Bickel *et al.* (1993, pp. 76–77). Define two functions α, β by

$$\alpha(u) = \int_0^1 \dot{l}_u^2(u, v) c_\theta(u, v) \mathrm{d}v, \qquad \beta(v) = \int_0^1 \dot{l}_v^2(u, v) c_\theta(u, v) \mathrm{d}u. \tag{4.48}$$

Note that by Proposition 4.7.6. of Bickel *et al.* (1993, p. 168), the sum space $\mathscr{R}(\dot{l}_g) + \mathscr{R}(\dot{l}_h)$ is closed if α and β satisfy

$$\alpha(u) \leqslant M\{u(1-u)\}^{-2}, \qquad 0 < u < 1,$$

and

$$\beta(v) \leqslant M\{v(1-v)\}^{-2}, \qquad 0 < v < 1.$$

Operating across (4.42) by $\dot{l}_g^{\mathrm{T}}$ yields

$$\dot{l}_g^{\mathrm{T}} \dot{l}_g a = \dot{l}_g^{\mathrm{T}} \dot{l}_g (\dot{l}_g^{\mathrm{T}} \dot{l}_g)^{-1} \dot{l}_g^{\mathrm{T}} (\dot{l}_\theta - \dot{l}_h b) \tag{4.49}$$

$$= \dot{l}_g^{\mathrm{T}} \dot{l}_\theta - \dot{l}_g^{\mathrm{T}} \dot{l}_h b. \tag{4.50}$$

Here

$$\dot{l}_g^{\mathrm{T}} \dot{l}_g a(u) = a(u) + \int_0^1 (1_{[u \leqslant s]} - s) \left\{ \int_0^s a \mathrm{d}I \right\} \alpha(s) \mathrm{d}s \tag{4.51}$$

for $a \in L_2^0(G)$, and

$$\dot{l}_g^{\mathrm{T}} \dot{l}_\theta(u) = \int_0^1 (1_{[u \leqslant s]} - s) \gamma(s) \mathrm{d}s, \tag{4.52}$$

with α as defined in (4.48) and

$$\gamma(u) = \int_0^1 \dot{l}_\theta \dot{l}_u(u, v) c_\theta(u, v) \mathrm{d}v.$$

So differentiation across (4.50) yields, with $A' = a$, $A'' = a'$,

$$A''(u) - \alpha(u)A(u) = -\gamma(u) - \frac{\partial}{\partial u}\{i_g^T i_h b\}. \tag{4.53}$$

To calculate the last term, we first calculate $i_g^T i_h b$. By (4.46) and the formula (4.45) for i_h, we obtain, with $B(v) = \int_0^v b(s)\mathrm{d}s$,

$$\begin{aligned}
i_g^T(i_h b)(u) &= \int_0^1 b(v)c_\theta(u,v)\mathrm{d}v \\
&\quad + \int_0^1 \dot{l}_v(u,v)B(v)c_\theta(u,v)\mathrm{d}v \\
&\quad + \int_0^1 b(v)\left\{\int_0^1 (1_{[u \leqslant s]} - s)\dot{l}_u(s,v)c_\theta(s,v)\mathrm{d}s\right\}\mathrm{d}v \\
&\quad + \int_0^1\int_0^1 (1_{[u \leqslant s]} - s)\dot{l}_u(s,v)\dot{l}_v(s,v)B(v)c_\theta(s,v)\mathrm{d}s\mathrm{d}v.
\end{aligned} \tag{4.54}$$

Differentiation of (4.54) with respect to u yields

$$\frac{\partial}{\partial u}\{i_g^T i_h b\}(u) = \int_0^1 B(v)\ddot{l}_{uv}(u,v)c_\theta(u,v)\mathrm{d}v. \tag{4.55}$$

By symmetry, we obtain

$$\frac{\partial}{\partial v}\{i_h^T i_g a\}(v) = \int_0^1 A(u)\ddot{l}_{uv}(u,v)c_\theta(u,v)\mathrm{d}u. \tag{4.56}$$

Let $K(u,v) = \ddot{l}_{uv}(u,v)c_\theta(u,v)$; then the coupled equations (4.42) and (4.43) become:

$$A''(u) - \alpha(u)A(u) = -\gamma(u) - \int_0^1 B(v)K(u,v)\mathrm{d}v, \tag{4.57}$$

$$B''(v) - \beta(v)B(v) = -\delta(v) - \int_0^1 A(u)K(u,v)\mathrm{d}u \tag{4.58}$$

with β as defined in (4.48) and

$$\delta(v) = \int_0^1 \dot{l}_\theta \dot{l}_v(u,v)c_\theta(u,v)\mathrm{d}u.$$

To this point, our development has involved rewriting the equations determining the projection of $\dot{l}_\theta$ onto the sum space $\mathscr{R}(\dot{l}_g) + \mathscr{R}(\dot{l}_h)$ for a general bivariate copula model. Now we specialize to the case of the normal copula family given by (1.7). In this case the corresponding density c_θ is (with ϕ_θ denoting the density of Φ_θ)

$$c_\theta(u,v) = \frac{\phi_\theta(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}. \tag{4.59}$$

