BAYESIAN INFERENCE UNDER LONG-RANGE DEPENDENCE: DAMAGE, DETECTION COMPUTATION, ASYMPTOTICS

by

Chuanbo Zang and Anirban DasGupta
Purdue University

Technical Report #98-01

Department of Statistics
Purdue University
West Lafayette, IN USA

January 1998

◇

BAYESIAN INFERENCE UNDER LONG-RANGE DEPENDENCE: DAMAGE, DETECTION, COMPUTATION, ASYMPTOTICS*

Chuanbo Zang Anirban DasGupta

Purdue University

Abstract

It has been known for some time that many standard procedures of iid statistical theory perform very poorly under long range dependence of the observations. Moreover, empirical evidence suggests that detection of such long range dependence is itself rather difficult. Most of the available results have concentrated on classical methods and finding alternatives using frequentist criteria.

In this article, we first present a set of results to formally show that distinguishing between different values of the long memory parameter is a hard problem; the Bayes factor and its asymptotic distribution are studied; next, point and interval estimation of the mean, estimation of the variance, and testing of hypotheses on the mean are addressed in a Bayesian framework. The Brown identities and the Brown-Gajek lower bounds on Bayes risk are used to derive the asymptotic efficiency of simply computable estimates, some of which do not need specification of the long memory parameter. Several surprising phenomena are established; for instance, the credible interval centered at $\overline{X}$ behaves randomly even if its length is adjusted for long memory and even when $n \to \infty$. So point and interval estimation behave differently.

The article also gives a number of results on how long memory can fundamentally affect the general quality of Bayesian inference. Particularly amazing is the fantastic sample size requirement for accurate inference. A number of examples and other computation illustrate the finite sample case and the theorems. The results pertain to a stationary ARIMA(0,d,0) Gaussian process.

*Research supported by a National Security Agency Grant.

1 Introduction

In statistical analysis, it is frequently assumed that the sample observations are independent, or that the correlation between the observations decays sufficiently rapidly. The first corresponds to the iid theory and the second to common time series models, such as ARMA and Markov processes. In the last fifteen years or so, scientists across a wide variety of disciplines have discovered that in many practical problems, data appear to indicate that the correlation decays much more slowly than these common models entertain, and if undetected, can completely invalidate practically all inferences. And paradoxically, it seems hard to detect it. Processes that exhibit such slowly decaying correlation are now commonly known as long memory processes, and the phenomenon is known as long range dependence. Substantial classical literature on various aspects of inference under long memory has grown in the last decade. See the lucid review article, Beran(1992), and the later publication Beran (1994), and the further references in these two sources. Also see Dahlhaus(1995) , Geweke and Porter-Hudak(1983), Koul(1992), Robinson(1994), and Yajima(1985, 1991), among others.

The corresponding Bayesian studies are lacking, in a broad theoretical sense. The general goal of this article is to present a collection of results on Bayesian decision theory under long range dependence. As in the classical case, a good amount of the results are asymptotic; but a fair amount, especially the examples, deal with the finite sample case. The principal ingredients of our methods and results are the Brown identity and the Brown-Gajek lower bounds for Bayes risk, techniques of Bayesian asymptotics, classic decision theory, and the existing classical literature on long range dependence. There is some use of other tools, but not substantial.

The results in this paper pertain to a stationary Gaussian ARIMA(0,d,0) process; it is one of the most common long memory processes. See Beran (1994). When $d \in (0, \frac{1}{2})$, the process exhibits long range dependence; $d = 0$ corresponds to white noise (independence). For this process, the spectral density equals

$$h(\lambda, d) = \frac{\sigma^2}{2\pi}(2\sin\frac{\lambda}{2})^{-2d}, \qquad \lambda \in (-\pi, \pi) \tag{1.1}$$

and the autocovariance function equals

$$\begin{aligned}
\gamma(k) &= \int_{-\pi}^{\pi} h(\lambda, d) e^{ik\lambda} d\lambda \\
&= (-1)^k \frac{\sigma^2 \Gamma(1-2d)}{\Gamma(1+k-d)\Gamma(1-k-d)}, \\
&\quad k = 0, \pm 1, \cdots, \pm(n-1); \qquad (1.2)
\end{aligned}$$

see Adenstedt(1974).

The topics addressed in this article fall in the following main categories: detection of long memory, the ill effect of long memory on the general quality of inference, point estimation of the stationary mean μ, interval estimation of μ, Bayesian testing about μ, and point estimation of the common variance σ^2. For the results on μ, the variance σ^2 is taken to be known, and set equal to 1. For the results on variance, both μ and σ^2 are assumed to be unknown.

In section 2, we address theoretically the problem of distinguishing between different values of d. This is believed to be notoriously difficult; see the discussions in Beran(1992). We consider the Bayes factor B_n for testing $H_0: \ d = d_0$ vs $H_1: \ d = d_1$ and by using certain martingale inequalities and by deriving the asymptotic distribution of B_n, we show that it is indeed hard to detect long memory or to distinguish between different values of d.

Section 3 gives a number of general results on how long memory fundamentally affects the quality of Bayesian inference. Long memory makes everything more difficult. In particular, it is shown that it always causes an increase in Bayes risk for estimation of μ for every prior and for every given sample size n; that astounding sample sizes are required to achieve the same accuracy that one could get with a small sample under independence; that in small samples where it is basically impossible to detect long memory it produces a seriously misleading Bayes estimate; in testing problems it causes an order of magnitude deflation in the Bayes factor in favor of the null and so can easily lead to erroneous conclusions. In the process, we give an interesting result on the variance of the BLUE, which we needed, but should be of independent interest in the general area.

Section 4 uses the Brown identity to derive the pairwise asymptotic relative efficiency of four estimates of μ:

$$\hat{\mu}_1 = \textit{sample mean} = \overline{X}_n,$$

$$\begin{aligned}
\hat{\mu}_2 &= \textit{posterior mean under assumption of independence of } x_1, ..., x_n, \\
\hat{\mu}_3 &= BLUE, \\
\hat{\mu}_4 &= \textit{posterior mean under correct model.}
\end{aligned} \tag{1.3}$$

Included here are the results that the BLUE is fully efficient, and the estimate $\hat{\mu}_2$ which completely ignores the dependence has an asymptotic efficiency $> .98$ for every d in $(0, \frac{1}{2})$. The latter result is not immediately obvious and actually is a bit of a surprise.

