Semiparametric Gaussian Copula Models:

Progress and Problems

Jon A. Wellner

University of Washington, Seattle

2015 IMS China,
Kunming

July 1-4, 2015

2015 IMS China Meeting, Kunming

Based on joint work with:

- Peter Hoff
- Xiaoyue (Maggie) Niu
- Chris Klaassen

Outline

- 0. Basics: notation and facts
- 1: Bivariate Gaussian copula models
- 2: $d-$variate Gaussian Copula models
- 3: Recent progress and results
- 4: Questions and open problems

0. Basics: notation and facts

Notation:

- $\Theta \subset \mathbb{R}^q$, $q \geq 1$; $\mathcal{F} = \{$all distribution functions on $\mathbb{R}\}$.
- Copulas: $\{C_\theta : \theta \in \Theta\} =$ a parametric family of distribution functions on $[0,1]^d$ with uniform marginal distributions $C_\theta(1,\ldots,1,u_j,1,\ldots,1) = u_j$ for $u_j \in (0,1)$ and $j = 1,\ldots,d$.
- Semiparametric copula distribution functions and measures: $F_{\theta,F_1,\ldots,F_d}(x_1,\ldots,x_d) = C_\theta(F_1(x_1),\ldots,F_d(x_d))$ for distribution functions F_j on $\mathbb{R}$,
  $P_{\theta,F_1,\ldots,F_d}(A) = \int_A dF_{\theta,F_1,\ldots,F_d}(\underline{x}),\ A \in \mathcal{B}^d$.
- Semiparametric copula model:
  $\mathcal{P} = \{P_{\theta,F_1,\ldots,F_d} :\ \theta \in \Theta,\ F_j \in \mathcal{F},\ j = 1,\ldots,d\}$.

Main focus here: multivariate Gaussian copulas

$$\Phi_\theta(\underline{x}) = P_\theta(\underline{X} \le \underline{x}) = \text{d.f. of } \; N_d(\underline{0}, \Sigma(\theta)),$$

where

$$\Sigma(\theta) = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1,d} \\ \rho_{12} & 1 & \rho_{23} & \cdots & \rho_{2,d} \\ \vdots & & & \vdots & \vdots \\ \rho_{1,d} & & & \rho_{d-1,d} & 1 \end{pmatrix}$$

and $\rho_{i,j} \equiv \rho_{i,j}(\theta)$. Then

$$\begin{aligned} C_\theta(\underline{u}) &= \Phi_\theta(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)), \\ c_\theta(\underline{u}) &= \frac{\phi_\theta(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))}{\prod_{j=1}^d \phi(\Phi^{-1}(u_j))}, \end{aligned}$$

for $\underline{u} = (u_1, \ldots, u_d) \in (0,1)^d$, and ...

$$F_{\theta,F_1,\ldots,F_d}(x_1,\ldots,x_d) = C_\theta(F_1(x_1),\ldots,F_d(x_d)), \quad \theta \in \Theta, \quad F_j \in \mathcal{F},$$

and $\mathcal{P}_d$ is a semiparametric Gaussian copula model based on c_θ.

Now suppose that we observe $\underline{X}_1,\ldots,\underline{X}_n$ i.i.d. with probability distribution $P_{\theta_0,F_{0,1},\ldots,F_{0,d}} \in \mathcal{P}_d$.

Questions:

- How well can we estimate $\theta \in \Theta$? (Lower bounds)
- Can we construct (rank-based) estimators achieving the lower bounds?

Since the model is invariant under monotone transformations on each axis, it is clear that the (multivariate) ranks are a maximal invariant.

More notation: let $\mathbf{X}$ denote the $n \times d$ matrix with rows $\underline{X}_1, \ldots, \underline{X}_n$. Let $\mathbf{R}(\mathbf{X}) : \mathbb{R}^{n \times d} \to \mathbb{R}^{n \times d}$ be the corresponding $n \times d$ matrix of ranks where $\mathbf{R} = (R_{i,j})$ and

$$R_{i,j} = \text{the rank of } X_{i,j} \text{ among } \{X_{1,j}, \ldots, X_{n,j}\}, \quad j = 1, \ldots, d.$$

Hoff (2007) has shown that the ranks $\mathbf{R}$ are partially sufficient in several senses, and it seems natural to try base inference procedures on them if possible.

