IMPROVING EFFICIENT MARGINAL ESTIMATORS IN BIVARIATE MODELS WITH PARAMETRIC MARGINALS

HANXIANG PENG AND ANTON SCHICK

Abstract. Suppose we have data from a bivariate model with parametric marginals. Efficient estimators of the parameters in the marginal models are generally not efficient in the bivariate model. In this article, we propose a method of improving these marginal estimators and demonstrate that the magnitude of this improvement can be as large as 100 percent in some cases.

1. Introduction

Let (X, Y) be a bivariate random vector with distribution Q. Let F denote the distribution of X and G the distribution of Y. We assume that F and G belong to regular parametric models, but that Q is unknown otherwise. Then $F = F_\alpha$ and $G = G_\beta$. The parameter α can typically be estimated efficiently by the maximum likelihood estimator $\hat{\alpha}_n$ based on a random sample $X_1, \dots, X_n$ from F_α. Similarly, the parameter β can be estimated efficiently by the maximum likelihood estimator $\hat{\beta}_n$ based on a random sample $Y_1, \dots, Y_n$ from G_β. If a random sample $(X_1, Y_1), \dots, (X_n, Y_n)$ from Q is observed, then the marginal estimator $\hat{\alpha}_n$ may no longer be efficient as there may be information about α contained in the data $Y_1, \dots, Y_n$. A similar observation applies to $\hat{\beta}_n$.

In this paper we propose a simple method of using possible information about α contained in $Y_1, \dots, Y_n$. More precisely, we propose to subtract from $\hat{\alpha}_n$ a properly weighted stochastic term that is an asymptotically unbiased estimator of zero. The resulting estimator has an asymptotic dispersion matrix that is at most as large as the asymptotic dispersion matrix of $\hat{\alpha}_n$. A variant of our approach has already be used in improving nonparametric estimators in the presence of a constraint, see e.g. Schick and Wefelmeyer (2008) for a recent overview of such methods.

We give the details for our method in Section 2. In Section 3 we discuss the magnitude of improvement possible by this method. There it is demonstrated that improvements of up to 100 percent are possible. Section 4 discusses exponential marginals.

We conclude this introduction by specifying the notion of a regular parametric model as used throughout this paper. We use regularity in the sense of Bickel et al (1993, page 12-13) to mean continuous Hellinger-differentiability with positive definite information matrices. Here is the formal definition.

Definition 1. A model $\mathscr{H} = \{H_\theta : \theta \in \Theta\}$ parametrized by an open subset Θ of $\mathbb{R}^k$ is *regular* if it is dominated by a measure ν, if the map $\theta \mapsto \sqrt{h_\theta} =$

Hanxiang Peng is supported by NSF Grant DMS 0707074.
Anton Schick was supported by NSF Grant DMS 0405791.

$\sqrt{dH_\theta/d\nu}$ is continuously differentiable in $L_2(\nu)$ (whose derivative we may write as $\theta \mapsto (1/2)\dot\kappa(\cdot,\theta)\sqrt{h_\theta}$ for some $\dot\kappa(\cdot,\theta) \in L_2^k(H_\theta)$ with $\int \dot\kappa(x,\theta)\,dH_\theta(x) = 0$), and if $J(\theta) = \int \dot\kappa(x,\theta)\dot\kappa^\top(x,\theta)\,dH_\theta(x)$ is positive for each θ. Then $\dot\kappa(\cdot,\theta)$ is called the *score function at* θ and $J(\theta)$ the *information matrix at* θ.

2. The Method

Let (X, Y) be a bivariate random vector with distribution Q. Let F denote the distribution of X and G the distribution of Y. We assume that F and G belong to parametric models, but that Q is unknown otherwise. Let $\mathscr{F} = \{F_s : s \in \Theta_1\}$ denote the parametric model for F with Θ_1 an open subset of $\mathbb{R}^{k_1}$ and $\mathscr{G} = \{G_t : t \in \Theta_2\}$ the parametric model for G, with Θ_2 an open subset of $\mathbb{R}^{k_2}$. We assume that the models $\mathscr{F}$ and $\mathscr{G}$ are regular. Since we assume that F belongs to $\mathscr{F}$ and G belongs to $\mathscr{G}$, there are $\alpha \in \Theta_1$ and $\beta \in \Theta_2$ such that $F = F_\alpha$ and $G = G_\beta$.