Bayesian interval estimation of μ is considered in Section 5. Three simple intervals are studied; an interval centered at $\overline{X}_n$ with length unadjusted for long memory, an interval centered at $\overline{X}_n$ with length adjusted for long memory, and an interval centered at the BLUE (with length adjusted, again). The posterior probability of the first interval is shown to converge in probability to zero (as expected), that of the third to the nominal specified level, and quite interestingly, the posterior probability of the second interval is shown to converge in law to a random variable! The density function of the limiting distribution is worked out explicitly, and it is seen that it acts essentially like a point mass at the nominal level : a reassuring property.

Point estimation of the variance σ^2 is considered in Section 6. The natural invariant loss $L(a, \theta) = (a - \theta)^2/\theta^2$ is used, where $\theta = \sigma^2$. By using the Brown-Gajek lower bounds on Bayes risk (Brown and Gajek (1990)), it is proved that the everyday estimate s^2 has zero asymptotic efficiency for all d in $(0, \frac{1}{2})$, not just large d. This is a contrast to the corresponding classical literature. The Brown-Gajek bounds also show that the UMVUE of σ^2, however, has asymptotic efficiency 1, a positive result.

The main results are therefore the following :

i. We show theoretically that detection of long memory is difficult; in particular, the asymptotic distribution of the Bayes factor (which is the likelihood ratio statistic as well) is derived;

ii. We give some results on how long memory fundamentally affects Bayesian inference. Long memory always causes an increase in the Bayes risk for every prior and for every sample size n. The Bayes risk obtainable with just 50 observations under independence requires roughly 7.75 billion observations when $d = 0.4$ (amusingly, it does not matter what prior for μ is being used);

iii. We show that asymptotically, various simply computable estimates have very high efficiency, and the BLUE is fully efficient, for estimating μ;

iv. We show that a variety of phenomena can occur for interval estimation of μ. Certain natural intervals can behave randomly even as $n \to \infty$;

v. We show that the variance estimation problem gives a slightly different set of phenomena: s^2 is very bad but the UMVUE is still fully efficient;

vi. We give a fairly good amount of computation and examples to highlight the practically important case of small to moderate n.

We thought it was necessary to formally present proofs of these natural questions. At least three of our results (Theorem 1, 3 and Lemma 3) are equally applicable to frequentist inference. We were also pleased to find these applications of the Brown identity and the Brown-Gajek lower bound which have previously been used principally for admissibility and minimaxity results. There are conditions on the prior densities necessary for the results; they are generally mild and stated in the sections. The very important problem of estimating d will be addressed separately where we will give a new estimate of d stemming from the calculations here.

2 Detection of Long Memory

2.1 Difficulty of Detection

There is ample empirical evidence that detecting long memory and distinguishing between different values of d with commonly available sample sizes are very difficult. See Beran(1992) and the discussions. Consider the problem of testing

$$H_0 : d = d_0 \qquad vs \qquad H_1 : d = d_1 \tag{2.1}$$

where $0 \leq d_0 < d_1 < \frac{1}{2}$.

The use of Bayes factors for testing is common in the Bayesian literature; see Berger(1986, 1996), Kass and Raftery(1995) and Kass and Wasserman(1995). For this case, the Bayes factor is just the likelihood ratio statistic. A large Bayes factor provides evidence in favor of H_0. The Jeffreys scale is often used to interpret the Bayes factor B_n (see Jeffreys (1961), Kass and Raftery (1995)):

B_n	**Strength of Evidence**
1-3	not worth more than a bare mention
3-20	positive
20-150	strong
> 150	very strong

We assume $\mu = EX_t = 0, \sigma^2 = 1$; then

$$B_n = \frac{f_0(\underline{X}_n)}{f_1(\underline{X}_n)} = \frac{|\Sigma_1|^{1/2}}{|\Sigma_0|^{1/2}} e^{\frac{\underline{X}_n'(\Sigma_1^{-1}-\Sigma_0^{-1})\underline{X}_n}{2}} \tag{2.2}$$

where f_i is the density of $\underline{X}_n$ under H_i, $i = 0, 1$, $\Sigma(d) = (\gamma_{k,l})_{k,l=1,\dots,n}$, $\gamma_{k,l} = \gamma(k-l)$ as defined in (1.2), and $\Sigma_i = \Sigma(d_i)$, $i = 0, 1$.

2.1.1 Martingale Bounds

Proposition 1. B_n is a submartingale under H_0.

Proof: $\frac{f_1(\underline{x}_n)}{f_0(\underline{x}_n)}$ is a martingale under H_0 (see Billingsley (1995)), and the reciprocal function is a convex function on R^+. So $B_n = \frac{f_0(\underline{x}_n)}{f_1(\underline{x}_n)}$ is a submartingale under H_0.

Corollary 1. Let $a > 0$. Then

$$P_{H_0}(B_n > a) \leq \frac{(|\Sigma_1^{-1}\Sigma_0| \ \ |2I - \Sigma_1^{-1}\Sigma_0|)^{-1/2}}{a}$$

Proof: By the maximal inequality (see Karlin and Taylor(1975), pp280)

$$P_{H_0}(max_{1\leq k\leq n} B_k > a) \leq \frac{1}{a} E_{H_0}(B_n)$$
$$\Rightarrow \ P_{H_0}(B_n > a) \leq \frac{1}{a} E_{H_0}(B_n) \tag{2.3}$$

By direct calculation,

$$E_{H_0}(B_n) = (|\Sigma_1^{-1}\Sigma_0| \ \ |2I - \Sigma_1^{-1}\Sigma_0|)^{-1/2} \tag{2.4}$$

Substituting (2.4) into (2.3), the corollary follows.