Then we obtain by straightforward calculation that (see also (4.7.92) and (4.7.93) in Bickel *et al.* 1993, p. 174)

$$\alpha(u) = \beta(u) = \frac{\theta^2}{1-\theta^2} \frac{1}{\phi^2(\Phi^{-1}(u))}, \tag{4.60}$$

$$\gamma(u) = \delta(u) = \frac{\theta}{1-\theta^2} \frac{\Phi^{-1}(u)}{\phi(\Phi^{-1}(u))} \tag{4.61}$$

and

$$\ddot{l}_{\mathrm{uv}}(u,v) = \frac{\theta}{1-\theta^2} \frac{1}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}. \tag{4.62}$$

Note that $\alpha(u) = o([u(1-u)]^{-2}$ as $u \to 0$ or $u \to 1$, and (a natural generalization of) Proposition 4.7.6 of Bickel *et al.* (1993, p. 168) holds and shows that $\mathscr{R}(\dot{l}_g) + \mathscr{R}(\dot{l}_h)$ is closed. Calculations will become simpler and more transparent in this present case if we transform back to y and z corresponding to normal marginal distributions, so we define $\tilde{A}$ and $\tilde{B}$ by

$$\tilde{A}(y) = A(\Phi(y)), \qquad A(u) = \tilde{A}(\Phi^{-1}(u)), \tag{4.63}$$

$$\tilde{B}(z) = B(\Phi(z)), \qquad B(v) = \tilde{B}(\Phi^{-1}(v)). \tag{4.64}$$

Then

$$A'(u) = \tilde{A}'(\Phi^{-1}(u)) \frac{1}{\phi(\Phi^{-1}(u))} \tag{4.65}$$

and

$$A''(u) = \tilde{A}''(\Phi^{-1}(u)) \frac{1}{\phi^2(\Phi^{-1}(u))} + \tilde{A}'(\Phi^{-1}(u)) \frac{\Phi^{-1}(u)}{\phi^2(\Phi^{-1}(u))}. \tag{4.66}$$

Using these in (4.57) and letting $y = \Phi^{-1}(u)$, we obtain a differential equation for $\tilde{A}$ with a coupling term involving $\tilde{B}$:

$$\tilde{A}''(y) + y\tilde{A}'(y) - \frac{\theta^2}{1-\theta^2} \tilde{A}(y)$$
$$= -\frac{\theta}{1-\theta^2} y\phi(y) - \int_{-\infty}^{\infty} \tilde{B}(z) \frac{\theta}{1-\theta^2} \frac{\phi_\theta(y,z)}{\phi(z)} \, \mathrm{d}z. \tag{4.67}$$

By symmetry, equation (4.58) becomes

$$\tilde{B}''(z) + z\tilde{B}'(z) - \frac{\theta^2}{1-\theta^2} \tilde{B}(z)$$
$$= -\frac{\theta}{1-\theta^2} z\phi(z) - \int_{-\infty}^{\infty} \tilde{A}(y) \frac{\theta}{1-\theta^2} \frac{\phi_\theta(y,z)}{\phi(y)} \, \mathrm{d}y. \tag{4.68}$$

To solve equations (4.67) and (4.68), we simply 'guess' the answer up to a constant c, and then solve for c: by taking $\tilde{A}(y) = cy\phi(y) = -c\phi'(y)$, $\tilde{B}(z) = cz\phi(z) = -c\phi'(z)$, it is easily

checked that $\tilde{A}$, $\tilde{B}$ satisfy (4.67) and (4.68) for $c = 2^{-1}\theta/(1 - \theta^2)$. This yields the efficient score function l_θ^* for θ:

$$l_\theta^*(u, v) = \dot{l}_\theta(u, v) - \dot{l}_g a(u, v) - \dot{l}_h b(u, v); \tag{4.69}$$

again it is a little bit easier to continue calculation on the y, z scales, and indeed we find, upon substitution, that

$$l_\theta^*(y, z) \equiv l_\theta^*(\Phi(y), \Phi(z)) = \frac{1}{(1 - \theta^2)^2} \left(yz - \frac{\theta}{2}(y^2 + z^2) \right). \tag{4.70}$$

Hence, with $(Y, Z) \sim \Phi_\theta$, the efficient information for θ in the bivariate normal copula model (1.7) is given by

$$I_*(\theta) = \mathrm{E}\{l_\theta^*(Y, Z)\}^2 = \frac{1}{(1 - \theta^2)^4} \mathrm{E} \left\{ YZ - \frac{\theta}{2}(Y^2 + Z^2) \right\}^2 = \frac{1}{(1 - \theta^2)^2}. \tag{4.71}$$

As already shown in Section 3, this means that the normal scale submodel of the bivariate normal copula is least favourable for estimation of θ.

