Large Sample Robustness Bayes Nets with Incomplete Information

Jim Smith and Ali Daneshkhah

Universities of Warwick and Strathclyde

Denmark PGM September 2008

Motivation

- We often worry about convergence of samplers etc. in a Bayesian analysis. How precise does the the prior on a BN have to be?
- In particular what is the overall effect of local and global independence assumptions on a given model?
- What are the overall inferential implications of using standard priors like product Dirichlets or product logistics?
- In general how hard do I need to think about these issues a priori when I know I will collect a large sample?

Messy Analyses

- Large BN - some expert knowledge incorporated.
- Nodes in our graph are systematically missing/ sample not random. Possible unidentifiablity even taking account of aliasing as $n \longrightarrow \infty$

$$
\begin{array}{ccccccccccccc}
 & & \theta_2 & & \theta_3 & & \theta_4 & & & \theta_5 & & & \\
 & & \downarrow & & \downarrow & & \downarrow & & & \downarrow & & & \\
\theta_1 & & \bullet & & \bullet & & \bullet & & & \bullet & & & \theta_6 \\
\downarrow & & \uparrow & \nearrow & & & \uparrow & \nearrow & & & & \swarrow & \\
\bullet & \leftarrow & \circ & \longleftarrow & \circ & \leftarrow & \circ & \rightarrow & \circ & \rightarrow & \bullet & & \\
 & \nearrow & & \nearrow & & \nearrow & \downarrow & & & \nwarrow & & & \\
\theta_7 & & \theta_8 & & \theta_9 & & \bullet & & & & & \theta_{11} & \\
 & & & & & & \uparrow & & & & & & \\
 & & & & & & \theta_{10} & & & & & &
\end{array}
$$

The Problems

- For a given prior only a **numerical** or **algebraic approximation** of posterior density. Just have **approximate summary statistics** (e.g. means, variances, sampled low dimensional margins, ...)
- Robustness issues: even for complete sampling. **Variation distance** $d_V\ (f, g) = \int |f - g|$ between two posteriors can **diverge quickly as sample size increases**, especially when the parameter space is large with outliers (Dawid, 1973) and more generally (Gustafson and Wasserman,1995).
- So when and how are posterior inferences strongly influenced by prior?
- **Local De Robertis** separations the key to addressing this issue!

About LDR

- Local De Robertis (LDR) separations are **easy to calculate** and extend natural parametrizations in exponential families.
- Have an intriguing **prior to posterior invariance** property.
- BN factorization of a density implies **linear** relationships between **clique** marginal separations and joint.
- **Bounds on the variation distance** between two posterior distributions associated with different priors **calculated explicitly** as a function of prior LDR bounds and posterior statistics associated with the functioning prior.
- **Bounds apply** posterior to an observed likelihood, even **when the sample density is misspecified**.

Contents

- De Robertis local Separations
- Some Properties of Local De Robertis Separations
- Some useful Theorems concerning LDR and BNs.
- What this means for the robustness of BN's

The Setting

- Let $g_0, (g_n)$ our **genuine** prior (posterior) density :$f_0, (f_n)$ our **functioning** prior (posterior) density
- Default for Bayes f_0 often products of Dirichlets
- $\mathbf{x}_n = (x_1, x_2, \ldots x_n)$, $n \geq 1$. with *observed* sample densities $\{p_n(\mathbf{x}_n|\boldsymbol{\theta})\}_{n\geq 1}$,
- With missing data, typically these sample densities are typically $\{p_n(\mathbf{x}_n|\boldsymbol{\theta})\}_{n\geq 1}$ (and hence f_n and g_n) intractable
- f_n therefore approximated either by drawing samples or algebraically.