Let us see an application. According to the Jeffreys scale, a strong indication of the null hypothesis will need a Bayes factor of 20 to 150. From the corollary, if $n = 100$, $d_0 = 0$ (independent data), and $d_1 = 0.1$, then

$$\begin{aligned} P(B_n > 20) &\leq 0.175365 \\ P(B_n > 50) &\leq 0.070146 \\ P(B_n > 100) &\leq 0.035073 \\ P(B_n > 150) &\leq 0.023382 \end{aligned}$$

In most applications, one is rather pleased to have 100 observations; the theoretical bounds above show the difficulty of detecting long range dependence with such sample sizes. Table 1 further illustrates the topic by listing $E(B_n)$ under H_0.

Table 1. $E(B_n)$

		n			
d_0	d_1	20	50	100	200
0	0.01	1.00277	1.00747	1.01541	1.03153
0	0.10	1.27459	1.86453	3.5073	12.384
0.10	0.30	2.79031	11.3125	112.645	10857.7
0.25	0.30	1.10394	1.23691	1.483	2.11677
0.40	0.49	3.12733	4.48716	7.8114	22.8132
0.40	0.499	10.1968	15.6311	30.19	108.011
0.49	0.499	2.34005	2.36652	2.395	2.43893

2.1.2 Asymptotic Distribution of B_n

Next, we will discuss the asymptotic distribution of B_n; note that this is also of interest in the classical sense, as B_n is also the likelihood ratio statistic. Now,

$$\begin{aligned} log(B_n) &= -\frac{1}{2}log|\Sigma_1^{-1}\Sigma_0| + \frac{1}{2}\underline{X}_n'(\Sigma_1^{-1} - \Sigma_0^{-1})\underline{X}_n \\ &= -\frac{1}{2}log|\Sigma_1^{-1}\Sigma_0| + \frac{1}{2}Q, \end{aligned} \tag{2.5}$$

where

$$Q = \underline{X}_n'(\Sigma_1^{-1} - \Sigma_0^{-1})\underline{X}_n \tag{2.6}$$

It is easily seen that Q can be written as

$$Q = \Sigma_{i=1}^{n} \lambda_i \chi_i^2 \tag{2.7}$$

where $\{\lambda_i\}$ are the eigenvalues of $\Sigma_1^{-1}\Sigma_0 - I$ and, $\{\chi_i^2\}$ are iid central chisquares with 1 degree of freedom.

And so,

$$EQ = \Sigma_{i=1}^{n} \lambda_i \tag{2.8}$$

$$VarQ = 2\Sigma_{i=1}^{n} \lambda_i^2 \tag{2.9}$$

The exact density of Q is an infinite series of Whittaker's functions (Provost (1988)). Any calculation or computing with it is impossible. The following theorem tells us the asymptotic distribution of Q and hence of $log(B_n)$.

Theorem 1. Under H_0,

$$\frac{Q - EQ}{\sqrt{VarQ}} = \frac{Q - \Sigma_{i=1}^{n} \lambda_i}{\sqrt{2\Sigma_{i=1}^{n} \lambda_i^2}} \xrightarrow{\mathcal{L}} N(0,1) \tag{2.10}$$

To prove this theorem, we need the following preparatory lemmas.

Lemma 1. $|\lambda_i| \leq 1, \quad for \quad all \quad i, \quad i = 1, 2, ..., n.$

Proof: Let

$$g(\lambda) = 2h(\lambda, d_1) - h(\lambda, d_0), \tag{2.11}$$

where $h(\lambda, d)$ is the spectral density of the process itself; see (1.1).

We claim that $g(\lambda)$ is a spectral density. Here is a short proof,

(i)

$$\begin{aligned} g(\lambda) &= \frac{1}{2\pi}[2(2sin\frac{\lambda}{2})^{-2d_1} - (2sin\frac{\lambda}{2})^{-2d_0}] \\ &= \frac{1}{2\pi}(2sin\frac{\lambda}{2})^{-2d_1}[2 - (2sin\frac{\lambda}{2})^{2(d_1-d_0)}] \\ &> 0 \quad (because \quad 0 < 2(d_1 - d_0) < 1) \end{aligned}$$

Obviously,

(ii) $g(\lambda) = g(-\lambda)$,

(iii) $\int_{-\pi}^{\pi} g(\lambda) d\lambda < \infty$.

So $g(\lambda)$ is a spectral density (see pp122, Brockwell and Davis (1991)).
Hence,

$$2\Sigma_1 - \Sigma_0 = ((\int_{-\pi}^{\pi} e^{i\lambda(k-l)} g(\lambda) d\lambda))_{k,l=1,2,\ldots,n}$$

is a covariance matrix and is positive definite.
Consequently,

$$\begin{aligned} & 2\underline{y}'\Sigma_1\underline{y} - \underline{y}'\Sigma_0\underline{y} > 0 && \forall\ \underline{y} \neq \underline{0}, \\ \Rightarrow\ & 2\underline{x}'\underline{x} - \underline{x}'\Sigma_1^{-1/2}\Sigma_0\Sigma_1^{-1/2}\underline{x} > 0 && \forall\ \underline{x} \neq \underline{0}, \\ \Rightarrow\ & \frac{\underline{x}'\Sigma_1^{-1/2}\Sigma_0\Sigma_1^{-1/2}\underline{x}}{\underline{x}'\underline{x}} < 2 && \forall\ \underline{x} \neq \underline{0}. \end{aligned} \tag{2.12}$$

Now let ξ_i be the eigenvalues of $\Sigma_1^{-1}\Sigma_0$. From (2.12),

$$\begin{aligned} & 0 \leq \xi_i < 2, \quad for \quad all \quad i, \quad i = 1, \ldots, n, \\ \Rightarrow\ & |\lambda_i| \leq 1, \quad for \quad all \quad i, \quad i = 1, \ldots, n. \end{aligned}$$

Lemma 2.

$$(1) \qquad \lim_{n\to\infty} \frac{1}{n} tr(\Sigma_1^{-1}\Sigma_0 - I) = \frac{1}{2\pi}\int_{-\pi}^{\pi} [(2sin\frac{\lambda}{2})^{2(d_1-d_0)} - 1] d\lambda,$$

$$(2) \qquad \lim_{n\to\infty} \frac{1}{n} tr(\Sigma_1^{-1}\Sigma_0 - I)^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} [(2sin\frac{\lambda}{2})^{2(d_1-d_0)} - 1]^2 d\lambda.$$

This lemma can be derived from Theorem 5.1 in Dahlhaus(1989). From this lemma, one has that $\Sigma_{i=1}^n \lambda_i \sim n$ and $\Sigma_{i=1}^n \lambda_i^2 \sim n$.