It should be emphasized that the short proof given in Section 3 was found only *after* we had performed the calculations presented in this section. Furthermore, we do not know solutions of the projection equations (4.57) and (4.58) for any other copula model. For example, it would be of interest to know more about the solution of (4.57) and (4.58) for the Clayton–Oakes and Frank models with

$$C_\theta(u, v) = \psi_\theta^{-1}(\psi_\theta(u) + \psi_\theta(v)) \tag{4.72}$$

and $\psi_\theta(u) = (u^{-\theta} - 1)/\theta$ or $\psi_\theta(u) = \log\{(1 - \theta)/(1 - \theta^u)\}$, respectively. Such calculations may be possible via calculation of eigenfunctions and eigenvalues of the integral operator(s) with kernel K appearing in (4.57) and (4.58). In the normal copula model considered here,

$$K(u, v) = \frac{\theta}{1 - \theta^2} \frac{\phi_\theta(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi^2(\Phi^{-1}(u))\phi^2(\Phi^{-1}(v))},$$

and hence Mehler's (1866) formula (6.89), which we will discuss in Section 6, yields an eigenexpansion of K (composed on Φ in each argument) and of the integral operators in (4.57) and (4.58).

5. Miscellanea

5.1. EXTENSION FROM m = 2 TO GENERAL m ⩾ 2

Now suppose that $X = (Y_1, \ldots, Y_m) \sim N_m(0, \Sigma)$ where $\Sigma \in \mathscr{S}$ can be regarded as a vector in $\mathbb{R}^{m(m+1)/2}$. It is well known that the maximum likelihood estimator $\hat{\Sigma}$ is an (asymptotically) efficient estimator of Σ in this regular parametric model. Since the population correlation coefficient ρ_{12} is a differentiable function of Σ, say $\rho_{12} = \psi_{12}(\Sigma)$, $\psi_{12}(\hat{\Sigma})$ is efficient in estimating $\psi_{12}(\Sigma) = \rho_{12}$. But $\hat{\Sigma}$ is the sample covariance matrix, and hence

$\psi_{12}(\hat{\Sigma})$ is the sample correlation coefficient. Consequently, the efficient influence function in estimating ρ_{12} equals the influence function of $\psi_{12}(\hat{\Sigma})$, which (as we know) equals the efficient influence function for estimating ρ_{12} within the bivariate normal model based on observing $X = (Y_1, Y_2)$. It follows that the results of Sections 3 and 4 for the case of $m = 2$ carry over immediately to the case $m > 2$: the normal scores rank correlation coefficient is (asymptotically) efficient for estimation of ρ_{12}, and similarly for the other correlation coefficients $\rho_{13}, \ldots, \rho_{(m-1)m}$.

5.2. DISCONTINUITY OF ρ_{M} AND ρ_{m}

One issue which appeared in Section 2 is that of identifying useful extensions of the parameter $\nu(P_{\theta,G,H}) = \theta$ beyond the normal copula model $\mathscr{P}$. As noted in Section 2 (and discussed further in Section 6), the maximum correlation coefficient equals $|\theta|$ on the normal copula model, so the maximum correlation coefficient $\rho_{\mathrm{M}}(P)$ gives an extension of $|\nu(P)|$ beyond the normal copula model $\mathscr{P}$. Similarly, the maximum monotone correlation coefficient $\rho_{\mathrm{m}}(P)$ extends $|\nu(P)|$ too.

Let the model $\mathscr{P}_{\mathrm{e}} \supset \mathscr{P}$ be an extension of $\mathscr{P}$. For the maximum correlation coefficient $\rho_{\mathrm{M}}(P)$ to be consistently estimable (uniformly on compact subsets of $\mathscr{P}_{\mathrm{e}}$ in the variational distance), it is necessary that $\rho_{\mathrm{M}}(P)$ be continuous on $\mathscr{P}_{\mathrm{e}}$; see, for example, Proposition 2.2.1.A of Bickel *et al.* (1993, p. 20). If $\mathscr{P}_{\mathrm{e}}$ is a sufficiently large extension of $\mathscr{P}$ this is not the case, as we will show in this subsection. In fact, we will prove the stronger result that both $\rho_{\mathrm{M}}(P)$ and $\rho_{\mathrm{m}}(P)$ are discontinuous on appropriate extensions $\mathscr{P}_{\mathrm{e}}$ of the core model of $\mathscr{P}$, that is, the normal model. Indeed, let $\mathscr{P}_{\mathrm{e}}$ be the class of all distributions on $\mathbb{R}^2$ with smooth density with respect to Lebesgue measure. Many definitions of smoothness will do for our proof of Theorem 5.1 below, for example if all partical derivatives of any order of the density exist.

