Summary of *Extending the Rank Likelihood for Semiparametric Copula Estimation*, by Peter Hoff

David Gerard

Department of Statistics
University of Washington
gerard2@uw.edu

May 2, 2013

Recall: General Social Survey data

INC	Respondent's Income
DEG	Highest degree obtained
CHILD	Number of children
PINC	Parent's income when respondent was 16
PDEG	Max(mother's degree, father's degree, na.rm $=$ T)
PCHILD	Number of siblings $+ 1$
AGE	Age of respondent

Recall: Which do we choose?

$$\begin{aligned} INC_i = &\beta_0 + \beta_1 CHILD_i + \beta_2 DEG_i + \beta_3 AGE_i \\ &+ \beta_4 PCHILD_i + \beta_5 PINC_i + \beta_6 PDEG_i + \epsilon_i \end{aligned}$$

or

$$\begin{aligned} CHILD_i \sim &\text{Pois}(\exp\{\beta_0 + \beta_1 INC_i + \beta_2 DEG_i + \beta_3 AGE_i \\ &+ \beta_4 PCHILD_i + \beta_5 PINC_i + \beta_6 PDEG_i\}) \end{aligned}$$

Will Sandwich Help?

Response	INC	CHILD	DEG
INC	NA	**1.103(0.112)**	7.025(<0.001)
CHILD	**0.005(0.00922)**	NA	-0.068(0.056)

Response	AGE	PCHILD	PINC	PDEG
INC	0.335(<0.001)	0.284(0.407)	4.070(0.001)	1.399(0.115)
CHILD	0.037(<0.001)	0.021(0.080)	-0.063(0.195)	-0.051(0.204)

Will Sandwich Help?

Response	INC	CHILD	DEG
INC	NA	**1.103(0.112)**	7.025(<0.001)
CHILD	**0.005(0.00922)**	NA	-0.068(0.056)

Response	AGE	PCHILD	PINC	PDEG
INC	0.335(<0.001)	0.284(0.407)	4.070(0.001)	1.399(0.115)
CHILD	0.037(<0.001)	0.021(0.080)	-0.063(0.195)	-0.051(0.204)

Using a sandwich estimator we have:

Response	INC	CHILD	...
INC	NA	**1.103(0.108)**	...
CHILD	**0.005(0.00939)**	NA	...

Recall: Setup

Let $y_{i,j}$ be the value of the j^{th} variable taken on by the i^{th} observational unit.

Gaussian Copula Sampling Model

$$\mathbf{z_1}, \ldots, \mathbf{z_n} | \mathbf{C} \sim \textit{i.i.d. multivariate normal}(\mathbf{0}, \mathbf{C})$$

$$y_{i,j} = F_j^{-1}[\Phi(z_{i,j})]$$

- Estimate association parameters without having to estimate marginals.
- Do this by only using the partial ordering induced by the data: $y_{k,j} < y_{i,j} \Rightarrow z_{k,j} < z_{i,j}$.
- A partial ordering is a total ordering without the *totality* condition (i.e., some elements may be incomparable).

The Extended Rank Likelihood

Let D be the set of $\mathbf{Z} := (z_{i,j})$'s that satisfy the partial ordering. Use the following "marginal likelihood" for inference:

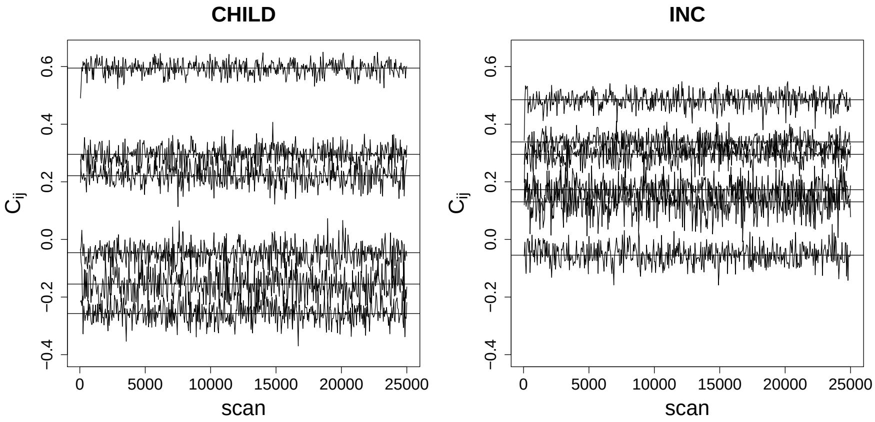
$$Pr(\mathbf{Z} \in D | \mathbf{C}, F_1, \ldots, F_p) = \int_D p(\mathbf{Z}|\mathbf{C}) d\mathbf{Z} = Pr(\mathbf{Z} \in D)$$