Proof of Theorem 1: From the representation (2.7), the moment generating function of $\frac{Q-EQ}{\sqrt{VarQ}}$ is (for $t \in (-\frac{1}{2}, \frac{1}{2})$)

$$M(t) = e^{-\frac{t\Sigma_{i=1}^n \lambda_i}{\sqrt{2\Sigma_{i=1}^n \lambda_i^2}}} \prod_{i=1}^n (1 - \frac{2\lambda_i t}{\sqrt{2\Sigma_{i=1}^n \lambda_i^2}})^{-1/2} \tag{2.13}$$

$$\Rightarrow\ logM(t) = -\frac{t\Sigma_{i=1}^n \lambda_i}{\sqrt{2\Sigma_{i=1}^n \lambda_i^2}} - \frac{1}{2}\Sigma_{i=1}^n log(1 - \frac{2\lambda_i t}{\sqrt{2\Sigma_{i=1}^n \lambda_i^2}}) \tag{2.14}$$

By a Taylor expansion,

$$\begin{aligned} & log(1 - \frac{2\lambda_i t}{\sqrt{2\Sigma_{i=1}^n \lambda_i^2}}) \\ = & -\frac{2\lambda_i t}{\sqrt{2\Sigma_{i=1}^n \lambda_i^2}} - \frac{(2\lambda_i t)^2}{2(2\Sigma_{i=1}^n \lambda_i^2)} + R_i, \end{aligned} \tag{2.15}$$

where

$$\begin{aligned}
|R_i| &\leq \Sigma_{k=3}^{\infty}\left|\frac{(2\lambda_i t)^k}{k(2\Sigma_{i=1}^n \lambda_i^2)^{k/2}}\right| \\
&\leq \Sigma_{k=3}^{\infty}\frac{|2t|^k}{(2\Sigma_{i=1}^n \lambda_i^2)^{k/2}} \quad (Lemma\ \ 1) \\
&= \frac{|2t|^3/(2\Sigma_{i=1}^n \lambda_i^2)^{3/2}}{1-|2t|/(2\Sigma_{i=1}^n \lambda_i^2)^{1/2}}. \qquad (2.16) \\
&\quad (\ \ Sum\ \ of\ \ a\ \ geometric\ \ progression\ \ )
\end{aligned}$$

Hence,

$$\begin{aligned}
|\Sigma_{i=1}^n R_i| &\leq \Sigma_{i=1}^n |R_i| \\
&\leq \frac{n|2t|^3/(2\Sigma_{i=1}^n \lambda_i^2)^{3/2}}{1-|2t|/(2\Sigma_{i=1}^n \lambda_i^2)^{1/2}} \\
&\rightarrow 0, \quad as\ \ n \rightarrow \infty \ \ (Lemma\ \ 2).
\end{aligned}$$

From (2.14), therefore,

$$\begin{aligned}
logM(t) &= -\frac{t\Sigma_{i=1}^n \lambda_i}{\sqrt{2\Sigma_{i=1}^n \lambda_i^2}} - \frac{1}{2}\Sigma_{i=1}^n[-\frac{2\lambda_i t}{\sqrt{2\Sigma_{i=1}^n \lambda_i^2}} - \frac{(2\lambda_i t)^2}{2(2\Sigma_{i=1}^n \lambda_i^2)} + R_i] \\
&= \frac{t^2}{2} - \frac{1}{2}\Sigma_{i=1}^n R_i \\
&\rightarrow \frac{t^2}{2}, \quad as\ \ n \rightarrow \infty.
\end{aligned}$$

The theorem follows.

We still have a bit of work to do for the asymptotic distribution of $logB_n$.

Corollary 2. Under H_0,

$$\frac{log(B_n) - m}{v} \xrightarrow{\mathcal{L}} N(0,1)$$

where

$$m = \frac{n}{4\pi}\int_{-\pi}^{\pi}\{log[(2sin\frac{\lambda}{2})^{2(d_0-d_1)}] + (2sin\frac{\lambda}{2})^{2(d_1-d_0)} - 1\}d\lambda, \qquad (2.17)$$

$$v^2 = \frac{n}{4\pi}\int_{-\pi}^{\pi}[(2sin\frac{\lambda}{2})^{2(d_1-d_0)} - 1]^2 d\lambda. \qquad (2.18)$$

Proof: From (2.5),

$$log(B_n) = \frac{1}{2}log\frac{|\Sigma_1|}{|\Sigma_0|} + \frac{1}{2}Q. \qquad (2.19)$$

By (5.38) in Beran(1994)

$$\lim_{n\to\infty} \frac{1}{n} log \frac{|\Sigma_1|}{|\Sigma_0|} = \frac{1}{2\pi} \int_{-\pi}^{\pi} log[(2sin\frac{\lambda}{2})^{2(d_0-d_1)}]d\lambda. \tag{2.20}$$

Then, by (2.20), Theorem 1 and Lemma 2, the corollary follows.

One can therefore use the lognormal(m, v) density

$$f_{B_n}(x) = \frac{1}{\sqrt{2\pi}vx} e^{-\frac{(logx-m)^2}{2v^2}}, \tag{2.21}$$

to approximate tail probabilities for the Bayes factor B_n. As discussed before, we believe that using the exact density of B_n would not even be possible. Table 2 lists these approximations for $P(B_n \geq a)$ using (2.21). For instance, if $n = 100$, $d_0 = 0$, and $d_1 = 0.1$, $P(B_n \geq 20) \approx 0.018$, very small.