1. Bivariate Gaussian copulas

Here $d = 2$ and $\theta \in \Theta = (-1, 1)$. **Klaassen and W (1997)** showed:

- $I_\theta(\mathcal{P}_2) = (1 - \theta^2)^{-2}$.
- Normal margins are least favorable.
- $\widehat{\theta}_n =$ normal scores rank correlation coefficient is asymptotically efficient:

$$\sqrt{n}(\widehat{\theta}_n - \theta) \rightarrow_d N(0, (1 - \theta^2)^2).$$

- $\widehat{\theta}_n$ is asymptotically equivalent to the maximum pseudo likelihood estimator $\widehat{\theta}^{ple}$: $\sqrt{n}(\widehat{\theta}_n - \widehat{\theta}_n^{ple}) = o_p(1)$ where

$$\widehat{\theta}_n^{ple} = \text{argmax}_{\theta \in \Theta} \ell_n(\theta, \mathbb{G}_n, \mathbb{H}_n)$$

  where $\mathbb{G}_n$, $\mathbb{H}_n$, are the marginal empirical distribution functions of the data. (Note that $\widehat{\theta}_n^{ple}$ is also a function of the ranks.)

1. Bivariate Gaussian copulas

Here with $\underline{X}_i = (Y_i, Z_i)$, $i = 1, \ldots, n$,

$$\begin{aligned}
\widehat{\theta}_n &= \frac{n^{-1}\sum_{i=1}^{n}\Phi^{-1}(\mathbb{G}_n^*(Y_i))\Phi^{-1}(\mathbb{H}_n^*(Z_i))}{n^{-1}\sum_{i=1}^{n}\Phi^{-1}\left(\frac{i}{n+1}\right)^2} \\
&= \frac{n^{-1}\sum_{i=1}^{n}\Phi^{-1}\left(\frac{R_{i,1}}{n+1}\right)\Phi^{-1}\left(\frac{R_{i,2}}{n+1}\right)}{n^{-1}\sum_{i=1}^{n}\Phi^{-1}\left(\frac{i}{n+1}\right)^2}
\end{aligned}$$

Asymptotic linearity:

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{\ell}_\theta(\underline{X}_i) + o_p(1)$$

where

$$\begin{aligned}
\widetilde{\ell}_\theta(y,z) &= I_\theta^{-1}\dot{\ell}_\theta^*(y,z) \\
&= \Phi^{-1}(G(y))\Phi^{-1}(H(z)) - \frac{\theta}{2}\left(\Phi^{-1}(G(y))^2 + \Phi^{-1}(H(z))^2\right).
\end{aligned}$$

2. Multivariate Gaussian copulas, $d > 2$

- When $\Sigma(\theta)$ is unstructured (i.e. $\theta = (\rho_{1,2}, \rho_{1,3}, \ldots, \rho_{1,d}, \ldots, \rho_{d-1,d}) \in [-1,1]^{d(d-1)/2}$), then the pseudo-likelihood estimator continues to be semiparametric efficient, as noted by Klaassen & W (1997), and Segers, von den Akker, Werker (2014).

- What if $d > 2$ and $\Sigma(\theta)$ is structured?

Examples:

- Example 1. (Exchangeable) $\Sigma(\theta) = (1-\theta)I_d + \theta \underline{1}\underline{1}^T$ with $\theta \in [-1/(d+1), 1)$. For example for $d = 4$

$$\Sigma(\theta) = \begin{pmatrix} 1 & \theta & \theta & \theta \\ \theta & 1 & \theta & \theta \\ \theta & \theta & 1 & \theta \\ \theta & \theta & \theta & 1 \end{pmatrix}.$$

- Example 2. (Circular) For $d = 4$,

$$\Sigma(\theta) = \begin{pmatrix} 1 & \theta & \theta^2 & \theta \\ \theta & 1 & \theta & \theta^2 \\ \theta^2 & \theta & 1 & \theta \\ \theta & \theta^2 & \theta & 1 \end{pmatrix}.$$

- Example 3. (Toeplitz). Here $\Sigma = (\sigma_{i,j})$ with $\sigma_{i,i} = 1$ for all i, $\sigma_{i,j} = \theta_{|i-j|}$ for $\theta = (\theta_1, \theta_2, \ldots, \theta_{d-1}) \in (-1, 1)^{d-1}$. For example, with $d = 4$,

$$\Sigma(\theta) = \begin{pmatrix} 1 & \theta_1 & \theta_2 & \theta_3 \\ \theta_1 & 1 & \theta_1 & \theta_2 \\ \theta_2 & \theta_1 & 1 & \theta_1 \\ \theta_3 & \theta_2 & \theta_1 & 1 \end{pmatrix}.$$

More background:

- Genest and Werker (2000): studied efficiency properties of pseudo-likelihood estimators for general semiparametric copula models:
  Conclusion: $\widehat{\theta}_n^{ple}$ is not efficient in general for (non-Gaussian) copulas.