We write $\dot\kappa_1(\cdot, s)$ for the score function of the model $\mathscr{F}$ at s, denote the corresponding information matrix by $J_1(s) = \int \dot\kappa_1(x,s)\dot\kappa_1^\top(x,s)\,dF_s(x)$ and set

$$\psi_1(x,s) = J_1^{-1}(s)\dot\kappa_1(x,s), \quad s \in \Theta_1.$$

We write $\dot\kappa_2(\cdot,t)$ for the score function of the model $\mathscr{G}$ at t, denote the corresponding information matrix by $J_2(t) = \int \dot\kappa_2(y,t)\dot\kappa_2^\top(y,t)\,dG_t(y)$ and set

$$\psi_2(y,t) = J_2^{-1}(t)\dot\kappa_2(y,t), \quad t \in \Theta_2.$$

Suppose we observe n independent copies $X_1, \dots, X_n$ of X only. Then an efficient estimator $\hat\alpha_n$ of α must satisfy

$$\hat\alpha_n = \alpha + \frac{1}{n}\sum_{j=1}^n \psi_1(X_j,\alpha) + o_p(n^{-1/2}). \tag{1}$$

Similarly, if only n independent copies $Y_1, \dots, Y_n$ of Y are observable, then an efficient estimator $\hat\beta_n$ of β must satisfy

$$\hat\beta_n = \beta + \frac{1}{n}\sum_{j=1}^n \psi_2(Y_j,\beta) + o_p(n^{-1/2}). \tag{2}$$

Typically maximum likelihood estimators possess these properties. We assume from now on that we have available estimators $\hat\alpha_n$ and $\hat\beta_n$ that satisfy (1) and (2).

Now suppose that we observe independent copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of the pair (X, Y). The above estimators use only the marginal information. Due to dependence, information about α may also be contained in the data $Y_1, \dots, Y_n$, and information about β may be contained in the data $X_1, \dots, X_n$. Thus the estimators $\hat\alpha_n$ and $\hat\beta_n$ may no longer be efficient and better estimators may be available. In this paper we demonstrate a simple approach of using the data $Y_1, \dots, Y_n$ to improve the estimator $\hat\alpha_n$ of α. By symmetry this also provides a method of improving the estimator $\hat\beta_n$ of β.

By our assumption on the models, $\mathscr{G}$ is dominated by a σ-finite measure μ. Write g_t for the density of G_t. Now let us look at a function W from $\mathbb{R} \times \Theta_2$ to $\mathbb{R}^m$ such that, for each t in Θ_2,

$$\int W(y,t)\,dG_t(y) = 0, \tag{3}$$

$$\int W(y,t)\dot{\kappa}_2^\top(y,t)\,dG_t(y) = 0, \tag{4}$$

$$\int W(y,t)W^\top(y,t)\,dG_t(y) = I_m, \tag{5}$$

with I_m the $m \times m$ identity matrix. Suppose also that

$$\int \|W(y,\beta_n)g_{\beta_n}^{1/2}(y) - W(y,\beta)g_\beta^{1/2}(y)\|^2\,d\mu(y) \to 0 \tag{6}$$

as $\beta_n \to \beta$. It then follows from Schick (2001) that

$$n^{-1/2}\sum_{j=1}^n [W(Y_j,\beta_n) - W(Y_j,\beta) + E[W(Y,\beta)\dot{\kappa}_2^\top(Y,\beta)](\beta_n - \beta)] = o_p(1)$$

for every sequence β_n in Θ_2 such that $n^{1/2}(\beta_n - \beta)$ is bounded. This even holds if we replace β_n by a discretized version of $\hat{\beta}_n$. To simplify notation we require that this holds without discretization. In view of (4) we then have

$$n^{-1/2}\sum_{j=1}^n W(Y_j,\hat{\beta}_n) = n^{-1/2}\sum_{j=1}^n W(Y_j,\beta) + o_p(1). \tag{7}$$

For a $k_1 \times m$ matrix D, consider the estimator $\hat{\alpha}_n(D)$ of α defined by

$$\hat{\alpha}_n(D) = \hat{\alpha}_n - \frac{1}{n}\sum_{j=1}^n DW(Y_j,\hat{\beta}_n).$$

It has expansion

$$\hat{\alpha}_n(D) = \alpha + \frac{1}{n}\sum_{j=1}^n [\psi_1(X_j,\alpha) - DW(Y_j,\beta)] + o_p(n^{-1/2}).$$