Table 2. Approximations to $P(B_n \geq a)$

$d_0 = 0, \quad d_1 = 0.1$

	n				
a	20	50	100	200	650
3	0.0245915	0.168308	0.364775	0.589975	0.903923
6	0.000341438	0.031099	0.162323	0.411175	0.853945
10	4.29842×10^{-6}	0.0056852	0.0726901	0.288547	0.807502
20	2.0623×10^{-9}	0.000296154	0.0180733	0.156334	0.731713
150	5.39602×10^{-24}	6.68171×10^{-10}	0.0000385165	0.0100744	0.455874

$d_0 = 0.4, \quad d_1 = 0.49$

	n				
a	20	50	100	200	650
3	0.0133153	0.125563	0.304465	0.524858	0.862851
6	0.0000769	0.0162069	0.112603	0.332335	0.793375
10	3.899×10^{-7}	0.00205183	0.041854	0.21218	0.730886
20	3.80093×10^{-11}	0.0000562934	0.00753345	0.097711	0.633246
150	9.16093×10^{-29}	7.68586×10^{-12}	3.92776×10^{-6}	0.00310943	0.323105

From all the results in this section, we get the common conclusion that detection (and discrimination) of long range dependence with sample sizes we commonly have is very very difficult. Especially difficult is distinguishing between two large values of d.

3 How Long Memory Affects Inference Itself

In this section, we give some basic results, examples and illustrative computation to explain how presence of long range dependence fundamentally affects inference on the mean μ. First we give a fairly surprising result, true for every sample size n. The result says that the presence of long range dependence makes estimation of μ fundamentally harder.

3.1 Effect on Variance of the BLUE and Bayes Risk

Theorem 2. Let $r(\pi) = r(\pi, n, d)$ denote Bayes risk under a prior π for estimation of μ under squared error loss. Then, for every n, $r(\pi)$ is increasing in the long memory parameter d.

The proof requires use of the following two lemmas. Of these, the first lemma should be of interest in nonBayesian inference as well.

Lemma 3. For any fixed n, The variance of the $BLUE$ of μ

$$V(n,d) = \frac{1}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}} = \frac{Beta(n, 1-2d)}{Beta(1-d, 1-d)} \tag{3.1}$$

is an increasing function of d.

Proof: The fundamental formula (3.1) is available in many articles, see, e.g. Samarov & Taqqu(1988). We prove by induction that, $V'(n,d)$, the derivative of $V(n,d)$ with respect to d, is always positive.

First we prove $V'(1,d) > 0$.

Note,

$$\begin{aligned} V(1,d) &= \frac{\Gamma(1-2d)}{\Gamma^2(1-d)} \\ &= \int_{-\pi}^{\pi} \frac{1}{2\pi}(2sin\frac{\lambda}{2})^{-2d} d\lambda \\ &= \int_{0}^{\pi} \frac{1}{\pi}(2sin\frac{\lambda}{2})^{-2d} d\lambda \end{aligned}$$

Hence,

$$V'(1,d) = -2\int_0^{\pi} \frac{1}{\pi} log(2sin\frac{\lambda}{2})\ (2sin\frac{\lambda}{2})^{-2d} d\lambda \tag{3.2}$$

and

$$V''(1,d) = 4\int_0^\pi \frac{1}{\pi}(log(2sin\frac{\lambda}{2}))^2\ (2sin\frac{\lambda}{2})^{-2d}d\lambda. \tag{3.3}$$

From (3.3), $V''(1,d) > 0$; on the other hand, from (3.2),

$$V'(1,0) = -\frac{2}{\pi}\int_0^\pi log(2sin\frac{\lambda}{2})d\lambda = 0.$$

$$(By\ exact\ integration\ on\ Mathematica).$$

Thus, $V'(1,d) > 0$ for all $0 < d < \frac{1}{2}$.

Now we show that $V'(n+1,d) > 0$ if $V'(n,d) > 0$. That will complete the proof.

By straightforward integration by parts,

$$\begin{aligned} V(n+1,d) &= \frac{n}{1-2d}(V(n,d) - V(n+1,d)) \\ \Rightarrow\ \ (1-2d+n)V(n+1,d) &= nV(n,d). \end{aligned} \tag{3.4}$$

By taking derivative with respect to d on both sides of (3.4), one gets

$$\begin{aligned} &-2V(n+1,d) + (1-2d+n)V'(n+1,d) = nV'(n,d) \\ \Rightarrow\ \ &V'(n+1,d) = \frac{1}{1-2d+n}[2V(n+1,d) + nV'(n,d)] \\ &> 0\ \ (by\ the\ induction\ hypothesis). \end{aligned}$$

This proves Lemma 3.

Now we state a second lemma which is of some basic interest in Bayesian statistics. The proof of this lemma is available in DasGupta(1997) and will not be duplicated here. In addition, the only proof we are aware of uses a completely different set of tools and will be out of place in this article.

Lemma 4. Consider the canonical problem of $Y \sim N(\mu, t)$ and estimation of μ under squared error loss. Let $\delta(Y) = \delta(Y,t)$ denote the Bayes estimate of μ under a fixed prior π and $r(\pi)$ the corresponding Bayes risk. Then

$$\frac{d}{dt}r(\pi) = E_m(\delta'(Y))^2, \tag{3.5}$$

where δ' denotes the derivative of δ with respect to Y and E_m denotes marginal expectation. In particular, the Bayes risk $r(\pi)$ is an increasing function of t.

Proof: See DasGupta(1997).

Lemma 3 and Lemma 4 lead to Theorem 2.

Proof of Theorem 2: The sufficient statistic $Y = \frac{\underline{1}'\Sigma^{-1}\underline{X}}{\underline{1}'\Sigma^{-1}\underline{1}}$ has the $N(\mu, t)$ distribution with $t = \frac{1}{\underline{1}'\Sigma^{-1}\underline{1}}$. By lemma 4, $r(\pi, n, d)$ increases with t and by lemma 3, t increases with d, for each fixed n. So the theorem follows.

3.2 Accuracy: The Fantastic Effect on Sample Size

Since the result just proved says that presence of long memory causes an increase in Bayes risk and so a loss of accuracy compared to independent data, if the same sample size is used, it is statistically interesting to ask how much larger the sample size needs to be to achieve the same Bayes risk as that under a given n_0 when the observations are independent. We call this the **Equivalent Sample Size** (not to be confused with an unrelated concept by the same name in Bayesian statistics).

