

2009

# Uniformly Hyper-Efficient Bayes Inference in a Class of Nonregular Problems

Danial J. Nordman

*Iowa State University, [dnordman@iastate.edu](mailto:dnordman@iastate.edu)*

Stephen B. Vardeman

*Iowa State University, [vardeman@iastate.edu](mailto:vardeman@iastate.edu)*

Melissa Ann Bingham

*University of Wisconsin - La Crosse*

Follow this and additional works at: [http://lib.dr.iastate.edu/imse\\_pubs](http://lib.dr.iastate.edu/imse_pubs)

 Part of the [Industrial Engineering Commons](#), [Statistics and Probability Commons](#), and the [Systems Engineering Commons](#)

The complete bibliographic information for this item can be found at [http://lib.dr.iastate.edu/imse\\_pubs/147](http://lib.dr.iastate.edu/imse_pubs/147). For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

# Uniformly Hyper-Efficient Bayes Inference in a Class of Non-Regular Problems

Daniel J. Nordman, Stephen B. Vardeman, and Melissa A. Bingham  
Iowa State University

August 12, 2008

## Abstract

We present a tractable class of non-regular continuous statistical models where 1) likelihoods have multiple singularities and maximum likelihood is intrinsically unavailable, but 2) Bayes procedures achieve convergence rates better than  $n^{-1}$  across the whole parameter space. In fact, for every  $p > 1$ , there is a member of the class for which the posterior distribution is consistent at rate  $n^{-p}$  uniformly in the parameter.

## 1 Introduction

It is common in models satisfying conventional regularity conditions to find that both likelihood-based and Bayes methods of inference have convergence rates of  $n^{-1/2}$ . In such problems, "super-efficiency" (including rates better than  $n^{-1/2}$ ) can be achieved at a relatively few points in a large parameter space by ad hoc modification of standard procedures, but on the whole, the  $n^{-1/2}$  rate is the standard benchmark for "efficiency" of inference. In this regard, see Chapter 8 of van der Vaart (1998), Chapter 6 of Lehmann and Casella (1998), Section 4.5 of Shao (2003), and Chapter 7 of Schervish (1995). Some non-regular models (like the Uniform  $(0, \theta)$  case) allow rates as fast as  $n^{-1}$  for both likelihood and Bayes methods, but these are usually considered pathological in that the parameter is a boundary point of the support of the model.

Here we present a class of simple examples in which ordinary one-sample likelihood-based methods are intrinsically not available, in that the likelihood has a singularity corresponding to every observation, but where Bayes methods have convergence rates far in excess of the usual  $n^{-1/2}$  benchmark for "efficiency" and even exceeding the  $n^{-1}$  rate known in problems where the parameter defines a model's support. And this not only at a few points of a continuous parameter space, but uniformly across it.

## 2 A Class of Families of Circular Distributions With Unbounded Densities and a Corresponding One-Sample Inference Problem

We consider continuous distributions on  $[-\pi, \pi)$  with probability densities

$$f(x|\mu, \alpha) = \frac{C(\alpha)}{||x - \mu|_{2\pi}|^\alpha} I[-\pi < x < \pi]$$

where  $I[\cdot]$  denotes the indicator function,

$$|y|_{2\pi} \equiv \begin{cases} y & \text{if } |y| < \pi \\ y - 2\pi & \text{if } y > \pi \\ y + 2\pi & \text{if } y < -\pi, \end{cases}$$

and for  $0 < \alpha < 1$

$$C(\alpha) = \left( \int_{-\pi}^{\pi} \frac{1}{|x|^\alpha} dx \right)^{-1}.$$

For fixed  $\alpha$ , the set of densities  $f(x|\mu, \alpha)$  for  $\mu \in [-\pi, \pi)$  specify what is essentially a "location family" on the unit circle, with the parameter  $\mu$  specifying the location/"direction" of greatest probability density. Our interest here is in the one-sample inference problem for  $\mu$ .

If  $X_1, X_2, \dots, X_n$  are iid with pdf  $f(x|\mu, \alpha)$ , the (stage- $n$ ) likelihood function

$$L_n(\mu) = \prod_{i=1}^n f(X_i|\mu, \alpha) \tag{1}$$

has a singularity at every  $X_i$ , the problem is non-regular, and there is, for example, no obvious way to use the principle of maximum likelihood in inference for  $\mu$ . However, Bayes inference for  $\mu$  is in principle straightforward. For sake of simplicity (and symmetry) we will consider a Uniform  $(-\pi, \pi)$  prior distribution for  $\mu$ , and under this prior the posterior density for  $\mu$  is proportional to the likelihood (1).