A Bayes Rule Identity

Let $\Theta(n) = \{\boldsymbol{\theta} \in \Theta : p(\mathbf{x}_n|\boldsymbol{\theta}) > 0\}$ For all $\boldsymbol{\theta} \in \Theta(n)$ then

$$\begin{aligned}
\log g_n(\boldsymbol{\theta}) &= \log g_0(\boldsymbol{\theta}) + \log p_n(\mathbf{x}_n|\boldsymbol{\theta}) - \log p_g(\mathbf{x}_n) \\
\log f_n(\boldsymbol{\theta}) &= \log f_0(\boldsymbol{\theta}) + \log p_n(\mathbf{x}_n|\boldsymbol{\theta}) - \log p_f(\mathbf{x}_n)
\end{aligned}$$

where
$p_g(\mathbf{x}_n) = \int_{\boldsymbol{\theta}\in\Theta(n)} p(\mathbf{x}_n|\boldsymbol{\theta}) g_0(\boldsymbol{\theta}) d\boldsymbol{\theta}$, $p_f(\mathbf{x}_n) = \int_{\boldsymbol{\theta}\in\Theta(n)} p(\mathbf{x}_n|\boldsymbol{\theta}) f_0(\boldsymbol{\theta}) d\boldsymbol{\theta}$,
(When $\boldsymbol{\theta} \in \Theta \backslash \Theta(n)$ set $g_n(\boldsymbol{\theta}) = f_n(\boldsymbol{\theta}) = 0$)
So

$$\log f_n(\boldsymbol{\theta}) - \log g_n(\boldsymbol{\theta}) = \log f_0(\boldsymbol{\theta}) - \log g_0(\boldsymbol{\theta}) + \log p_g(\mathbf{x}_n) - \log p_f(\mathbf{x}_n)$$

From Bayes Rule to LDR

For any subset $A \subseteq \Theta(n)$ let

$$d_A^L(f,g) \triangleq \sup_{\boldsymbol{\theta} \in A} (\log f(\boldsymbol{\theta}) - \log g(\boldsymbol{\theta})) - \inf_{\boldsymbol{\phi} \in A} (\log f(\boldsymbol{\phi}) - \log g(\boldsymbol{\phi}))$$

Then since

$$\log f_n(\boldsymbol{\theta}) - \log g_n(\boldsymbol{\theta}) = \log f_0(\boldsymbol{\theta}) - \log g_0(\boldsymbol{\theta}) + \log p_g(\mathbf{x}_n) - \log p_f(\mathbf{x}_n)$$

for *any* sequence $\{p(\mathbf{x}_n|\boldsymbol{\theta})\}_{n\geq 1}$ - however complicated -

$$d_A^L(f_n, g_n) = d_A^L(f_0, g_0)$$

Isoseparation

$$d_A^L(f_n, g_n) = d_A^L(f_0, g_0)$$

- So for $A \subseteq \Theta(n)$ the posterior approximation of f_n to g_n is identical in quality to that of f_0 to g_0.

Isoseparation

$$d_A^L(f_n, g_n) = d_A^L(f_0, g_0)$$

- So for $A \subseteq \Theta(n)$ the posterior approximation of f_n to g_n is identical in quality to that of f_0 to g_0.
- When $A = \Theta(n)$ this property (De Robertis,1978) used for density ratio metrics and the specification of neighbourhoods.

Isoseparation

$$d_A^L(f_n, g_n) = d_A^L(f_0, g_0)$$

- So for $A \subseteq \Theta(n)$ the posterior approximation of f_n to g_n is identical in quality to that of f_0 to g_0.
- When $A = \Theta(n)$ this property (De Robertis,1978) used for density ratio metrics and the specification of neighbourhoods.
- Trivially posterior distances between densities can be calculated effortlessly from priors.

Isoseparation

$$d_A^L(f_n, g_n) = d_A^L(f_0, g_0)$$

- So for $A \subseteq \Theta(n)$ the posterior approximation of f_n to g_n is identical in quality to that of f_0 to g_0.
- When $A = \Theta(n)$ this property (De Robertis,1978) used for density ratio metrics and the specification of neighbourhoods.
- Trivially posterior distances between densities can be calculated effortlessly from priors.
- Separation of two priors lying in standard families can usually be expressed explicitly and always explicitly bounded.

Some notation

We will be especially interested in small sets A.

- Let $B(\boldsymbol{\mu};\rho)$ denote the open ball centred at $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)$ and of radius ρ
- Let
$$d^L_{\mu;\rho}(f,g) \triangleq d^L_{B(\boldsymbol{\mu};\rho)}(f,g)$$
- For any subset $\Theta_0 \subseteq \Theta$, let
$$d^L_{\Theta_0;\rho}(f,g) = \sup_{\mu \in \Theta_0} d^L_{\mu;\rho}(f,g)$$
- Obviously for any $A \subseteq B(\boldsymbol{\mu};\rho)$, $\boldsymbol{\mu} \in \boldsymbol{\Theta}_0 \subseteq \Theta$,
$$d^L_A(f,g) \leq d^L_{\Theta_0;\rho}(f,g)$$

Separation of two Dirichlets

.

- Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, $\theta_i, \alpha_i > 0, \sum_{i=1}^{k} \theta_i = 1$
- Let $f_0(\boldsymbol{\theta}|\boldsymbol{\alpha}_f)$ and $g_0(\boldsymbol{\theta}|\boldsymbol{\alpha}_g)$ be Dirichlet$(\boldsymbol{\alpha})$ so that

$$f_0(\boldsymbol{\theta}|\boldsymbol{\alpha}_f) \propto \prod_{i=1}^{k} \theta_i^{\alpha_{i,f}-1}, \quad g_0(\boldsymbol{\theta}|\boldsymbol{\alpha}_g) \propto \prod_{i=1}^{k} \theta_i^{\alpha_{i,g}-1}$$

- Let $\boldsymbol{\mu}_n = (\mu_{1,n}, \mu_{2,n}, \ldots, \mu_{k,n})$ be the mean of f_n If $\rho_n < \mu_n^0 = \min\{\mu_n : 1 \leq i \leq k\}$

$$d^L_{\mu;\rho_n}(f_0, g_0) \leq 2k\rho_n \left(\mu_n^0 - \rho_n\right)^{-1} \overline{\alpha}(f_0, g_0)$$

where

$$\overline{\alpha}(f_0, g_0) = k^{-1} \sum_{i=1}^{k} |\alpha_{i,f} - \alpha_{i,g}|$$

is the average distance between hyperparameters of f_0 and g_0.

Where Separations might be large

$$d^L_{\mu;\rho_n}(f_0, g_0) \leq 2\rho_n \left(\mu_n^0 - \rho_n\right)^{-1} \sum_{i=1}^{k} \left|\alpha_{i,f} - \alpha_{i,g}\right|$$

- So $d^L_{\mu;\rho_n}(f_0, g_0)$ is uniformly bounded whenever μ_n all away from 0 and converging approximately linearly in n.
- OTOH if f_n tends to mass near a zero probability, then even when $\overline{\alpha}(f, g)$ is small, it can be shown that at least some likelihoods will force the variation distance between the posterior densities to stay large for increasing n: Smith(2007). The smaller the smallest probability tended to the slower any convergence.

BN's with local and global independence

If functioning prior $f(\boldsymbol{\theta})$ and genuine prior $g(\boldsymbol{\theta})$ factorize on subvectors $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots \boldsymbol{\theta}_k\}$ so that

$$f(\boldsymbol{\theta}) = \prod_{i=1}^{k} f_i(\boldsymbol{\theta}_i), \quad g(\boldsymbol{\theta}) = \prod_{i=1}^{k} g_i(\boldsymbol{\theta}_i)$$

where $f_i(\boldsymbol{\theta}_i)$ $(g_i(\boldsymbol{\theta}_i))$ are the functioning (genuine) margin on$\boldsymbol{\theta}_i$, $1 \leq i \leq k$, then (like K-L separations)

$$d_A^L(f, g) = \sum_{i=1}^{k} d_{A_i}^L(f_i, g_i)$$

So local prior distances grow linearly with no. of defining conditional probability vectors.

Some conclusions

- BN's with larger nos of edges intrinsically **less stable**
- However - like K-L - marginal densities are never more separated than their joint densities - so if a utility is only on a particular margin then these distances may be much less.
- Bayes Factors automatically select simpler models but note also inferences of a more complex model tends to be more sensitive to wrongly specified priors.

Disaster?

- There are certain features in the prior which will always endure.
- If there is a point where locally LDR diverges - in a sense which violates the condition above then it is possible to construct a "regular" likelihood such that the variation distance between posteriors remains bounded away from zero as $n \longrightarrow \infty$.
- However if the mass is converging on to a small set because then we can focus on a small set A
- Usually $d_A^L(f_0, g_0)$ is small when A lies in a small ball.

Salvation!