Theorem 3. Let $n_0 \geq 1$ be fixed. For any given $d > 0$, let $n(\pi, d)$ denote the sample size under which the Bayes risk $r(\pi)$ equals the Bayes risk obtainable with n_0 observations under independence. Then, $n(\pi, d)$ depends only on d, but not on the prior π, and as $d \to \frac{1}{2}$, it is of the exact order

$$n(\pi, d) \sim 4e^{-\gamma}[\frac{n_0}{\pi(1-2d)}]^{\frac{1}{1-2d}}, \tag{3.6}$$

where $\gamma(= .577215665)$ is the Euler constant.

Remark 1. $n(\pi, d)$ also has the following interpretation: it is the sample size necessary for the variance of the BLUE of μ to equal the variance obtainable with n_0 observations under independence.

Remark 2. From the expression in (3.6), it is clear how astounding the sample size requirement is. As $d \to \frac{1}{2}$, $n(\pi, d)$ grows faster than any exponential rate; we may call this **Super Exponential Growth.**

Proof: *Step 1.* The sufficient statistic $Y = \frac{\underline{1}'\Sigma^{-1}\underline{X}}{\underline{1}'\Sigma^{-1}\underline{1}}$ has the $N(\mu, t)$ distribution with $t = \frac{1}{\underline{1}'\Sigma^{-1}\underline{1}}$. For n_0 independent observations, t becomes $\frac{1}{n_0}$. It will follow from Lemma 4 that the Bayes risks are equal iff

$$\underline{1}'\Sigma^{-1}\underline{1} = n_0 \tag{3.7}$$

This already means $n(\pi, d)$ only depends on d.

Step 2. For convenience, let $1-2d=x, n=n(\pi,d)$. Obviously, $n \to \infty$ as $x \to 0$(i.e. as $d \to \frac{1}{2}$).

With this notation, equation (3.7) becomes

$$\frac{Beta(\frac{1+x}{2},\frac{1+x}{2})}{Beta(n,x)} = n_0,$$

i.e.

$$\frac{\Gamma(n+x)}{\Gamma(n)\Gamma(x)} = \frac{n_0}{Beta(\frac{1+x}{2},\frac{1+x}{2})} \tag{3.8}$$

$$\Longleftrightarrow \quad \frac{\Gamma(n+x)}{\Gamma(n)} = \frac{n_0\,\Gamma(x)}{Beta(\frac{1+x}{2},\frac{1+x}{2})} = h(x) \quad (say) \tag{3.9}$$

Step 3. Use now the facts (see Lebedev (1965), pp13):

$$\Gamma(n) = n^{n-1/2}e^{-n+\epsilon_1}\sqrt{2\pi}$$

and

$$\Gamma(n+x) = (n+x)^{n+x-1/2}e^{-(n+x)+\epsilon_2}\sqrt{2\pi}$$

where

$$|\epsilon_1| \le \frac{1}{12n}, \quad |\epsilon_2| \le \frac{1}{12(n+x)}. \tag{3.10}$$

Then (3.9) becomes

$$\frac{(n+x)^{n+x-1/2}e^{-(n+x)+\epsilon_2}}{n^{n-1/2}e^{-n+\epsilon_1}} = h(x).$$

Taking logarithm on both sides,

$$(n+x-\frac{1}{2})log(n+x) - (n-\frac{1}{2})logn - x + \epsilon_2 - \epsilon_1 = logh(x)$$

$$\Rightarrow \quad nlog(1+\frac{x}{n}) + xlogn + xlog(1+\frac{x}{n}) - \frac{1}{2}log(1+\frac{x}{n}) - x + \epsilon_2 - \epsilon_1 = logh(x) \tag{3.11}$$

Step 4. From (3.11),

$$\begin{aligned} logh(x) &= n[\frac{x}{n}+o(\frac{x}{n})] + xlogn + x[\frac{x}{n}+o(\frac{x}{n})] - \frac{1}{2}[\frac{x}{n}+o(\frac{x}{n})] - x + \epsilon_2 - \epsilon_1 \\ &= xlogn + o(x) + \epsilon_2 - \epsilon_1 \\ &\Rightarrow \quad logn = \frac{logh(x)}{x} - o(1) + \frac{\epsilon_1-\epsilon_2}{x} \\ &\Rightarrow \quad n = [h(x)]^{\frac{1}{x}}e^{-o(1)}e^{\frac{\epsilon_1-\epsilon_2}{x}}. \end{aligned} \tag{3.12}$$

Step 5. We now claim

$$\frac{\epsilon_1 - \epsilon_2}{x} \to 0, \qquad as \;\; x \to 0. \tag{3.13}$$

By (3.10)

$$|\epsilon_1 - \epsilon_2| \le \frac{1}{12n} + \frac{1}{12(n+x)},$$
$$\Rightarrow \;\; |\frac{\epsilon_1 - \epsilon_2}{x}| \le \frac{1}{12nx} + \frac{1}{12(n+x)x}.$$

If we can prove

$$\lim_{x \to 0}(nx) = \infty, \tag{3.14}$$

then (3.13) follows. So we will now prove (3.14).

Step 6. By (3.8)

$$\frac{(n-1+x)(n-2+x)\cdots(1+x)x\Gamma(x)}{(n-1)(n-2)\cdots 2 \cdot 1 \cdot \Gamma(x)} = \frac{n_0}{Beta(\frac{1+x}{2}, \frac{1+x}{2})}. \tag{3.15}$$

On a bit of algebra, (3.15) will give

$$\begin{aligned}
(n-1)x &= \frac{n_0}{Beta(\frac{1+x}{2}, \frac{1+x}{2})} \prod_{i=1}^{n-1} \frac{i}{i-1+x} - x^2 \\
&= \frac{n_0}{Beta(\frac{1+x}{2}, \frac{1+x}{2})} \prod_{i=1}^{n-1} (1 + \frac{1-x}{i-1+x}) - x^2 \\
&\ge \frac{n_0}{Beta(\frac{1+x}{2}, \frac{1+x}{2})} (1 + \Sigma_{i=1}^{n-1} \frac{1-x}{i-1+x}) - x^2 \\
&\ge \frac{n_0}{Beta(\frac{1+x}{2}, \frac{1+x}{2})} [1 + (1-x)\Sigma_{i=1}^{n-1} \frac{1}{i}] - x^2 \\
&\to \infty, \qquad as \;\; x \to 0,
\end{aligned}$$

and so $nx \to \infty$ as $x \to 0$.