Gibbs Sampler

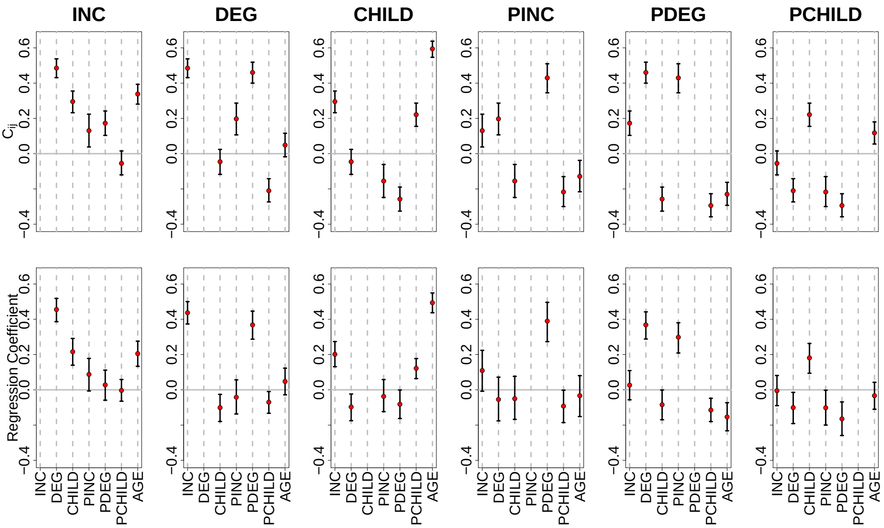
Full conditionals of the $z_{i,j}$'s are easy to derive, so we can implement a Gibb's sampler (using covariance matrix rather than correlation matrix, but doesn't matter for estimation).

1. Sample $z_{i,j}|\mathbf{Z}_{[-i,-j]}, \mathbf{V}$ from a truncated normal with the bounds set by the partial ordering and the conditional mean and variance found in the usual way: $\sigma_j^2 = \mathbf{V}_{[j,j]} - \mathbf{V}_{[j,-j]}\mathbf{V}_{[-j,-j]}^{-1}\mathbf{V}_{[-j,j]}$ and $\mu = \mathbf{V}_{[j,-j]}\mathbf{V}_{[-j,-j]}\mathbf{Z}_{[i,-j]}^T$.
2. Sample $\mathbf{V}$ from an inverse-Wishart distribution (if you use the conjugate prior).
3. Let $\mathbf{C}_{[i,j]} = \mathbf{V}_{[i,j]}/\sqrt{\mathbf{V}_{[i,i]}\mathbf{V}_{[j,j]}}$

Some Trace Plots



95% Posterior Credible Intervals



Visualizing Conditional Dependencies of GSS data.