- When n is large A will lie in a small ball with high probability
- it is usually reasonable to assume that f_0 and g_0 for A lying in a small ball $d_A^L(f_0, g_0)$ is small.
- Can usually assume for open balls $B(\boldsymbol{\mu}; \rho)$ centred at $\boldsymbol{\mu}$ and of radius ρ, $f_0, g_0 \in \mathcal{F}(\Theta_0, M(\Theta_0), p(\Theta_0))$ meaning

$$\begin{aligned}
\sup_{\boldsymbol{\theta},\boldsymbol{\phi} \in B(\boldsymbol{\mu};\rho))} |\log f_0(\boldsymbol{\theta}) - \log f_0(\boldsymbol{\phi})| &\leq M(\Theta_0)\rho^{0.5p(\Theta_0)} \\
\sup_{\boldsymbol{\theta},\boldsymbol{\phi} \in B(\boldsymbol{\mu};\rho))} |\log g_0(\boldsymbol{\theta}) - \log g_0(\boldsymbol{\phi})| &\leq M(\Theta_0)\rho^{0.5p(\Theta_0)}
\end{aligned}$$

A simple smoothness/roughness condition

- When $p(\Theta_0) = 2$ just demands that $\log f_0$ and $\log g_0$ both have bounded derivatives within the set Θ_0 - used to determine where f_n concentrates its mass. Then it is easily shown (see Smith and Rigat,2008) that

$$d^L_{\Theta_0,\rho}(f,g) \leq 2M(\Theta_0)\rho^{1/2p(\Theta_0)}$$

- So **rate** of convergence to zero of $d^L_{\Theta_0,\rho}(f,g)$ governed by the "roughness" parameter $p(\Theta_0)$.
- This is the always true for densities with inverse polynomial tails like the **Student t density**. If densities have tighter tails than this then is also true provided continuously differentiable on a closed bounded interval Θ_0.
- For continuous f, g when Θ_0 closed and bounded (so no divergence due to outliers) $d^L_{\Theta_0,\rho}(f,g)$ converges to zero.

Introducing smoothness accidentally

Consider the typical hierarchical models used in e.g. BUGS

$$\begin{array}{ccccc} \mathbf{X}_1 & & & & \mathbf{X}_2 \\ \uparrow & & & & \uparrow \\ \theta_1 & \leftarrow & \theta & \rightarrow & \theta_2 \end{array}$$

e.g. $i = 1, 2$, $\theta_i = \theta \div \varepsilon_i$ where ε_i is an independent error term, (Gaussian, Student t) etc. provided the error term is smooth then this automatically forces the prior margin $g_0(\theta_1, \theta_2)$ to be smooth (even if θ if discrete) regardless of the smoothness of θ.

Moral: nearly all conventional hierarchical BN's with enough depth have implicit priors on parameters of the likelihood are smooth in the sense above (making them robust in the sense below).

But why worry about LDR separation?

- Without the LDR condition above large sample variation convergence cannot hold in general.
- Conversely with a regularity condition and a technical devise convergence will happen.   .
- **Regularity Condition**. Call a genuine prior *c-rejectable* if the ratio of marginal likelihood $\frac{p_g(\mathbf{x})}{p_f(\mathbf{x})} < c$.

If f_0 does not explain the data much better than g_0 we would expect this ratio to be small - certainly not c- rejectable for a moderately large values of $c \geq 1$.

A Second Tail convergence condition

- Say density f $\Lambda-$*tail dominates* a density g if

$$\sup_{\boldsymbol{\theta}\in\Theta} \frac{g(\boldsymbol{\theta})}{f(\boldsymbol{\theta})} = \Lambda < \infty$$

  When $g(\boldsymbol{\theta})$ is bounded then this condition requires that the tail convergence of g is no slower than f.
- Condition met provided f_0 is chosen to have a flatter tail than g_0.
- Note: flat tailed priors recommended for robustness on other grounds e.g. O'Hagan and Forster (2003)

A typical result(Smith and Rigat(2007)

Theorem

If the genuine prior g_0 *is not* c *rejectable with respect to* f_0 *,* f_0 $\Lambda-$*tail dominates* g_0 *and* $f_0, g_0 \in \mathcal{F}(\Theta_0, M(\Theta_0), p(\Theta_0))$*.then*

$$d_V(f_n, g_n) \le \inf_{\rho>0} \{ T_n(1, \rho_n) + 2T_n(2, \rho_n) : B(\boldsymbol{\mu}_n, \rho_n) \subset \Theta_0 \} \tag{1}$$

where

$$T_n(1, \rho_n) = \exp d^L_{\boldsymbol{\mu},\rho}(f, g) - 1 \le \exp \left\{ 2M\rho_n^{p/2} \right\} - 1$$

and

$$T_n(2, \rho_n) = (1 + c\Lambda) F_n (\boldsymbol{\theta} \notin B(\boldsymbol{\mu}_n; \rho_n))$$