In what follows, we provide evidence that under the  $\mu$  model, a distribution specified by a pdf proportional to  $L_n(\cdot)$  converges to a point mass at  $\mu$  at a rate  $n^{-1/(1-\alpha)}$  uniformly in  $\mu$ . This says that not only are Bayes methods straightforward, but their convergence rate is much better than anything one might naively expect (especially given the "regular" and "efficient" language routinely applied in problems where the  $n^{-1/2}$  rate holds).

## 3 Theoretical Evidence for Hyper-Efficiency

The one-sample problem just described is symmetric and it is obvious that whatever can be said about the behavior of the posterior density under the  $\mu = 0$  version of the model is equally true (at least upon applying the data

transformation  $[\cdot - \mu]_{2\pi}$ ) about the posterior density for any other  $\mu$ . So we consider the behavior of the log-likelihood near  $\mu = 0$  under the  $\mu = 0$  version of the model.

Let

$$l_n(\mu) = \ln(L_n(\mu)) = \sum_{i=1}^n \ln(f(X_i|\mu, \alpha))$$

Then,

$$\begin{aligned} l_n(0) - l_n(\delta) &= -\alpha \sum_{i=1}^n \left\{ \ln |[X_i]_{2\pi}| - \ln |[X_i - \delta]_{2\pi}| \right\} \\ &= \alpha \sum_{i=1}^n \left\{ \ln |[X_i - \delta]_{2\pi}| - \ln |X_i| \right\} \end{aligned} \quad (2)$$

represents the difference in the stage- $n$  log posterior density at 0 and at  $\delta$ .

Under our contention that the posterior distribution is consistent for  $\mu = 0$  under the iid  $f(x|0, \alpha)$  model at rate  $n^{-1/(1-\alpha)}$ , we should expect that for any constant  $c \neq 0$ , random differences

$$l_n(0) - l_n\left(\frac{c}{n^{1/(1-\alpha)}}\right) \quad (3)$$

go neither to 0 nor to  $\infty$  in probability. In fact, it is possible to prove that, under the  $\mu = 0$  model, these differences (3) converge in distribution. On the other hand, the differences (2) will converge trivially to zero at any rate for  $\delta$  faster than  $n^{-1/(1-\alpha)}$ . These two convergence results, made formal in Proposition 1, together provide evidence that  $n^{-1/(1-\alpha)}$  is exactly the right scale at which to view the posterior density to see something non-trivial.

**Proposition 1** *Let  $c \neq 0$  and suppose  $X_1, X_2, \dots$  are iid with pdf  $f(x|0, \alpha)$ . Then, there is a non-degenerate probability distribution  $G_c$  such that, as  $n \rightarrow \infty$ ,*

$$l_n(0) - l_n\left(\frac{c}{n^{1/(1-\alpha)}}\right) \xrightarrow{D} G_c.$$

*Alternatively, if  $p > 1/(1-\alpha)$ , then*

$$l_n(0) - l_n\left(\frac{c}{n^p}\right) \xrightarrow{P} 0.$$

**Proof.** For  $0 < |\delta| < \pi$  and  $\iota = \sqrt{-1}$ , let  $\phi_{n,\delta}(t) \equiv \mathbb{E} \exp\{it[l_n(0) - l_n(\delta)]\}$ ,  $t \in \mathbb{R}$ , denote the characteristic function of (2) under the  $\mu = 0$  model. Given  $t \in \mathbb{R}$ , also define complex-valued functions of a real-valued argument  $y \geq 0$  as

$$\begin{aligned} g_t(y) &= C(\alpha) \int_{y/\pi}^{\infty} (1 - \exp\{it\alpha \ln(x+1)\}) x^{\alpha-2} dx \\ h_t(y) &= C(\alpha) \int_{y/(\pi-y)}^{\infty} (1 - \exp\{it\alpha \ln|x-1|\}) x^{\alpha-2} dx, \end{aligned}$$

which satisfy  $\sup\{|g_t(y)| + |h_t(y)| : y \geq 0\} < \infty$ , since the integrands above are bounded by the integrable function  $2C(\alpha)x^{\alpha-2}[I(x > 1/2) + x|t|I(x \leq 1/2)]$ ,  $x \in (0, \infty)$ . Considering the  $n = 1$  case, direct integration yields