- Chen, Fan, and Tsyrennikov (2006) constructed semiparametric efficient estimators for general multivariate copula models using parametric sieve methods. Their estimators of θ are not based solely on the multivariate ranks

Questions:

- Do Maximum Likelihood Estimators based on rank likelihoods achieve semiparametric efficiency for general multivariate copula models?
- Do alternative estimators based on ranks achieve semiparametric efficiency?
- Are the pseudo maximum likelihood estimators semiparametric efficient for structured Gaussian copula models?

For $\theta \in \Theta \subset \mathbb{R}^q$ with $q < d(d-1)/2$, let

$$L(\theta; \mathbf{R}) \text{ denote the likelihood of the ranks } \mathbf{R},$$
$$L(\theta, \psi; \mathbf{X}) \text{ denote the likelihood of the data } \mathbf{X},$$

where $\psi \in \Psi$ denotes parameters for the marginal transformations. For fixed $\theta \in \Theta$, $\psi \in \Psi$ let

$$\lambda_{\mathbf{R}}(t) \equiv \log \frac{L(\theta + t/\sqrt{n}; \mathbf{R})}{L(\theta; \mathbf{R})},$$
$$\lambda_{\mathbf{X}}(t, s) \equiv \log \frac{L(\theta + t/\sqrt{n}, \psi + s/\sqrt{n}; \mathbf{X})}{L(\theta, \psi; \mathbf{X})}.$$

Theorem 1. (Hoff-Niu-W, 2014) Let $\{F_{\theta,\psi}(\underline{x}) : \ \theta \in \Theta, \ \psi \in \Psi\}$ be an absolutely continuous copula model where, for given θ and t there exist ψ and s such that under i.i.d. sampling from $F_{\theta,\psi}$. Suppose that:
(1) $\lambda_{\mathbf{X}}(t,s)$ satisfies Local Asymptotic Normality (LAN):

$$\lambda_{\mathbf{X}}(t,s) \rightarrow_d Z$$

(2) There exists an $\mathbf{R}$-measurable approximation $\lambda_{\hat{\mathbf{X}}}(t,s)$ such that $\lambda_{\hat{\mathbf{X}}}(t,s) - \lambda_{\mathbf{X}} \rightarrow_p 0$.
Then $\lambda_{\mathbf{R}}(t) \rightarrow_d Z$ under i.i.d. sampling from any population with copula $C_\theta(\cdot)$ equal to that of $F(\cdot;\theta,\psi)$ and arbitrary absolutely continuous marginal distributions.

Conclusion: To show that the local likelihood ratio of the ranks satisfies LAN (from which an information bound follows for procedures based on the ranks follows), we need to construct suitable rank-measurable approximations of the local likelihood ratios of the data for parametric submodels.

Let $\underline{X}_1, \ldots, \underline{X}_n$ be i.i.d. from a member $P_{\theta,\psi}$ of a collection of $N_d(0, \Sigma_{\theta,\psi})$ where θ parameterizes the correlations and ψ are the variance parameters. Then

$$\lambda_{\mathbf{X}}(t,s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \underline{X}_i^T A \underline{X}_i + c(\theta, \psi, t, s) + o_p(1)$$

where $A = A_{t,s,\theta,\psi}$. A natural rank-based approximation is

$$\lambda_{\widehat{\mathbf{X}}}(t,s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \underline{\widehat{X}}_i^T A \underline{\widehat{X}}_i + c(\theta, \psi, t, s)$$

where

$$\widehat{X}_{i,j} \equiv \sqrt{Var(X_{i,j})} \Phi^{-1} \left(\frac{R_{i,j}}{n+1} \right).$$

This leads to the following theorem:

Theorem 2. (Hoff, Niu, & W, 2014) Let $\underline{X}_1, \ldots, \underline{X}_n$ be i.i.d. $N_d(0, C)$ where C is a correlation matrix and let $\widehat{X}_{i,j} = \Phi^{-1}(R_{i,j}/(n+1))$. Let A be a matrix such that the diagonal entries of $AC + A^T C$ are zero. Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\underline{\widehat{X}}_i^T A \underline{\widehat{X}}_i - \underline{X}_i^T A \underline{X}_i\} = o_p(1).$$