Thus $n^{1/2}(\hat{\alpha}_n(D) - \alpha)$ converges in distribution to a normal random vector with mean zero and the same dispersion matrix $\Psi(D)$ as $\psi_1(X,\alpha) - DW(Y,\beta)$. It is straightforward to check that this dispersion matrix is minimized by

$$D = D_* = E[\psi_1(X,\alpha)W^\top(Y,\beta)],$$

resulting in the minimal dispersion matrix

$$\Psi(D_*) = J_1^{-1}(\alpha) - D_*D_*^\top.$$

Since $D_*D_*^\top$ is non-negative definite, the estimator $\hat{\alpha}_n(D_*)$ is asymptotically no more dispersed than the estimator $\hat{\alpha}_n$. In the case that X and Y are independent, there is no information on α contained in the data $Y_1, ..., Y_n$ and no improvement over $\hat{\alpha}_n$ is possible. In this scenario, we have $D_* = 0$ and $\Psi(D_*) = J_1^{-1}(\alpha)$.

The matrix D_* is unknown and must be estimated. This can be done by

$$\hat{D}_* = \frac{1}{n}\sum_{j=1}^n \psi_1(X_j,\hat{\alpha}_n)W^\top(Y_j,\hat{\beta}_n).$$

This estimator is consistent if

$$\frac{1}{n}\sum_{j=1}^n \|\psi_1(X_j,\hat{\alpha}_n) - \psi_1(X_j,\alpha)\|^2 = o_p(1), \tag{8}$$

and

$$\frac{1}{n}\sum_{j=1}^{n}\|W(Y_j,\hat\beta_n)-W(Y_j,\beta)\|^2=o_p(1). \tag{9}$$

Thus, under assumptions (7)-(9) we find that the estimator

$$\hat\alpha_n^*=\hat\alpha_n-\hat D_*\frac{1}{n}\sum_{j=1}^{n}W(Y_j,\hat\beta_n)$$

satisfies

$$\hat\alpha_n^*=\alpha+\frac{1}{n}\sum_{j=1}^{n}[\psi_1(X_j,\alpha)-D_*W(Y_j,\beta)]+o_p(n^{-1/2}). \tag{10}$$

Hence $\sqrt{n}(\hat\alpha_n^*-\alpha)$ is asymptotically normal with zero mean vector and dispersion matrix $\Psi(D_*)$. Thus $\hat\alpha_n^*$ improves upon $\hat\alpha_n$ unless $D_*=0$.

We have already seen that $D_*=0$ if X and Y are independent. But $D_*=0$ can also happen with dependent data. For example, if X and Y have normal marginals and are jointly normal, then the maximum likelihood estimators of the means and variances based on the marginal observations coincide with the likelihood estimators based on the joint distribution and the marginal estimators are also efficient for the bivariate data and thus cannot be improved.

In the next section we shall discuss the magnitude of the improvements in some special cases. There we shall see that improvements of up to 100 percent are possible.

Remark 1. Our method can also be used if the marginal distribution G is known. In this case we replace $W(Y_j,\hat\beta_n)$ by $W(Y_j)$ where W needs to only satisfy

$$\int W\,dG=0 \quad\text{and}\quad \int WW^\top\,dG=I_m.$$

3. On the Magnitude of the Improvement

To see how large a gain can be achieved with our improvement procedure we consider now a bivariate location model. Let f be a *symmetric* density that has finite Fisher information $I(f)$ for location. This means that f is absolutely continuous and

$$I(f)=\int \ell_f^2(x)f(x)\,dx$$

is finite, where $\ell_f=-f'/f$ is the score function for location. We take $\mathscr{F}$ to be the location model $\mathscr{F}=\{F_s:s\in\mathbb{R}\}$ generated by f in which $dF_s(x)=f(x-s)\,dx$. It is well known that this model is regular with

$$\dot\kappa_1(x,s)=\ell_f(x-s),\quad J_1(s)=I(f),\quad \psi_1(x,s)=\ell_f(x-s)/I(f).$$

For $\mathscr{G}$ we take the location model generated by a *symmetric* density g with finite Fisher information $I(g)$ for location. Then $\mathscr{G}$ is regular with

$$\dot\kappa_2(y,t)=\ell_g(y-t),\quad J_2(t)=I(g),\quad \psi_2(y,t)=\ell_g(y-t)/I(g).$$