Step 7. Therefore, by combining (3.12) and (3.13), we have $n \sim [h(x)]^{\frac{1}{x}}$.

Step 8. In the final step, we reduce $[h(x)]^{\frac{1}{x}}$ to the form (3.6) stated in the theorem.

By using the definition (3.9) of $h(x)$,

$$h(x) = \frac{n_0 \; x \; \Gamma^2(x)}{[\Gamma(\frac{1}{2} + \frac{x}{2})]^2}. \tag{3.16}$$

By the Duplication formula for gamma functions (see Lebedev(1965), pp4),

$$\Gamma(\frac{1}{2}+\frac{x}{2}) = \frac{\Gamma(x)\sqrt{\pi}\ 2^{1-x}}{\Gamma(\frac{x}{2})}$$

$$\Rightarrow \quad \frac{\Gamma(x)}{\Gamma(\frac{1}{2}+\frac{x}{2})} = \frac{\Gamma(\frac{x}{2})2^{x-1}}{\sqrt{\pi}}.$$

Hence, from (3.16),

$$[h(x)]^{\frac{1}{x}} = [\frac{n_0\ x\ 2^{2x-2}}{\pi}\Gamma^2(\frac{x}{2})]^{\frac{1}{x}} \tag{3.17}$$

In (3.17), write $\Gamma(\frac{x}{2}) = \frac{2}{x}\Gamma(1+\frac{x}{2})$. This will mean:

$$\begin{aligned} [h(x)]^{\frac{1}{x}} &= [\frac{n_0\ x\ 2^{2x-2}\ 4}{\pi x^2}\Gamma^2(1+\frac{x}{2})]^{\frac{1}{x}} \\ &= 4(\frac{n_0}{\pi x})^{\frac{1}{x}}[\Gamma(1+\frac{x}{2})]^{\frac{2}{x}} \end{aligned} \tag{3.18}$$

On Using a Taylor expansion for $\Gamma(1+\frac{x}{2})$ arround $x = 0$ and the fact that $\Gamma'(1)$ equals $-\gamma$ (the negative of the Euler constant), (3.18) will finally yield

$$\begin{aligned} [h(x)]^{\frac{1}{x}} &\sim 4(\frac{n_0}{\pi x})^{\frac{1}{x}}(1-\frac{\gamma}{2}x)^{\frac{2}{x}} \\ &\sim 4e^{-\gamma}(\frac{n_0}{\pi x})^{\frac{1}{x}}, \end{aligned}$$

as asserted in the statement of the theorem.

The following table gives some numerical values; these are exact, not coming from Theorem 3. The numbers are amazing; to get the same accuracy as one would get with 50 observations under independence, one will need 7.75 billion observations if $d = 0.4$!!

Table 3. Equivalent Sample Size

	n			
d	10	30	50	100
.10	17	66	125	297
.25	110	986	2736	10943
.40	2.48×10^6	6.03×10^8	7.75×10^9	2.48×10^{11}

3.3 Severe Bias: An Example

Although a large number of our subsequent results will show that asymptotically, for every fixed d, simple estimates like $\overline{X}$ perform well, the case of practical sample sizes is important.

The following little table is illustrative of how a short observed series can give a totally misleading Bayes estimate for μ.

Example 1. Consider estimation of μ by using the popular Double Exponential prior with density $\frac{1}{2}e^{-|\mu|}$. The following table gives the correct Bayes estimate under $d = 0$, 0.25, and 0.49, for $n = 30$. At such sample sizes, detection of long memory is extremely difficult, and yet the table below shows how seriously the estimate is affected if one unknowingly assumes independence ($d = 0$). In the following, y denotes the $BLUE$.

Table 4. Posterior Mean with Double Exponential Prior

n	d	y						
		0	0.5	1.0	1.5	2.0	2.5	3.0
30	0	0	0.47	0.97	1.47	1.97	2.47	2.97
	0.25	0	0.37	0.82	1.31	1.81	2.31	2.81
	0.49	0	0.05	0.10	0.15	0.21	0.26	0.31

3.4 Effect On Bayes Factors

In this section we give a result that says that presence of long range dependence will cause an order of magnitude deflation in the Bayes factor in favor of a point null hypothesis on the stationary mean μ. Thus , if the long memory is undetected, a Bayesian will see a much smaller Bayes factor in favor of the null than he/she is used to and can easily be deceived into an erroneous conclusion. Theorem 4 explains it; the rate of growth of the Bayes factor slows down with d.

The following table illustrates the point. In this table, the expected Bayes factor under the null:$\mu = 0$, is given for selected n and d; the prior on the alternative is $N(0,1)$.

Table 5. Average Bayes Factor In Favor of Null

d	n			
	30	100	300	1000
0	3.97	7.12	12.28	22.38
.10	2.96	4.66	7.16	11.53
.25	1.84	2.35	3.01	3.98
.40	1.16	1.22	1.28	1.36