Easy to bound $F_n (\boldsymbol{\theta} \notin B(\boldsymbol{\mu}_n; \rho_n))$ in many ways explicitly using Chebychev type inequalities: Smith (2007). Example of bound is given below, specified in terms of the posterior means and variances of the vector of parameters under f_n routinely approximated.

An Example of an Explicit Bound

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ and $\mu_{j,n}, \sigma^2_{jj,n}$ denote the mean and variance of θ_j, $1 \leq j \leq k$ under f_n. Using Chebychev bounds in Tong (1980), p153), writing $\boldsymbol{\mu}_n = (\mu_{1,n}, \mu_{2,n}, \ldots \mu_{k,n})$

$$F_n \left(\boldsymbol{\theta} \notin B(\boldsymbol{\mu}_n; \rho_n)\right) \leq k\rho_n^{-2} \sum_{j=1}^{k} \sigma^2_{jj,n}$$

where writing $\sigma_n^2 = k \max_{1 \leq j \leq k} \sigma^2_{j,n}$ this implies

$$T_n(2, \rho_n) \leq c\Lambda\sigma_n^2\rho_n^{-2}$$

- e.g. if $\sigma_n^2 \leq n^{-1}\sigma^2$ for some value σ^2, $T_n(2, \rho_n) \to 0$ provided $\rho_n^2 \leq n^r \rho^2$ where $0 < r < 1$.
- In practice for a given data set we just have an approximate value of σ_n^2 we can plug in.

Inference on margins separation

When $A_1$is a restriction of A to $\boldsymbol{\theta}_1$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $f_1(\boldsymbol{\theta}_1), g_1(\boldsymbol{\theta}_1)$ contin. margins of $f(\boldsymbol{\theta})$ and $g(\boldsymbol{\theta})$, resp. then

$$d^L_{A_1}(f_1, g_1) \leq d^L_A(f, g)$$

- If f_n converges on a margin, then even if the model is unidentified, provided $f_0, g_0 \in \mathcal{F}(\Theta_0, M(\Theta_0), p(\Theta_0))$, then for large n, f_n will be a good surrogate for g_n.
- BN's with interior systematically hidden variables are unidentified. However if a utility function is only on manifest variables, in standard scenarios under above conditions $d_V(f_{1,n}, g_{1,n}) \longrightarrow 0$ at a rate of at least $\sqrt[3]{n}$.
- Instability only on posteriors of functions of probabilities associated with the hidden variables conditional on the manifest variables.

A Simple Example: The Star tree

$$
\begin{array}{ccccccccc}
\theta_5 & \longrightarrow & \bullet & & & & \bullet & \longleftarrow & \theta_6 \\
 & & \downarrow & & & & \downarrow & & \\
\theta_2 & \longrightarrow & \circ & & \theta_1 & & \circ & \leftarrow & \theta_3 \\
 & & \downarrow & \nwarrow & \downarrow & \nearrow & \downarrow & & \\
\theta_7 & \longrightarrow & \bullet & & \odot & & \bullet & \leftarrow & \theta_8 \\
 & & & & \downarrow & & & & \\
 & & & & \circ & & & & \\
 & & & \swarrow & \uparrow & \searrow & & & \\
\theta_9 & \rightarrow & \bullet & & \theta_4 & & \bullet & \leftarrow & \theta_{10}
\end{array}
$$

- d -sep. tells us $\theta_1 \coprod X | \boldsymbol{\theta}_{\backslash 1}$. So what we put in as a prior for $\theta | \boldsymbol{\theta}_{\backslash 1}$ is what we get out
- However model $\Rightarrow \theta_1$ a **function** of $\boldsymbol{\theta}_{\backslash 1}$ (up to aliasing), so actually no deviation consistent with the model.