As the underlying joint distribution we take the bivariate probability measure Q which has a density q of the form

$$q(x,y) = (1+\rho u(x-\alpha)v(y-\beta))f(x-\alpha)g(y-\beta), \quad x,y\in\mathbb{R}, \tag{11}$$

for some (unknown) reals α and β and some ρ in $[-1,1]$. Here u and v are measurable functions with values in $[-1,1]$ such that $\int u(x)f(x)\,dx = 0$ and $\int v(y)g(y)\,dy = 0$. Some of the copula models are of this form. Indeed, the Farlie-Gumbel-Morgenstern copula model is of this type. See Hutchinson and Lai (1990) and Nelsen (1999) for properties and applications of this copula model as well as for additional references.

We take $W(y,t)$ to be of the form $w(y-t)$ for some function w. The above distribution Q allows now for a simple calculation of D_*. Indeed, under Q, we calculate

$$\begin{aligned} I(f)D_* &= \rho E[u(X-\alpha)\ell_f(X-\alpha)]E[v(Y-\beta)w(Y-\beta)] \\ &= \rho\int u(x)\ell_f(x)f(x)\,dx\int v(y)w(y)g(y)\,dy. \end{aligned}$$

The required conditions (3)-(6) are implied if we take w so that

$$\int w(y)g(y)\,dy = 0, \qquad \int w^2(y)g(y)\,dy = 1, \qquad \int w(y)g'(y)\,dy = 0. \tag{12}$$

A possible choice for w is

$$w_*(x) = 2\mathbf{1}_{[-c,c]}(x) - 1 = \begin{cases} 1, & |x|\le c, \\ -1, & |x|>c, \end{cases}$$

with c the third quartile of g. Indeed, then the first part of (12) follows from the choice of c, the second part follows from the fact that $|w_*| = 1$, and the third part follows as w_* is even and g' is odd. The asymptotic relative efficiency (ARE) of the proposed estimator $\hat{\alpha}_n^*$ to the marginal estimator $\hat{\alpha}_n$ is the ratio of the asymptotic variance $1/I(f) - D_*^2$ of $\hat{\alpha}_n^*$ and the asymptotic variance $1/I(f)$ of $\hat{\alpha}_n$:

$$\text{ARE} = 1 - I(f)D_*^2 = 1 - \frac{\rho^2}{I(f)}\Big(\int u(x)\ell_f(x)f(x)\,dx\int v(y)w(y)g(y)\,dy\Big)^2.$$

As $|v|$ is bounded by 1, we see that

$$\text{ARE} \ge 1 - \frac{\rho^2}{I(f)}\Big(\int u(x)\ell_f(x)f(x)\,dx\Big)^2$$

with equality if $v = w = w_*$. Note that $u = \operatorname{sign}(\ell_f)$ is an odd function in view of the symmetry of f so that $\int \operatorname{sign}(\ell_f(x))f(x)\,dx = 0$ and that $|u| = 1$. Note also that this u maximizes $|\int u(x)\ell_f(x)f(x)\,dx|$ subject to $|u|\le 1$ and $\int u(x)f(x)\,dx = 0$. Thus we obtain

$$\text{ARE} \ge 1 - \frac{\rho^2}{I(f)}\Big(\int |f'(x)|\,dx\Big)^2 \tag{13}$$

with equality if $u = \operatorname{sign}(\ell_f)$ and $v = w = w_*$.

Example 1. Let us choose f and g to be the standard normal density p given by $p(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$. This density has finite Fisher information $I(p) = 1$ and score function $\ell_p(x) = x$. Based on the marginal observations $X_1,\dots,X_n$, the maximum likelihood estimator $\hat{\alpha}_n$ is the sample mean. The sample mean is efficient in the (standard) normal location model $\mathscr{F}$ with asymptotic variance 1. For $v = w = w_*$ and $u(x) = \operatorname{sign}(x)$, we obtain

$$\mathrm{ARE} = 1 - \rho^2 \frac{2}{\pi}\Big(\int_0^\infty x \exp(-x^2/2)\,dx\Big)^2 = 1 - \frac{2\rho^2}{\pi}.$$

Thus the efficiency gain can be as large as is $2/\pi$ or about 63.7 percent.