Theorem 5.1 (Discontinuity of ρ_{M} and ρ_{m}). *Both $\rho_{\mathrm{M}}(P)$ and $\rho_{\mathrm{m}}(P)$ are (weakly) discontinuous functionals on $\mathscr{P}_{\mathrm{e}}$ at any $P_0 \in \mathscr{P}_{\mathrm{e}}$ with $\rho_{\mathrm{M}}(P_0) < 1$ and $\rho_{\mathrm{m}}(P_0) < 1$, respectively. Furthermore, $\rho_{\mathrm{M}}(P)$ and $\rho_{\mathrm{m}}(P)$ are lower semi-continuous on $\mathscr{P}_{\mathrm{e}}$ and hence continuous at those P_0 with $\rho_{\mathrm{M}}(P_0) = 1$ and $\rho_{\mathrm{m}}(P_0) = 1$, respectively.*

Remark 5.1. Discontinuity of ρ_{M} at any bivariate distribution with independent marginals was proved by Kimeldorf and Sampson (1978): their Theorem 1 (p. 897) exhibits a sequence of distributions P_n on the unit square in $\mathbb{R}^2$ (now known as 'shuffles of min' distributions) which satisfy:

- $\rho_{\mathrm{M}}(P_n) = 1$ for all $n = 1, 2, \ldots$.
- $P_n \to_{\mathrm{d}} Uniform([0,1]^2)$.

Since $\rho_{\mathrm{M}}(Uniform([0,1]^2)) = 0$, this proves that ρ_{M} is discontinuous at 'independence'.

This was strengthened by Mikusiński *et al.* (1991), who show that shuffles of min are dense in the collection of all copulas on $[0,1]^2$ in the sense of Kolmogorov (supremum

norm) distance between distribution functions. Stated another way, this says that for copula C on $[0, 1]^2$ with arbitrary maximal correlation $\rho_M(C)$, there exists a sequence of copulas $\{C_n\}$ on $[0, 1]^2$ with $\rho_M(C_n) = 1$ for all n and $\| C_n - C \|_\infty \to 0$. Since $\| C_n - C \|_\infty \to 0$ implies that $C_n \to_d C$, this implies that ρ_M is weakly discontinuous at every copula C, and hence also at every bivariate distribution P.

Now we turn to ρ_m. Preservation of $\rho_m(P) = 1$ under weak convergence was proved by Kimeldorf and Sampson (1978): their Theorem 3 (p. 899) can be rephrased as follows. If $P_n \to_d P$ and $\rho_m(P_n) = 1$ for each n, then $\rho_m(P) = 1$. (This is *not* the same as continuity of ρ_m at 1, which would assert that if $P_n \to_d P$ and $\rho_m(P) = 1$, then $\rho_m(P_n) \to 1$.)

Proof. First consider $\rho_m(P)$. Fix P_0 and $\delta > 0$ so that $\rho_m(P_0) + \delta < 1$. Let a, b be monotone functions satisfying

$$\rho_m(P_0) \leqslant \int a(y)b(z)\mathrm{d}P_0(y, z) + \delta \tag{5.73}$$

with

$$\mathrm{E}_0 a(Y) = \mathrm{E}_0 b(z) = 0, \qquad \mathrm{E}_0 a^2(Y) = \mathrm{E}_0 b^2(Z) = 1. \tag{5.74}$$

Without loss of generality we may assume that both $a : \mathbb{R} \to \bar{\mathbb{R}}$ and $b : \mathbb{R} \to \bar{\mathbb{R}}$ are continuous and unbounded on the support of Y and Z, respectively. We may even assume them to be strictly increasing. (If $a : \mathbb{R} \to \bar{\mathbb{R}}$ is not continuous, unbounded, or strictly increasing, first truncate a to obtain $a_1(y) = (-M) \vee (a(y) \wedge M)$ with M sufficiently large such that $\int(a_1 - a)^2 \mathrm{d}F$ is sufficiently small; here F is the marginal distribution function of Y. Second, convolve a_1 with a smooth density, getting $a_2(y) = \int a_1(y + \varepsilon u)k(u)\mathrm{d}u$; here k is, for instance, the logistic density. If $\varepsilon > 0$ is sufficiently small, then $\int(a_2 - a_1)^2 \mathrm{d}F$ is small. Note that a_2 is continuous and strictly increasing unless $a = 0$. Furthermore, define $a_3(y) = a_2(y) + \varepsilon\{(F(y))^{-1/4} + (1 - F(y))^{-1/4}\}$, for y in the support of Y, with ε sufficiently small. Finally, normalize a_3 such that the resulting a_4 satisfies (5.74). For appropriate choices M and ε it also satisfies (5.73).)