$$\phi_{1,\delta}(t) = 1 - |\delta|^{1-\alpha} [g_t(|\delta|) + h_t(|\delta|)] + E_t(\delta), \quad t \in \mathbb{R}, \quad (4)$$

where, for  $\pi_\delta = \pi - 2^{-1}|\delta|$ ,

$$E_t(\delta) = C(\alpha) \int_{-|\delta|}^{|\delta|} \frac{\exp\{it\alpha[\ln(\pi_\delta - x) - \ln(\pi_\delta + x)]\}}{(\pi_\delta + x)^\alpha} dx + \frac{1}{2} \left[ \left(1 - \frac{|\delta|}{\pi}\right)^{1-\alpha} - 1 \right].$$

As  $\delta \rightarrow 0$ , both  $g_t(|\delta|) \rightarrow g_t(0)$  and  $h_t(|\delta|) \rightarrow h_t(0)$  follow by the Dominated Convergence Theorem while  $E_t(\delta) = O(|\delta|) = o(|\delta|^{1-\alpha})$  holds (for fixed  $t$ ).

Setting  $\delta_n = c/n^{1/(1-\alpha)}$  for  $\delta$  in (4), we may then deduce that the characteristic function  $\phi_{n,\delta_n}(t) = [\phi_{1,\delta_n}(t)]^n$  of (3) converges

$$\begin{aligned} \phi_{n,\delta_n}(t) &= \left(1 - \frac{|c|^{1-\alpha}}{n} [g_t(|\delta_n|) + h_t(|\delta_n|)] + o(1/n)\right)^n \\ &\rightarrow \phi(t) \equiv \exp\{-|c|^{1-\alpha} [g_t(0) + h_t(0)]\}, \quad t \in \mathbb{R}, \end{aligned}$$

as  $n \rightarrow \infty$ . It can be checked that  $\phi(\cdot)$  is continuous and nonnegative definite (as the limit of characteristic functions which are inherently nonnegative definite) with  $\phi(0) = 1$ . Hence, by the Bochner-Khinchine theorem (Chung, 1974),  $\phi(\cdot)$  is a legitimate characteristic function for some distribution  $G_c$ , establishing the first result of Proposition 1.

By setting  $\delta_n = c/n^p$  in (4), the characteristic function of  $l_n(0) - l_n(c/n^p)$  also converges as  $n \rightarrow \infty$ ,

$$\phi_{n,\delta_n}(t) = \left[1 + O\left(\frac{1}{n^{p(1-\alpha)}}\right)\right]^n = [1 + o(1/n)]^n \rightarrow 1, \quad t \in \mathbb{R}.$$

This implies that  $l_n(0) - l_n(c/n^p)$  converges in distribution to zero under the  $\mu = 0$  model, proving the second assertion of Proposition 1. ■

## 4 Simulation Evidence of Hyper-Efficiency

To provide concrete illustration of the uniform hyper-efficiency of Bayes methods suggested by Proposition 1, we used a Metropolis-Hastings algorithm to simulate from posteriors for the one-sample problem described in Section 2 for sample sizes  $n = 10, 100, 1000$ , and  $10000$  for each of the values  $\alpha = .3\bar{3}, .5$ , and  $.6\bar{6}$  (these parameters correspond to rates of  $n^{-1.5}, n^{-2}$ , and  $n^{-3}$ ). We made central 90% credible intervals for  $\mu$  based on the simulated posteriors. Table 1 contains the corresponding estimated coverage probabilities and median interval lengths for the  $(\alpha, n)$  pairs (based on 1000 simulated data sets for each pair and an MCMC sample of size 20000 from each posterior). Figure 1 is a scatterplot of

Table 1: Estimated coverage probabilities and median 90% interval lengths for the  $(\alpha, n)$  pairs

$(\alpha, n)$	Coverage	Median Length
$(.3\bar{3}, 10)$	0.938	4.064969
$(.3\bar{3}, 100)$	0.895	0.225804
$(.3\bar{3}, 1000)$	0.889	0.007573
$(.3\bar{3}, 10000)$	0.910	0.000244
$(.5, 10)$	0.929	1.374810
$(.5, 100)$	0.890	0.018718
$(.5, 1000)$	0.863	0.000144
$(.5, 10000)$	0.895	$1.746534 \times 10^{-6}$
$(.6\bar{6}, 10)$	0.846	0.105236
$(.6\bar{6}, 100)$	0.839	0.000122
$(.6\bar{6}, 1000)$	0.866	$1.378716 \times 10^{-7}$
$(.6\bar{6}, 10000)$	0.886	$2.122745 \times 10^{-10}$

log median interval length versus log sample size for the three values of  $\alpha$ , with best fitting lines of the form

$$\ln(\text{median length}) = \beta - \left( \frac{1}{1 - \alpha} \right) \ln(n)$$

added (we fit only the coefficients  $\beta$ ). These provide clear empirical evidence of the hyper-efficiency of the Bayes methods.