Example 2. Let us choose f and g to be the Cauchy density given by $p(x) = 1/(\pi(1+x^2))$. This density has finite Fisher information for location $I(p) = 1/2$ and score function $\ell_p(x) = 2x/(1+x^2)$. Based on the marginal observations $X_1, ..., X_n$, the maximum likelihood estimator $\hat\alpha_n$ is a solution to the score equation

$$\sum_{i=1}^n \frac{X_i - \hat\alpha_n}{1 + (X_i - \hat\alpha_n)^2} = 0.$$

This estimator $\hat\alpha_n$ is efficient in the Cauchy-location model $\mathscr{F}$ with asymptotic variance 2. The third quartile of the Cauchy density p is $c = 1$. For $v = w = w_* = 2\mathbf{1}_{[-1,1]} - 1$ and $u(x) = \operatorname{sign}(x)$, we have

$$\mathrm{ARE} = 1 - 2\rho^2\Big(2\int_0^\infty \frac{2x}{\pi(1+x^2)^2}\,dx\Big)^2 = 1 - \frac{8\rho^2}{\pi^2}.$$

Thus the efficiency gain can be as large as is $8/\pi^2$ or about 81 percent.

Example 3. Let us choose f and g to be the Laplace (or double exponential) density p given by $p(x) = (1/2)\exp(-|x|)$. This density has finite Fisher information $I(p) = 1$ and score function $\ell_p(x) = \operatorname{sign}(x)$. The maximum likelihood estimator $\hat\alpha_n$ is the sample median of $X_1, \dots, X_n$. It is efficient in the Laplace location model $\mathscr{F}$ and has asymptotic variance 1. The third quartile of the Laplace density p is $c = \log 2$. For $v = w = w_* = 2\mathbf{1}_{[-\log 2, \log 2]} - 1$ and $u(x) = \operatorname{sign}(x)$, we obtain

$$ARE = 1 - \rho^2\Big(\int_0^\infty e^{-x}\,dx\Big)^2 = 1 - \rho^2.$$

Thus the efficiency gain can be as large as 1 or 100 percent.

4. Exponential marginals

Let us now assume that $\mathscr{F} = \mathscr{G} = \{F_s : s > 0\}$, where F_s is the exponential distribution with mean s, i.e., F_s has density $f_s(x) = f(x/s)/s$, where

$$f(x) = \exp(-x)\mathbf{1}_{(0,\infty)}(x), \quad x \in \mathbb{R}.$$

This model is regular with score function $\dot\kappa(x, s) = (x - s)/s^2$ and information $J(s) = 1/s^2$. Thus $\psi_1(x, \alpha) = x - \alpha$ and $\psi_2(y, \beta) = y - \beta$. The maximum likelihood estimator based on a random sample from this model is the sample mean and is efficient. Thus we take $\hat\alpha_n = \bar X = (X_1 + \cdots + X_n)/n$ and $\hat\beta_n = \bar Y_n = (Y_1 + \cdots + Y_n)/n$.

Here we take W to be of the form $W(y, t) = w(y/t)$ for some function w. The required conditions (3)-(6) are implied if w satisfies

$$\int w(y)f(y)\,dy = 0, \quad \int w^2(y)f(y)\,dy = 1, \quad \int w(y)(y-1)f(y)\,dy = 0. \tag{14}$$

Possible choices for w are given by $w = w_{(c,d)} = (\mathbf{1}_{(c,d)} - p)/\sqrt{p(1-p)}$, where $p = \exp(-c) - \exp(-d)$ and $0 < c < 1 < d < \infty$ are chosen to satisfy $c\exp(-c) = d\exp(-d)$.

We shall look at two different joint distributions Q. The first joint distribution has distribution function H given by

$$H(x,y) = \gamma F(x/\alpha)F(y/\beta) + (1-\gamma)\min(F(x/\alpha),\, F(y/\beta)), \quad x,y \in \mathbb{R}, \tag{15}$$

for some $0 < \gamma < 1$, where F is the distribution function of f. The second joint distribution has a density q of the form

$$q(x,y) = (1 + \rho u(x/\alpha)v(y/\beta))f(x/\alpha)f(y/\beta)/(\alpha\beta), \quad x,y \in \mathbb{R}, \tag{16}$$

for some ρ in $[-1,1]$. Here u and v are measurable functions with values in $[-1,1]$ such that $\int u(x)f(x)\,dx = 0$ and $\int v(y)f(y)\,dy = 0$.