- The proof of Theorem 2 is based on some classical results of de Wet and Venter (1972).
- It remains to apply the results of Theorems 1 and 2 to the setting of Gaussian copulas:

Theorem 3. (Hoff, Niu, & W, 2014). Suppose that $\{\Sigma(\theta) : \theta \in \Theta \subset \mathbb{R}^q\}$ is a collection of positive definite correlation matrices such that $\Sigma_{i,j}(\theta)$ is continuously differentiable with respect to each θ_k, $1 \le k \le q$. If $\underline{X}_1, \ldots, \underline{X}_n$ are i.i.d. $P_{\theta,\psi}$ with absolutely continuous marginals and Gaussian copula C_θ for some $\theta \in \Theta$, then the local likelihood ratio of the ranks $\lambda_{\mathbf{R}}(t)$ satisfies LAN:

$$\lambda_{\mathbf{R}}(t) \rightarrow_d N(-(1/2)t^T I_{\theta\theta\cdot\psi} t, t^T I_{\theta\theta\cdot\psi} t)$$

where $I_{\theta\theta\cdot\psi}$ is the information for θ in the Gaussian model with correlation matrix $\Sigma(\theta)$ and precisions ψ.

Summary: Let $B(\theta) \equiv \Sigma^{-1}(\theta)$. Then, for $q = 1$,

- The efficient score function ℓ_θ^* is, with $\underline{y} = (\Phi^{-1}(F_1(x_1), \ldots, \Phi^{-1}(F_d(x_d))$:

$$\ell_\theta^*(y) \;=\; \dot{\ell}_\theta - I_{\theta\psi} I_{\psi\psi}^{-1} \dot{\ell}_\psi = \frac{1}{2} y^T \left\{ \frac{\psi}{d} \text{tr}(B_\theta C) B - \psi B_\theta \right\} y.$$

- The efficient influence function $\tilde{\ell}_\theta$ for θ is, with $\underline{y} = (\Phi^{-1}(F_1(x_1), \ldots, \Phi^{-1}(F_d(x_d))$:

$$\begin{aligned} \tilde{\ell}_\theta(y) &= I^{-1}_{\theta\theta\cdot\psi} \ell^*_\theta(y), \qquad \text{where} \\ I_{\theta\theta\cdot\psi} &= (1/2)\{\mathrm{tr}(B_\theta C B_\theta C) - \mathrm{tr}(B_\theta C)^2/d\}. \end{aligned}$$

Consequences:

- No information concerning θ is lost (asymptotically) by reducing to the ranks $\mathbf{R}$.
- Gaussian marginals are least favorable.
- The information bounds for estimation of θ in such a Gaussian copula model are given in terms of $I^{-1}_{\theta\theta\cdot\psi}$.

The efficient influence function $\widetilde{\ell}_\theta(\underline{x})$ can be shown to be

$$\widetilde{\ell}_\theta(\underline{x}) = I_{\theta\theta}^{-1}\left\{\dot{\ell}_\theta(\underline{x}) - I_{\theta\psi}\widetilde{\ell}_\psi(\underline{x})\right\}$$

The influence function of the pseudo likelihood estimator is given by

$$\psi_\theta(\underline{x}) = I_{\theta\theta}^{-1}\left(\dot{\ell}_\theta(\underline{x}) - \sum_{j=1}^{d} W_j(x_j)\right)$$

where

$$W_j(x_j) = \int_{(0,1)^d}\left(\frac{\partial^2}{\partial\theta\partial u_j}\log c_\theta(\underline{u})\right)\left(1\{\Phi(x_j) \le u_j\} - u_j\right)c_\theta(\underline{u})d\underline{u}.$$

Corollary: The maximum pseudo likelihood estimator is semi-parametric efficient if

$$\sum_{j=1}^{d} W_j(x_j) = \frac{1}{2}\mathrm{tr}\left(\mathbf{B}\boldsymbol{\Sigma}_\theta\{I - \mathrm{diag}(\underline{x}\circ\underline{x})\}\right).$$

When $q = 1$ (and then $\psi \in \mathbb{R}$), this simplifies to

$$\widetilde{\ell}_\psi(\underline{x}) = \frac{1}{d} \sum_{j=1}^{d} (1 - x_j^2).$$