For $\varepsilon > 0$ we define the sets A_ε and B_ε by

$$A_\varepsilon = \{(y, z) : \varepsilon^{-2} \leqslant a(y) \leqslant \varepsilon^{-2}(1 + \varepsilon^4)^{1/2}, \varepsilon^{-2} \leq b(z) \leq \varepsilon^{-2}(1 + \varepsilon^4)^{1/2}\}, \tag{5.75}$$

$$B_\varepsilon = \{(y, z) : -\varepsilon^{-2}(1 + \varepsilon^4)^{1/2} \leqslant a(y) \leqslant -\varepsilon^{-2}, -\varepsilon^{-2}(1 + \varepsilon^4)^{1/2} \leqslant b(z) \leqslant -\varepsilon^{-2}\}. \tag{5.76}$$

We choose $P_{1\varepsilon} \in \mathscr{P}_e$ such that

$$P_{1\varepsilon}(A_\varepsilon) = P_{1\varepsilon}(B_\varepsilon) = \tfrac{1}{2}. \tag{5.77}$$

This may be done in such a way that both $a(Y)$ and $b(Z)$ have mean 0 under $P_{1\varepsilon}$. Note that

$$\mathrm{E}_{1\varepsilon} a^2(Y) \leqslant \varepsilon^{-4}(1 + \varepsilon^4), \qquad \mathrm{E}_{1\varepsilon} a(Y)b(Z) \geqslant \varepsilon^{-4}. \tag{5.78}$$

Now we define $P_\varepsilon \in \mathscr{P}_e$ by $P_\varepsilon = (1 - \varepsilon^4)P_0 + \varepsilon^4 P_{1\varepsilon}$. Then $\mathrm{E}_\varepsilon a(Y) = \mathrm{E}_\varepsilon b(Z) = 0$, and by (5.78) we obtain

$$\rho_m(P_\varepsilon) \geqslant \frac{\mathrm{E}_\varepsilon a(Y)b(Z)}{\sqrt{\mathrm{E}_\varepsilon a^2(Y)\mathrm{E}_\varepsilon b^2(Z)}} \qquad \geqslant \frac{1 - \varepsilon^4)(\rho_m(P_0) - \delta) + 1}{(1 - \varepsilon^4) + (1 + \varepsilon)^4}. \tag{5.79}$$

In view of $\rho_\mathrm{m}(P_0) + \delta < 1$, this yields

$$\liminf_{\varepsilon \downarrow 0} \rho_\mathrm{m}(P_\varepsilon) \geqslant \frac{1 + \rho_\mathrm{m}(P_0) - \delta}{2} > \rho_\mathrm{m}(P_0). \tag{5.80}$$

Since $P_\varepsilon \to_\mathrm{d} P_0$ as $\varepsilon \downarrow 0$, $P \to \rho_\mathrm{m}(P)$ is not weakly continuous at P_0.

The same arguments without the monotonicity restrictions yield a proof for the discontinuity of $\rho_\mathrm{M}(P)$ at P_0.

Again, fix P_0 and $\delta > 0$. Choose bounded continuous monotone functions a and b such that

$$\rho_{P_0}(a(Y), b(Z)) \geqslant \rho_\mathrm{m}(P_0) = \delta. \tag{5.81}$$

For any sequence $\{P_n\}$ converging weakly to P_0 we have

$$\rho_\mathrm{m}(P_n) \geqslant \rho_{P_n}(a(Y), b(Z)) \to \rho_{P_0}(a(Y), b(Z)) \tag{5.82}$$

as $n \to \infty$. Since δ is arbitrary, the second part of the theorem follows from (5.81) and (5.82). Subsequently, the continuity property is implied by $\rho_\mathrm{m}(P) \leqslant 1$. $\square$

5.3. EFFICIENT ESTIMATION OF G AND H?

It would be very interesting to know information bounds and efficient estimators for estimation of the marginal distribution functions G and H in the bivariate normal copula model treated here, or in other copula models. It is clear that the marginal distributions $\mathbb{G}_n$ and $\mathbb{H}_n$ provide $\sqrt{n}$-consistent estimators of G and H, respectively, but because of the parametric dependence structure these will be inefficient in general. One approach to calculation of an information bound for estimation of $G(x_0)$ for a fixed x_0 involves projection of the influence function $1_{(-\infty, x_0]} - G(x_0)$ of the inefficient estimator $\mathbb{G}_n(x_0)$ onto the tangent space $\dot{\mathscr{P}} = [\dot{l}_\theta] + \dot{\mathscr{P}}_g + \dot{\mathscr{P}}_h$ of the model. We have not yet succeeded in solving the (coupled) differential equations connected to this projection problem.

6. Mehler's formula and the Gebelein–Lancaster theorem

In his well-known paper 'On measures of dependence', Rényi (1959) includes the following postulate among his list of requirements of a measure of dependence $\delta(Y, Z)$ between two random variables Y and Z: 'If the joint distribution of Y and Z is normal, then $\delta(Y, Z) = |\rho(Y, Z)|$ where $\rho(Y, Z)$ is the correlation coefficient of Y and Z.' In his discussion of the

maximum correlation ρ_M, Rényi attributes the verification of this postulate in the case of ρ_M to Gebelein (1941). This seems to be one thread in what we shall call the 'continental European' history of the maximum correlation and its properties, which seems to have begun with Gebelein (1941) and continued with Richter (1949), Sarmanov (1958), Rényi (1959) and Bell (1962).