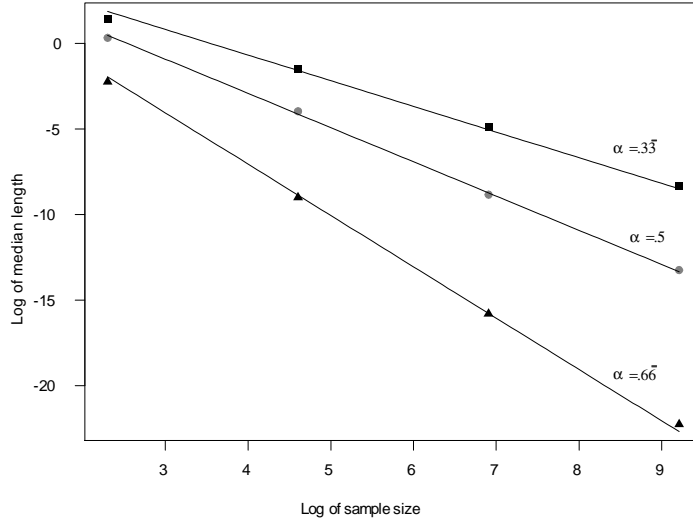


Figure 1: Log median interval length versus log sample size along with best fitting lines for  $\alpha = .33, .5$ , and  $.66$

## 5 Conclusion

We first encountered the phenomenon illustrated here in the related but more complicated context of Bingham, Vardeman, and Nordman (2008) considering inference on  $3 \times 3$  rotation matrices. There we empirically found Bayes methods to dominate quasi-likelihood methods (that standard arguments show to have convergence rate  $n^{-1/2}$ ), and thank Prof. Michael Stein for pointing out to us that our simulations suggested that the Bayes rate of convergence is  $n^{-1}$  in the matrix inference scenario. That realization led us to look for a simple class of examples illustrating the potential hyper-efficiency of Bayes methods in non-regular problems where there are no boundary-of-support-set issues. Now, while we don't know how far the basic notions here extend, we do not expect that the phenomenon we illustrate is particularly rare or pathological. Rather, we think that standard "regularity" (that really refers to mathematical convenience) and "efficiency" language potentially misleads us into expecting that relatively slow rates associated with smooth problems are somehow ideal.

Contrast between the present set of examples and von Mises circular models is perhaps instructive. The family of distributions on  $[-\pi, \pi)$  with probability densities

$$v(x|\mu, \kappa) = D(\kappa) \exp(\kappa \cos(x - \mu)) I[-\pi < x < \pi]$$

where

$$D(\kappa) = \left( \int_{-\pi}^{\pi} \exp(\kappa \cos(x)) dx \right)^{-1}$$

for  $\mu \in [-\pi, \pi)$  and fixed value of  $\kappa \geq 0$  is a “regular” location/direction model on the unit circle and the rate  $n^{-1/2}$  holds for both maximum likelihood and Bayes procedures.

So it is not the symmetry of our models nor the compactness of the (closure of) the parameter space that enables the huge convergence rates. Rather, it is the singularities in the likelihood. These prove to be not something to somehow smooth away (by, for example, employing a smooth quasi-likelihood), but rather something to be exploited. And the capacity to exploit such singularities in an automatic and effortless manner is arguably yet another virtue of Bayes technology.

## 6 References

Bingham, M.A., Vardeman, S.B., and Nordman, D.J. (2008). “Bayes One-Sample and One-Way Random Effects Analyses for 3-D Orientations with Application to Materials Science,” submitted to *The Annals of Applied Statistics*.

Chung, K. L. (1974). *A Course in Probability Theory*, 2nd edition. Springer-Verlag, New York.

Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd edition. Springer-Verlag, New York.

Schervish, M.J. (1995). *Theory of Statistics*. Springer-Verlag, New York.

Shao, J. (2003). *Mathematical Statistics*, 2nd edition. Springer-Verlag, New York.

van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.