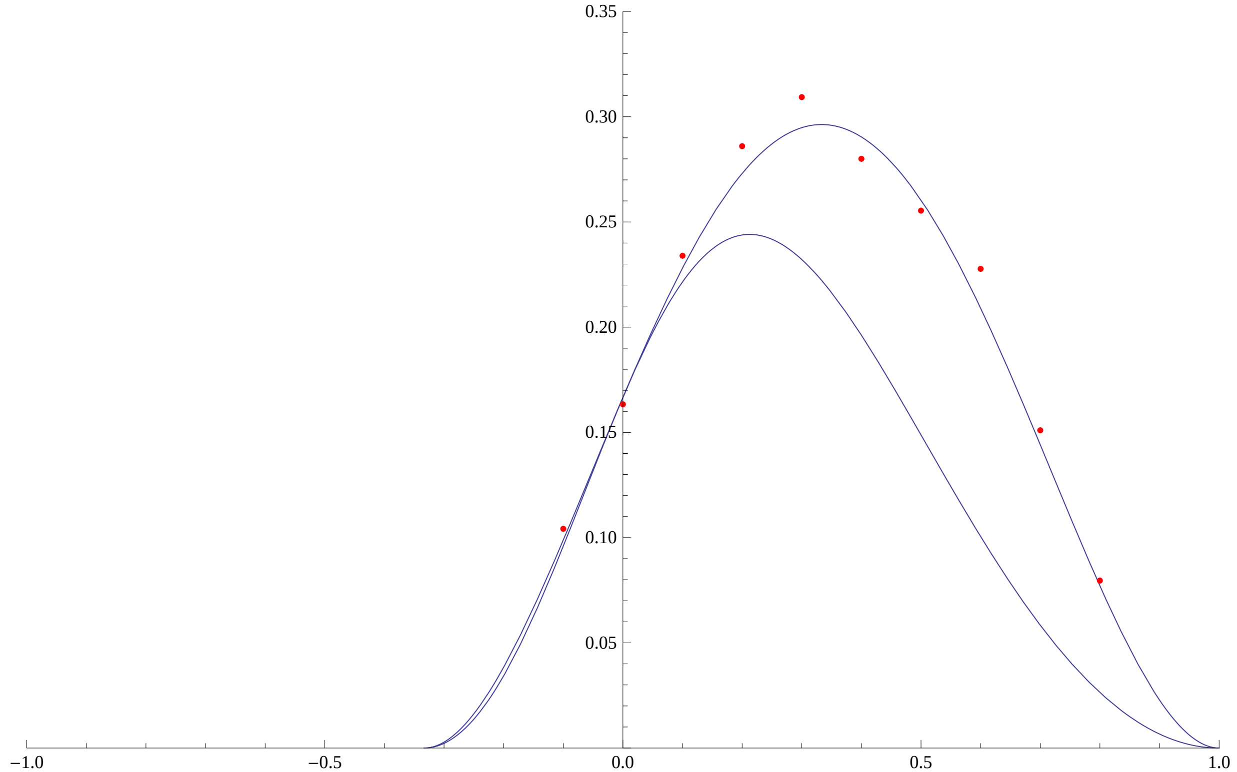
Examples, continued:

- Example 1. (Exchangeable) $\Sigma(\theta) = (1-\theta)I_d + \theta \underline{1}\underline{1}^T$. For $d = 4$, calculation yields

$$
\begin{aligned}
I^{-1}_{\theta\theta\cdot\psi} &= \frac{1}{6}(1 + 2\theta - 3\theta^2), \\
\widetilde{\ell}_\theta(\underline{x}) &= \frac{1}{12}\left\{2 \sum_{1\le i<j\le 4} x_i x_j - 3\theta \sum_{j=1}^{4} x_j^2\right\}, \quad \text{and} \\
-I_{\theta\psi}\widetilde{\ell}_\psi(\underline{x}) &= \frac{6\theta}{1+2\theta-3\theta^2}\frac{1}{4}\sum_{j=1}^{4}(x_j^2 - 1) \\
&= \frac{3\theta/2}{1+2\theta-3\theta^2}\sum_{j=1}^{4}(x_j^2 - 1) = \sum_{j=1}^{4} W_j(x_j),
\end{aligned}
$$

so the pseudo-likelihood estimator is semiparametric efficient.

Figure 1, Example 1: Information bounds and Monte-carlo variance of p-mle: red, $n = 800$.