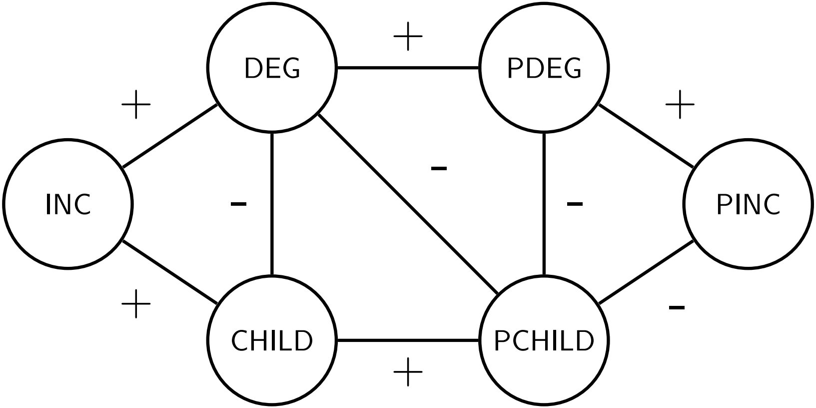


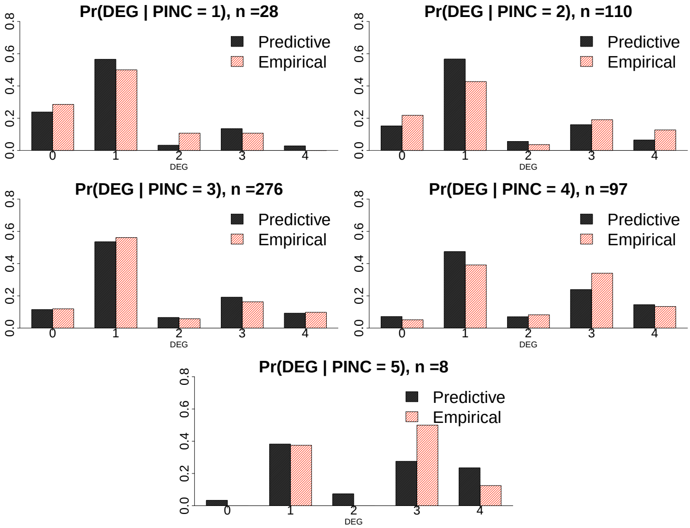
Figure: Reduced conditional dependence graph for the General Social Survey data.

Posterior Predictive

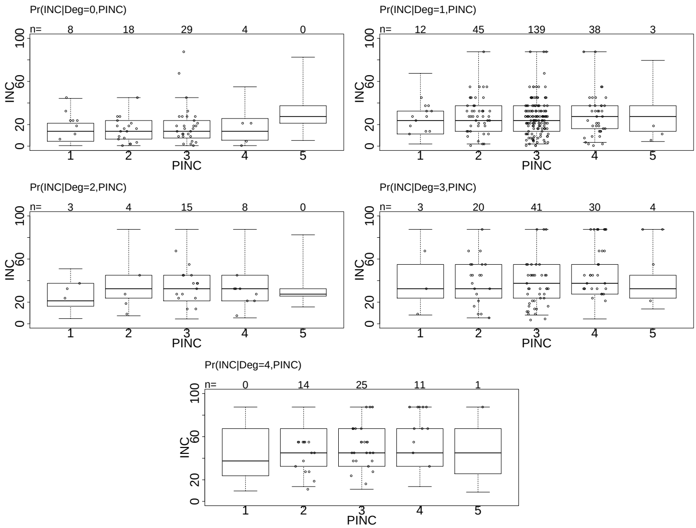
- We can also sample from a posterior predictive distribution to do inference on y's.
- Compare to empirical distributions for model checking.

1. sample $\mathbf{C} \sim p(\mathbf{C}|\mathbf{Z} \in D)$;
2. sample $\mathbf{z} \sim$ multivariate normal$(\mathbf{0}, \mathbf{C})$;
3. set $y_j = \hat{F}_j^{-1}(\Phi(z_j))$.

Posterior Predictive and Empirical distributions of DEG given PINC



Posterior Predictive and Empirical distributions of INC given DEG and PINC



Partial Sufficiency

- The paper proves that ranks are "G-sufficient" and "L-Sufficient" when we have continuous marginals.
- However, when the data are discrete the partial ordering does not have either of these properties.

Groups

Definition

A collection $\mathcal{G}$ of 1-1 transformations of $\mathcal{X}$ (the sample space) is a group if

1. For all $g_1, g_2 \in \mathcal{G}$, $g_1 g_2 \in \mathcal{G}$
2. For all $g \in \mathcal{G}, g^{-1} \in \mathcal{G}$.

Definition

Two points $x_1, x_2 \in \mathcal{X}$ are equivalent if there exists a $g \in \mathcal{G}$ such that $x_1 = gx_2$. The sets of equivalent points are the orbits of $\mathcal{G}$.

Definition

A function M is said to be maximally invariant if it is in 1-1 correspondence with the orbits of $\mathcal{G}$. i.e. $M(gx) = M(x)$ for all $g \in \mathcal{G}$ and $M(x_1) = M(x_2) \Rightarrow x_2 = gx_1$ for some $g \in \mathcal{G}$.

Groups

- $\mathcal{G}$ induces a group $\bar{\mathcal{G}}$ over the parameter space.
- Let $\bar{\mathcal{G}} = \{\bar{g} : \Omega \rightarrow \Omega \text{ s.t. } \bar{g}\theta = \theta' \text{ if } X \sim P_\theta \text{ and } gX \sim P_{\theta'}\}$.
- There are also maximally invariant parameters for $\bar{\mathcal{G}}$.