Departures from Parameter Independence

$$
\begin{aligned}
f(\boldsymbol{\theta}) &= f_1(\theta_1)\prod_{i=2}^{k} f_{i|.}(\theta_i|\boldsymbol{\theta}_{pa_i}) \\
g(\boldsymbol{\theta}) &= g_1(\theta_1)\prod_{i=2}^{k} g_{i|.}(\theta_i|\boldsymbol{\theta}_{pa_i})
\end{aligned}
$$

we then have the inequality

$$
d_A^L(f,g) \leq \sum_{i=2}^{k} d_{A[i]}^L(f_{[i]}, g_{[i]})
$$

where $f_{[i]}, g_{[i]}$ are respectively the margin of f and g on the space $\Theta[i]$ of the i^{th} variable and its parents. So distances bounded by sums on distances on cliques margins.

Uniformly A Uncertain

Suppose g is uniformly A uncertain and factorises as f and

$$\sup_{g} \sup_{\boldsymbol{\theta}_i, \boldsymbol{\phi}_i \in A[i]} \left\{ \log f_{i|}(\boldsymbol{\theta}) - \log g_{i|}(\boldsymbol{\theta}) - \log f_{i|}(\boldsymbol{\phi}) + \log g_{i|}(\boldsymbol{\phi}) \right\}$$

is not a function of $\boldsymbol{\theta}_{pa_i}$ $2 \leq i \leq n$, then we can write

$$d_A^L(f, g) = \sum_{i=1}^{k} d_{A[i]}^{L*}(f_{i|}, g_{i|})$$

- Separation between the joint densities f and g sum of the separation between its component conditionals $f_{i|}$ and $g_{i|}$ $1 \leq i \leq k$.
- Bounds can be calculated *even* when the likelihood destroys the factorisation of the prior. So the critical property we assume here is the fact that we believe a priori that f respects the same factorisation as g.

Conclusions

- Bayesian inference on BN's is most stable to prior settings the simpler the model

Conclusions

- Bayesian inference on BN's is most stable to prior settings the simpler the model
- For large samples general total variation robustness is lost when posterior masses concentrate near a zero probability.

Conclusions

- Bayesian inference on BN's is most stable to prior settings the simpler the model
- For large samples general total variation robustness is lost when posterior masses concentrate near a zero probability.
- However robustness can sometimes be retrieved if that probability is not appear in a utility function.

Conclusions

- Bayesian inference on BN's is most stable to prior settings the simpler the model
- For large samples general total variation robustness is lost when posterior masses concentrate near a zero probability.
- However robustness can sometimes be retrieved if that probability is not appear in a utility function.
- Even for moderate sized samples, explicit bounds on the effects of priors can be calculated on line.

Conclusions

- Bayesian inference on BN's is most stable to prior settings the simpler the model
- For large samples general total variation robustness is lost when posterior masses concentrate near a zero probability.
- However robustness can sometimes be retrieved if that probability is not appear in a utility function.
- Even for moderate sized samples, explicit bounds on the effects of priors can be calculated on line.
- In regular problems, these bounds usually contract surprisingly quickly as data increases.

A Few references

Daneseshkhah,A (2004) "Estimation in Causal Graphical Models" PhD Thesis University of Warwick.

DeRobertis, L. (1978) "The use of partial prior knowledge in Bayesian inference" Ph.D. idssertation, Yale Univ.

Gustafson, P. and Wasserman, L. (1995) "Local sensitivity diagnostics for Bayesiain inference" Annals Statist ,23 , , 2153 - 2167

French, S. and Rios Insua, D.(2000) "Statistical Decision Theory" Kendall's Library of Statistics Arnold

O'Hagan, A and Forster, J (2004) "Bayesian Inference" Kendall's Advanced Theory of Statistics, Arnold

A few more References

Smith, J.Q."Local Robustness of Bayesian Parametric Inference and Observed Likelihoods" CRiSM Res Rep 07-08

Smith, J.Q. and Rigat, F.(2008) "Isoseparation and Robustness in Finite Parameter Bayesian Inference" CRiSM Res Rep

Smith,J.Q. and Croft, J. (2003) "Bayesian networks for discrete multivariate data" J of Multivariate Analysis 84(2), 387 -402

Tong, Y.L.(1980) "Probability Inequalities in Multivariate Distributions" Academic Press New York

Wasserman, L.(1992a) "Invariance properties of density ratio priors" Ann Statist, 20, 2177- 2182