On the other hand, there was a strong development of maximum correlation (or canonical correlation) in England, especially for discrete variables. The introduction of correlation seems to have begun with Galton (1888) and Pearson (1896); see Stigler (1986, pp. 297–299 and 342). Study of maximum correlation took off with Hirschfeld (1935), Fisher (1940) and Maung (1942). Lancaster (1957) renewed the investigation, and independently proved Gebelein's result in Lancaster (1958) (where he references Mehler 1866). Lancaster (1963) and Eagleson (1964) contain related results, and by Lancaster (1969) the 'continental' and English developments have united: Lancaster (1969) references Rényi (1959).

To the best of our knowledge, the only textbook containing a proof of the theorem noted by Rényi (1959) is Kendall and Stuart (1973), where the theorem is attributed to Lancaster (1957).

It seems fairly clear that one major reason for the lack of contact between the two literatures ('continental Europe' and 'English') was the Second World War. That it took until 1969 for this contact or bridging to occur attests to the depth of the division. Because the proof in Kendall and Stuart is quite brief and apparently not well known, we include a proof here of the theorem due to Gebelein (1941) and Lancaster (1957).

Theorem 6.1. *If the joint distribution of X and Y is normal, then $\rho_M(X, Y) = |\rho(X, Y)|$. Moreover, the supremum is attained for (and only for) linear transformations a and b of X and Y, respectively.*

Proof. The following formulae appear in Mehler (1866), pp. 173–174):

$$E(x, y) = \frac{1}{\sqrt{1 - \rho^2}} \exp\left(\frac{-\rho^2(x^2 + y^2) + 2\rho xy}{1 - \rho^2}\right)$$

$$= \frac{e^{x^2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(t - iy)^2 - (x + i\rho t)^2} dt$$

$$= \frac{e^{x^2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(t - iy)^2} \sum_{n=0}^{\infty} \frac{(i\rho t)^n}{n!} \frac{\partial^n}{\partial x^n} e^{-x^2} dt \tag{6.83}$$

$$= \sum_{n=0}^{\infty} \frac{(2\rho)^n}{n!} \tilde{H}_n(x) \tilde{H}_n(y),$$

with

$$\tilde{H}_n(x) = (-2)^{-n} e^{x^2} \frac{\partial^n}{\partial x^n} e^{-x^2} = \frac{(-i)^n}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^n e^{-(t - ix)^2} dt. \tag{6.84}$$

The Hermite polynomials $H_n(x)$ are defined (cf. Feller 1971, p. 532) by

$$H_n(x) = (-1)^n \frac{1}{\phi(x)} \frac{\partial^n}{\partial x^n} \phi(x) \tag{6.85}$$

where $\phi(x) = (2\pi)^{-1/2} e^{-x^2/2}$ is the standard normal density. So

$$\tilde{H}_n\left(\frac{1}{2}\sqrt{2}x\right) = (-2)^{-n} \frac{1}{\phi(x)} \frac{\partial^n}{\partial x^n} \phi(x) 2^{n/2} = 2^{-n/2} H_n(x), \tag{6.86}$$

and

$$E\left(\frac{1}{2}\sqrt{2}x, \frac{1}{2}\sqrt{2}y\right) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho^2(x^2+y^2) - 2\rho xy}{2(1-\rho^2)}\right) \tag{6.87}$$

$$= \sum_{n=0}^{\infty} \frac{\rho^n}{n!} H_n(x) H_n(y). \tag{6.88}$$

If $(X, Y) \sim \Phi_\rho$ then the density ϕ_ρ of (X, Y) equals

$$\phi_\rho(x,y) = \frac{1}{2\pi} E\left(\frac{1}{2}\sqrt{2}x, \frac{1}{2}\sqrt{2}y\right) \exp\left(-\frac{(1-\rho^2)(x^2+y^2)}{2(1-\rho^2)}\right)$$

$$= E\left(\frac{1}{2}\sqrt{2}x, \frac{1}{2}\sqrt{2}y\right) \phi(x)\phi(y)$$

$$= \left\{1 + \sum_{n=1}^{\infty} \frac{\rho^n}{n!} H_n(x) H_n(y)\right\} \phi(x)\phi(y)$$

$$= \left\{1 + \sum_{n=1}^{\infty} \rho^n H_n^*(x) H_n^*(y)\right\} \phi(x)\phi(y), \tag{6.89}$$

where $H_n^* \equiv (n!)^{-1/2} H_n$, $n = 0, 1, 2, \ldots$, the normalized Hermite polynomials, form a complete orthonormal system with respect to ϕ (see, for example, Abramowitz and Stegun 1972, Chapter 22). Either of these last two expansions is commonly known now as *Mehler's expansion* or *Mehler's formula*. For further information on general expansions of this type and references to related literature, see Buja (1990).