- Example 2. (Circular) For $d = 4$, calculation yields

$$
\begin{aligned}
I_{\theta\theta\cdot\psi} &= \frac{4}{(1-\theta^2)^2}, \\
\widetilde{\ell}_\theta(\underline{x}) &= \frac{1}{8(1-\theta^2)}\Bigg\{(1+\theta^2)\sum_{j=i+1,i+3} x_i x_j \\
&\qquad - \; 2\theta \sum_{j=1}^{4} x_j^2 - 2\theta \sum_{j=i+2} x_i x_j\Bigg\}, \quad \text{and} \\
-I_{\theta\psi}\widetilde{\ell}_\psi(\underline{x}) &= \text{a complicated quadratic in } x_j\text{'s and cubic in } \theta \\
&\neq \sum_{j=1}^{4} W_j(x_j) = -\frac{\theta}{1-\theta^2}\sum_{j=1}^{4}(x_j^2-1).
\end{aligned}
$$

so the pseudo-likelihood estimator is not semiparametric efficient.

Figure 1, Example 2: Information bound and variance of p-mle

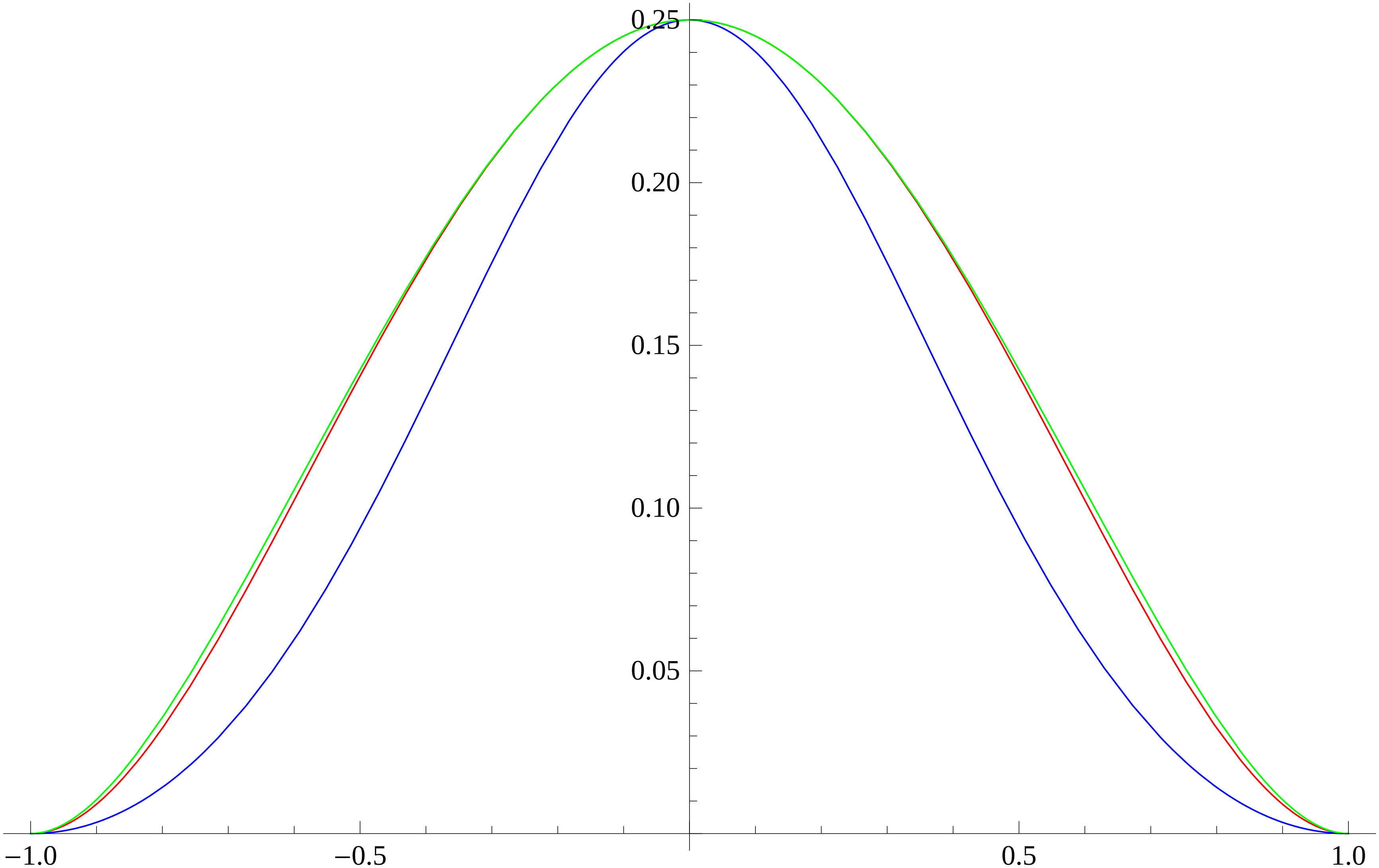
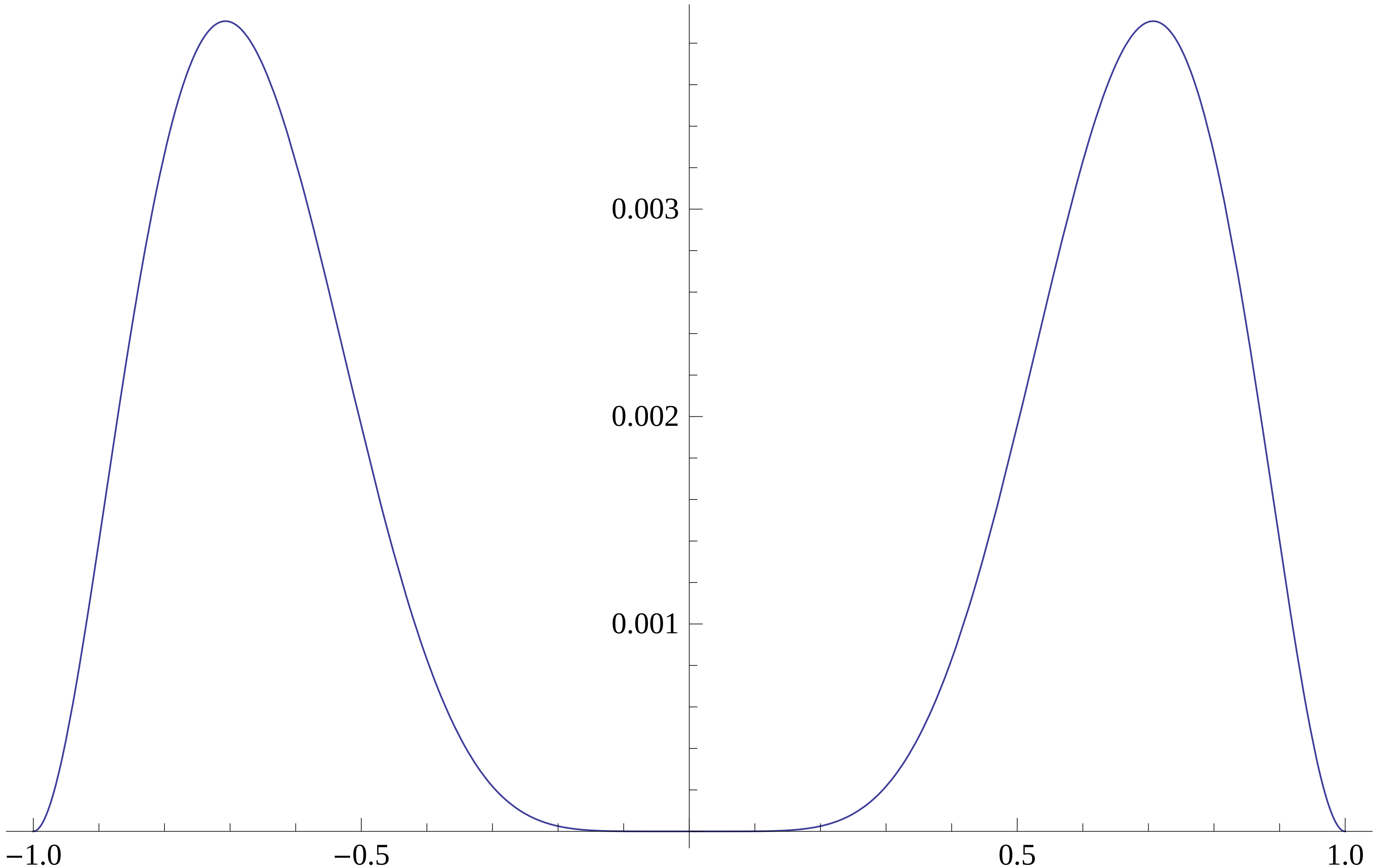


Figure 2, Example 2: Difference, variance of p-mle and Information bound



Summary:

- Information bounds for (structured) multivariate Gaussian models are available and computable.
- Gaussian marginal distributions are least favorable.
- The pseudo likelihood estimator is not always semiparametric efficient (but perhaps not missing efficiency by much).

Questions:

- Can we construct rank-based semiparametric efficient estimators?
- Are the pseudo likelihood estimators sometimes seriously inefficient?