We shall see that for the first joint distribution our method does not provide any improvement. Indeed, H is the distribution function of $(\alpha\tilde{X},\, \beta(\mathbf{1}[U \le \gamma]\tilde{Y} + \mathbf{1}[U > \gamma]\tilde{X})$ where $\tilde{X}$, $\tilde{Y}$ and U are independent, $\tilde{X}$ and $\tilde{Y}$ are exponentially distributed with mean 1, and U is uniformly distributed on $(0,1)$. Thus we can express

$$\begin{aligned} D_* &= \alpha E[(\tilde{X}-1)w(\mathbf{1}[U \le \gamma]\tilde{Y} + \mathbf{1}[U > \gamma]\tilde{X})] \\ &= \alpha(E[\mathbf{1}[U \le \gamma](\tilde{X}-1)w(\tilde{Y})] + E[\mathbf{1}[U > \gamma](\tilde{X}-1)w(\tilde{X})]. \end{aligned}$$

In view of the independence of $U, \tilde{X}, \tilde{Y}$ and the last condition in (14), we find $D_* = 0$. Thus no improvement is possible.

For the second distribution we calculate

$$\begin{aligned} D_* &= \rho E[u(X/\alpha)(X-\alpha)]E[v(Y/\beta)w(Y/\beta)] \\ &= \alpha\rho \int_0^\infty u(x)(x-1)\exp(-x)\,dx \int_0^\infty v(y)w(y)\exp(-y)\,dy. \end{aligned}$$

The asymptotic relative efficiency of the proposed estimator $\hat{\alpha}_n^*$ to the marginal estimator $\hat{\alpha}_n$ is the ratio of the asymptotic variance $\alpha^2 - D_*^2$ of $\hat{\alpha}_n^*$ and the asymptotic variance α^2 of $\hat{\alpha}_n$, namely,

$$\mathrm{ARE} = 1 - \frac{D_*^2}{\alpha^2} = 1 - \rho^2\Big(\int_0^\infty u(x)(x-1)\exp(-x)\,dx \int_0^\infty v(y)w(y)\exp(-y)\,dy\Big)^2.$$

Let us now take $u = u_* = \mathbf{1}_{(1,\infty)} - \mathbf{1}_{(0,1-\delta)}$ with $\delta = \log(e-1)$. Then the integral involving u becomes $\delta_* = \exp(-1) + (1-\delta)\exp(\delta - 1) = 1 - (1-e^{-1})\log(e-1)$, and accordingly,

$$\mathrm{ARE} = 1 - \rho^2\delta_*^2\Big(\int_0^\infty v(y)w(y)\exp(-y)\,dy\Big)^2.$$

For the choice $w = w_{(c,d)}$ we obtain

$$\mathrm{ARE} \ge 1 - \rho^2\delta_*^2\frac{\sqrt{p(1-p)}}{\max(p,1-p)}$$

with equality for $v = (\mathbf{1}_{(c,d)} - p)/\max(p,1-p)$. If c and d are such that $p = 1/2$, then $\mathrm{ARE} = 1 - \rho^2\delta_*^2$. Since $\delta_*^2 \simeq 0.4327$, we see that improvements of up to 56.7 percent are possible.

References

[1] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and adaptive estimation in semiparametric models.* Johns Hopkins Univ. Press, Baltimore.

[2] Hutchinson, T. P., and Lai, C. D. (1990) *Continuous bivariate distributions, emphasising applications.* Rumsby Scientific Publishing, Adelaide.

[3] Nelsen, R. (1999). An introduction to copulas. *Lecture Notes in Statistics*, 139. Springer-Verlag, New York.

[4] Hanxiang Peng and Anton Schick (2004). Estimation of Linear Functionals of a Bivariate Probability with Unknown Parametric Marginals. *Statistics & Decisions.* **22** 61-78

[5] Schick, A. (2001). On asymptotic differentiability of averages. *Statist. Probab. Lett.*, **51**, 15-23.

[6] Schick, A. and Wefelmeyer, W. (2008). Some developments in semiparametric models. To appear in: *J. Stat. Theory Pract.*

Hanxiang Peng, Department of Mathematical Sciences, Purdue School of Sciences, Indiana University-Purdue University at Indianapolis, Indianapolis, IN 46202-3216, USA

Anton Schick, Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902-6000, USA