Now we turn to the maximum correlation coefficient

$$\rho_M(X, Y) = \sup_{a,b} \rho(a(X), b(Y)). \tag{6.90}$$

For functions a, b with $\text{var}(a(X)) < \infty$, $\text{var}(b(Y)) < \infty$, we can expand in terms of the normalized Hermite polynomials H_n^*:

$$a(x) = \sum_{m=0}^{\infty} \alpha_m H_m^*(x), \qquad b(y) = \sum_{n=0}^{\infty} \beta_n H_n^*(y). \tag{6.91}$$

Then, using Mehler's expansion (6.89),

$$\text{cov}_\rho(a(X), b(Y)) = \text{cov}_\rho\left(\sum_{m=1}^{\infty} \alpha_m H_m^*(X), \sum_{n=1}^{\infty} \beta_n H_n^*(Y)\right) \tag{6.92}$$

$$= \text{E}\left\{\sum_{m=1}^{\infty} \alpha_m H_m^*(U) \sum_{n=1}^{\infty} \beta_n H_n^*(V) \sum_{j=1}^{\infty} \rho^j H_j^*(U) H_j^*(V)\right\} \tag{6.93}$$

where U and V are independent $N(0,1)$ random variables. By the orthonormality of the h_m^*s this yields

$$\text{cov}_\rho(a(X), b(Y)) = \sum_{n=1}^{\infty} \alpha_n \beta_n \rho^n. \tag{6.94}$$

Since the marginals are standard normal,

$$\text{var}(a(X)) = \text{E}\left(\sum_{m=1}^{\infty} \alpha_m H_m^*(X)\right)^2 = \sum_{m=1}^{\infty} \alpha_m^2, \tag{6.95}$$

$$\text{var}(b(Y)) = \sum_{n=1}^{\infty} \beta_n^2. \tag{6.96}$$

Consequently

$$\rho(a(X), b(Y)) = \frac{\sum_{n=1}^{\infty} \alpha_n \beta_n \rho^n}{\left\{\sum_{m=1}^{\infty} \alpha_m^2 \sum_{n=1}^{\infty} \beta_n^2\right\}^{1/2}}. \tag{6.97}$$

By the Cauchy–Schwarz inequality this yields

$$\rho(a(X), b(Y)) \leqslant \left\{\frac{\sum_{n=1}^{\infty} \beta_n^2 (\rho^2)^n}{\sum_{n=1}^{\infty} \beta_n^2}\right\}^{1/2} \leqslant |\rho|, \tag{6.98}$$

where the last inequality is an equality if $\beta_n = 0, n \geqslant 2, \beta_1 \neq 0$. Consequently

$$\sup_{a,b} \rho(a(X), b(Y)) = |\rho| = \rho(X, Y) \vee \rho(X, -Y) \tag{6.99}$$

where equality holds if a and b are linear functions of X and Y, respectively. □

Acknowledgements

References

Abramowitz, M. and Stegun, I.A. (1972) *Handbook of Mathematical Functions* (9th edn). New York: Dover.

Barron, A.R. (1986) Entropy and the central limit theorem. *Ann. Probab.*, **14**, 336–342.

Bell, C.B. (1962) Mutual information and maximal correlation as measures of dependence. *Ann. Math. Statist.*, **33**, 587–595.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993) *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: Johns Hopkins University Press.

Borovkov, A.A. and Utev, A. (1983) On an inequality and a related characterization of the normal distribution. *Theory Probab. Appl.*, **28**, 209–218.

Buja, A. (1990) Remarks on functional canonical variates, alternating least squares methods and ACE. *Ann. Statist.*, **18**, 1032–1069.

Chernoff, H. (1981) A note on an inequality involving the normal distribution. *Ann. Probab.*, **9**, 533–535.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141–151.

Clayton, D. and Cuzick, J. (1985) Multivariate generalizations of the proportional hazards model (with discussion). *J. Roy. Statist. Soc. Ser B*, **34**, 187–220.

Clayton, D. and Cuzick, J. (1986) The semiparametric Pareto model for regression analysis of survival times. *Papers on Semiparametric Models* MS-R8614, pp. 19–31. Amsterdam: Centrum voor Wiskunde en Informatica.

Eagleson, G.K. (1964) Polynomial expansions of bivariate distributions. *Ann. Math. Statist.*, **35**, 1208–1215.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, Volume II (2nd edn). New York: Wiley.

Fisher, R.A. (1940) The precision of discriminant functions. *Ann. Eugenics London*, **10**, 422–429.

Fréchet, M. (1943) Sur l'extension de certaines evaluations statistiques au cas de petits échantillons. *Rev. Inst. Internat. Statist.* **11**, 182–205.

Galton, F. (1888) Co-relations and their measurement, chiefly from anthropological data. *Proc. Roy. Soc. London*, **45**, 135–145.

Gebelein, H. (1941) Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. *Zeit. Angew. Math. Mech.*, **21**, 364–379.

Genest, C. (1987) Frank's family of bivariate distributions. *Biometrika*, **74**, 549–555.

Genest, C. and MacKay, J. (1986a) The joy of copulas: bivariate distributions with uniform marginals. *Amer. Statist.*, **40**, 280–283.