Segers, van den Akker, and Werker (2014) give affirmative answers to both questions!

Recent progress and results

Segers, van den Akker, and Werker (2014) give affirmative answers to both questions!

Rank-based semiparametric efficient estimators:

via a "one-step" method:

- Start with a $\sqrt{n}-$consistent rank based estimator $\widehat{\theta}_n^0$; e.g the pseudo likelihood estimator $\widehat{\theta}_n^{ple}$.
- Construct the natural one-step estimator starting from $\widehat{\theta}_n^0$ and based on the efficient score function $\dot{\ell}_\theta^*$.

Recent progress and results

Inefficiency of pseudo likelihood estimator $\widehat{\theta}_n^{ple}$:

Example 3: (Toeplitz correlation model) Suppose that $\theta = (\theta_1, \ldots, \theta_{d-1}) \in (-1,1)^{d-1}$ and $\Sigma = (\sigma_{i,j})_{i,j=1}^d = (\sigma_{i,j}(\theta)$ where $\sigma_{i,i} = 1$ and $\sigma_{i,j}(\theta) = \theta_{|i-j|}$ for $j \neq i$. For example: when $d = 3$, $\theta = (\theta_1, \theta_2) \in (-1,1)^2$ and

$$\Sigma(\theta) = \begin{pmatrix} 1 & \theta_1 & \theta_2 \\ \theta_1 & 1\theta_1 & \\ \theta_2 & \theta_1 & 1 \end{pmatrix};$$

when $d = 4$, $\theta = (\theta_1, \theta_2, \theta_3) \in (-1,1)^3$ and

$$\Sigma(\theta) = \begin{pmatrix} 1 & \theta_1 & \theta_2 & \theta_3 \\ \theta_1 & 1 & \theta_1 & \theta_2 \\ \theta_2 & \theta_1 & 1 & \theta_1 \\ \theta_3 & \theta_2 & \theta_1 & 1 \end{pmatrix}.$$

Segers, vd Akker, and Werker (2014) show that:

Recent progress and results

- For $d = 3$ the Pseudo-Likelihood Estimator (PLE) $\widehat{\theta}_n^{ple}$ is semiparametric efficient.
- For $d = 4$, $\widehat{\theta}_n^{ple}$ is not efficient, and some times severely so. When $\theta = (0.494546, -0.450276, -0846249)$, the asymptotic relative efficiencies of the PLE with respect to the information bound are

$$(18.3\%,\ 19.8\%, 96.9\%).$$

- The PLE is semiparametric efficient for a large class of "factor models": if θ is a $d \times q$ matrix, $q < d$, $\Theta =$ an open subset of $\{\theta \in \mathbb{R}^{d\times q} : \quad (\theta\theta^T)_{jj} < 1,\ j = 1, \ldots, d\}$ and

$$\Sigma(\theta) \equiv \theta\theta^T + (I_d - \text{diag}(\theta\theta^T)).$$

4: Questions and open problems

- Semiparametric efficient estimation of the marginal distributions?
  - ▷ Can we improve on the marginal empirical distribution functions? (Apparently not known even for bivariate Gaussian copula model?)
  - ▷ Asymptotic behavior of the sieve estimators of Chen, Fan, and Tsyrennikov (2006)?
- Asymptotic behavior of the MLE's of θ based on the rank likelihood. (Rank likelihood is difficult to compute!)
- Rank-based semiparametric efficient estimators of θ for non-Gaussian copula's?
- Asymptotic theory for P. Hoff's "extended rank likelihood" (Hoff 2007, 2008)?
- What happens under model miss-specification? (Remember David X. Li (2000)!)

References & Cautions

Selected references:

- Chen, Fan, and Tsyrennikov (2006). JASA.
- Genest, Ghoudi, and Rivest (1995). Biometrika.
- Hoff (2007). Ann. Appl Stat.
- Hoff, Niu, and W (2014). Bernoulli
- Klaassen and W (1997). Bernoulli
- Liu, Han, Yuan, Lafferty, and Wasserman (2012). Ann. Statist.
- Segers, van den Akker, and Werker (2014). Ann. Statist.

References & Cautions

Cautions:

- Li, David X. (2000).
  On Default Correlation: a copula function approach.
  Journal of Fixed Income **9** (4): 4354.
- Salmon, F. (2009).
  Recipe for Disaster: The Formula That Killed Wall Street.
  Wired **17** (3).
- Mikosch (2006). Copulas: tales and facts.
  Extremes **9**, 3-20.

Xièxiè!