G-Sufficiency

- Under a group of transformations, $\mathcal{G}$, the maximally invariant statistic is called "G-sufficient" for the maximally invariant parameter.
- The ranks are the maximally invariant statistics under the group of continuous strictly increasing functions [Lehmann and Romano, 2005, pp 215 - 216]. You can also put the correlation matrix in a 1-1 correspondence with the induced group $\bar{\mathcal{G}}$'s orbits (but only if the marginals are continuous).
- Hence, the ranks are "G-sufficient" for the correlation matrix.
- Intuition: if we assume the marginals are unknown, then applying strictly increasing continuous functions to the data should not change the estimation problem

Simple Example

- Let $X_1, \ldots, X_n \sim i.i.d. N(\mu, \sigma^2)$, $\Omega = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ and $\mathcal{X} = \mathbb{R}$.
- Let $\mathcal{G} = \{g : g(\mathbf{x}) = \mathbf{x} + a\mathbf{1}, a \in \mathbb{R}\}$
- Then $g(\mathbf{X}) = (X_1 + a, \ldots, X_n + a)$ s.t. $X_i \sim$ i.i.d. $N(\mu + a, \sigma^2)$ and $\bar{\mathcal{G}} = \left\{\bar{g} : \bar{g}\begin{pmatrix} \mu \\ \sigma \end{pmatrix} = \begin{pmatrix} \mu + a \\ \sigma \end{pmatrix}\right\}$.
- Orbits of Ω under $\bar{\mathcal{G}}$ are defined by σ. So σ is the maximally invariant parameter.
- Orbits of $\mathcal{X}$ under $\mathcal{G}$ are defined by the differences from the mean $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$, so the differences are G-sufficient for σ and inference should only be based on the differences.
- Can extend these ideas to *minimal* G-sufficient ("maximal invariant reduction of minimal sufficient statistics") [Barnard, 1963]. Here, s^2.

L-Sufficiency

- For $\{F_1, \ldots, F_p\} \in \mathcal{F}$ the marginal distributions, a statistic $t(\mathbf{Y})$ is L-sufficient for $\mathbf{C}$ if
  1. $t(\mathbf{Y}_0) = t(\mathbf{Y}_1) \Rightarrow \sup_{\{F_1,\ldots,F_p\}\in\mathcal{F}} p(\mathbf{Y}_0|\mathbf{C}, F_1, \ldots, F_p) = \sup_{\{F_1,\ldots,F_p\}\in\mathcal{F}} p(\mathbf{Y}_1|\mathbf{C}, F_1, \ldots, F_p)$; and
  2. $p(t(\mathbf{Y})|\mathbf{C}, F_1, \ldots, F_p) = p(t(\mathbf{Y})|\mathbf{C})$
- If there are no nuisance parameters, L-sufficiency becomes "full sufficiency".
- Intuition: partition generated by L-sufficient statistic is "at least as fine" as the partition generated by the M.L.E. of **C** [Rémon, 1984] (i.e. MLE is function of ranks alone).

Future Work

- Develop maximum likelihood approach using the extended rank likelihood.
  - Lots of trouble with this. [Pettitt, 1982] used an approximation for the rank likelihood in the standard multivariate normal setting.
- Run simulation study against the competitors [Genest et al., 1995] (just plugged in ECDF); [Olsson, 1979] (estimate threshholding values when number of categories is known)
- Develop technique for other copulas, then simulate and see what info is lost by using the wrong copula.

References

Barnard, G. A. (1963).
Some logical aspects of the fiducial argument.
Journal of the Royal Statistical Society. Series B (Methodological), pages 111–114.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995).
A semiparametric estimation procedure of dependence parameters in multivariate families of distributions.
Biometrika, 82(3):543–552.

Hoff, P. D. (2007).
Extending the rank likelihood for semiparametric copula estimation.
The Annals of Applied Statistics, pages 265–283.

Lehmann, E. E. L. and Romano, J. P. (2005).
Testing statistical hypotheses.
Springer Science+ Business Media.

Olsson, U. (1979).
Maximum likelihood estimation of the polychoric correlation coefficient.
Psychometrika, 44(4):443–460.

Pettitt, A. (1982).
Inference for the linear model using a likelihood based on ranks.
Journal of the Royal Statistical Society. Series B (Methodological), pages 234–243.

Rémon, M. (1984).
On a concept of partial sufficiency: L-sufficiency.
International Statistical Review/Revue Internationale de Statistique, pages 127–135.