Genest, C. and MacKay, R.J. (1986b) Copules archimédiennes et familles de lois bidimensionelles dont les marges sont données. *Canad. J. Statist.*, **14**, 145–159.

Genest, C., Ghoudi, K. and Rivest, L.-P. (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**, 543–552.

Hájek, J. and Šidák, Z. (1967) *Theory of Rank Tests*. Prague: Academia.

Hirschfeld, A.O. (1935) A connection between correlation and contingency. *Proc. Cambridge Phil. Soc.*, **31**, 520–524.

Hoeffding, W. (1940) Maßstabinvariante Korrelationstheorie. *Schriften Math. Inst. Angew. Math. Univ. Berlin*, **5**, 179–233.

Hoeffding, W. (1994) *The Collected Works of Wassily Hoeffding* (N.I. Fisher and P.K. Sen, eds). New York: Springer-Verlag.

Hu, Hui-Lin (1995) Large sample theory for pseudo-maximum likelihood estimates in semiparametric models. Ph.D. thesis, in progress.

Huber, P.J. (1981) *Robust Statistics*. New York: Wiley.

Joe, H. (1993) Parametric families of multivariate distributions with given marginals. *J. Multivariate Anal.*, **46**, 262–282.

Kendall, M.G. and Stuart, A. (1973) *The Advanced Theory of Statistics*, Vol. 2 (3rd edn). New York: Hafner.

Kimeldorf, G. and Sampson, A.R. (1975a) One parameter families of bivariate distributions with fixed marginals. *Comm. Statist. Theory Methods*, **4**, 293–301.

Kimeldorf, G. and Sampson, A.R. (1975b) Uniform representation of bivariate distributions. *Comm. Statist. Theory Methods*, **4**, 617–627.

Kimeldorf, G. and Sampson, A.R. (1978) Monotone dependence. *Ann. Statist.*, **6**, 895–903.

Klaassen, C.A.J. (1981) *Statistical Performance of Location Estimators*. Math. Centre Tract 133. Amsterdam: Centrum voor Wiskunde en Informatica.

Klaasen, C.A.J. (1985) On an inequality of Chernoff. *Ann. Probab.*, **13**, 966–974.

Klaassen, C.A.J. (1988) Efficient estimation in the Clayton–Cuzick model for survival data. Preprint, University of Leiden.

Lancaster, H.O. (1957) Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika*, **44**, 289–292.

Lancaster, H.O. (1958) The structure of bivariate distributions. *Ann. Math. Statist.*, **29**, 719–736.

Lancaster, H.O. (1963) Correlations and canonical forms of bivariate distributions. *Ann. Math. Statist.*, **34**, 532–538.

Lancaster, H.O. (1969) *The Chi-squared Distribution*. New York: Wiley.

Marshall, A.W. and Olkin, I. (1988) Families of multivariate distributions. *J. Amer. Statist. Assoc.*, **83**, 934–841.

Maung, K. (1942) Measurement of association in a contingency table with special reference to the pigmentation of hair and eye colors of Scottish school children. *Ann. Eugenics London*, **11**, 189–223.

Mehler, F.G. (1866) Ueber die Entwicklung einer Funktion von beliebeg vielen Variablen nach Laplaceschen Funktionen höherer Ordnung. *J. Reine Angew. Math.*, **66**, 161–176.

Mikusiński, P., Sherwood, H. and Taylor, M.D. (1991) Probabilistic interpretations of copulas. In G. Dall'Aglio, S. Kotz and G. Salinetti (eds), *Advances in Probability Distributions with Given Marginals*. Dordrecht: Kluwer.

Pearson, K. (1896) Mathematical contributions to the theory of evolution, III: on regression, heredity and panmixia. *Philos. Trans. Roy. Soc. London Ser. A*, **187**, 253–318.

Rényi, A. (1959) On measures of dependence. *Acta Math. Acad. Sci. Hungar.*, **10**, 441–451.

Richter, H. (1949) Zur Maximalkorrelation. *Z. Angew. Math. Mech.*, **29**, 127–128.

Ruymgaart, F.H. (1974) Asymptotic normality of nonparametric tests for independence. *Ann. Statist.*, **2**, 892–910.

Ruymgaart, F.H., Shorack, G.R. and Van Zwet, W.R. (1972) Asymptotic normality of nonparametric tests for independence. *Ann. Math. Statist.*, **43**, 1122–1135.

Sarmanov, O.V. (1958) Maximum correlation coefficient (non-symmetrical case). *Dokl. Akad. Nauk*, **121**, 52.

Schweizer, B. (1991) Thirty years of copulas. In G. Dall'Aglio, S. Kotz and G. Salinetti (eds), *Advances in Probability Distributions with Given Marginals*. Dordrecht: Kluwer.

Sklar, A. (1959) Fonctions de répartition à *n* dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, **8**, 229–231.

Stigler, S. (1986) *The History of Statistics*. Cambridge, MA: Harvard University Press.