Theorem 4. For the ARIMA(0, d, 0) process, let $H_0 : \mu = 0$ and $H_1 : \ \mu \neq 0$. Suppose, conditionally on H_1 being true, μ has a prior density $\pi(\mu)$, which can be defined continuously

at $\mu = 0$ and that this value $\pi(0) > 0$. Let B_d be the Bayes factor in favor of H_0. Then, as $n \longrightarrow \infty$,

$$-2log\frac{\sqrt{2\pi}\pi(0)B_d}{\sqrt{c(d)}n^{1/2-d}} \xrightarrow{\mathcal{L}} \chi^2(1)$$

under H_0 for every d in $[0,\frac{1}{2})$; in particular, the norming is different for $d = 0$ and $d > 0$, In the above,

$$c(d) = \frac{\Gamma^2(1-d)}{\Gamma(1-2d)\Gamma(2-2d)}$$

Proof: The argument is standard and so we only briefly describe it. B_d is defined as

$$B_d = \frac{\sqrt{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{2}y^2}}{\sqrt{2\pi}m(y)} \tag{3.19}$$

where y denotes the BLUE and $m(y)$ denotes the marginal density. By virtue of the continuity and strict positivity of the prior density at 0, $m(y) \xrightarrow{P} \pi(0)$ under H_0. The stated result will therefore follow from the following:

$$\begin{aligned}
& B_d = \frac{\frac{\sqrt{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{2}y^2}}{\sqrt{2\pi}}}{\pi(0)}\frac{\pi(0)}{m(y)} \\
\Rightarrow \quad & logB_d + log\pi(0) = \frac{1}{2}log(\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}) - \frac{1}{2}log(2\pi) - \frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{2}y^2 + o_p(1) \\
\Rightarrow \quad & logB_d + log\pi(0) + \frac{1}{2}log(2\pi) - \frac{1}{2}logc(d) - \frac{1}{2}logn^{1-2d} \\
& = -\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{2}y^2 + o_p(1), \\
& (because\ \ \underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}} \sim c(d)n^{1-2d})
\end{aligned} \tag{3.20}$$

from which the result follows on a little algebra.

4 Point Estimation of μ

4.1 A Positive Result: Asymptotic Unanimity

The first result is a prelude to the asymptotic efficiency results of the subsequent sections and shows that there is asymptotic unanimity in a fairly strong sense among the four estimates listed in (1.3). We have the following result.

Theorem 5. Let $\hat{\mu}_i,\ \ 1 \leq i \leq 4$ be the four estimates presented in (1.3). Suppose the prior density $\pi(\mu)$ is symmetric and unimodal and $|\frac{\pi'(\mu)}{\pi(\mu)}| \leq a|\mu| + b$ for some $a,\ b \geq 0$, and $\int |\mu|\pi(\mu)d\mu < \infty$. Then $\hat{\mu}_i - \hat{\mu}_j$ converges in L_1 to 0 under the marginal as well as under each fixed μ, for every pair $i,\ j$. In particular, $\hat{\mu}_i - \hat{\mu}_j$ converges to 0 in probability.

Remark. Actually a stronger result can be proved giving the rate of convergence as well. Also note that the conjugate normal and the double exponential prior each satisfies the conditions on $\pi(\mu)$.

Proof: We will only show $E_\mu|\hat{\mu}_1 - \hat{\mu}_3| \to 0$ and $E_\mu|\hat{\mu}_2 - \hat{\mu}_4| \to 0$. The others are similar. Since $\hat{\mu}_1 - \hat{\mu}_3|\mu \sim N(0, \frac{\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}}{n^2} - \frac{1}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}})$, clearly $E_\mu|\hat{\mu}_1 - \hat{\mu}_3| = \sqrt{\frac{2}{\pi}}\sqrt{\frac{\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}}{n^2} - \frac{1}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}}$ $\to 0$, as $\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}} = O(n^{1+2d})$ and $\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}} = O(n^{1-2d})$.

Next,

$$\begin{aligned}
\hat{\mu}_2 &= \int \mu \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n(\mu-\overline{x})^2}{2}} \pi(\mu)d\mu \Big/ \int \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n(\mu-\overline{x})^2}{2}} \pi(\mu)d\mu \\
&= \overline{x} + \frac{1}{n} \frac{\int e^{-\frac{n(\mu-\overline{x})^2}{2}} \pi'(\mu)d\mu}{\int e^{-\frac{n(\mu-\overline{x})^2}{2}} \pi(\mu)d\mu} \qquad (4.1)
\end{aligned}$$

(integration by parts; see Brown(1986)).

Similarly,

$$\hat{\mu}_4 = y + \frac{1}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}} \frac{\int e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}(\mu-y)^2}{2}} \pi'(\mu)d\mu}{\int e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}(\mu-y)^2}{2}} \pi(\mu)d\mu}, \qquad (4.2)$$

where as before y denotes the $BLUE$.

Hence,

$$\begin{aligned}
\hat{\mu}_2 - \hat{\mu}_4 &= \overline{x} - y + \frac{1}{n} \frac{\int e^{-\frac{n(\mu-\overline{x})^2}{2}} \pi'(\mu)d\mu}{\int e^{-\frac{n(\mu-\overline{x})^2}{2}} \pi(\mu)d\mu} - \frac{1}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}} \frac{\int e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}(\mu-y)^2}{2}} \pi'(\mu)d\mu}{\int e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}(\mu-y)^2}{2}} \pi(\mu)d\mu} \\
&= \hat{\mu}_1 - \hat{\mu}_3 + R_1 + R_2 \ \ (say) \qquad (4.3)
\end{aligned}$$

We have already seen that $E_\mu|\hat{\mu}_1 - \hat{\mu}_3| \to 0$. If we can prove the expectations of $R_1,\ R_2$ go to 0, the proof will be complete.

Now,

$$\begin{aligned}
|R_1| &\leq \frac{1}{n} \int e^{-\frac{n(\mu-\overline{x})^2}{2}} |\pi'(\mu)| d\mu \Big/ \int e^{-\frac{n(\mu-\overline{x})^2}{2}} \pi(\mu)d\mu \\
&\leq \frac{1}{n} \int e^{-\frac{n(\mu-\overline{x})^2}{2}} (a|\mu| + b)\pi(\mu)d\mu \Big/ \int e^{-\frac{n(\mu-\overline{x})^2}{2}} \pi(\mu)d\mu
\end{aligned}$$

$$
\begin{aligned}
&= \frac{b}{n} + \frac{a}{n}\frac{E[|\mu|\pi(\mu)|\mu \sim N(\overline{x}, \frac{1}{n})]}{E[\pi(\mu)|\mu \sim N(\overline{x}, \frac{1}{n})]} \\
&\leq \frac{b}{n} + \frac{a}{n}\frac{E[|\mu||\mu \sim N(\overline{x}, \frac{1}{n})] \;\; E[\pi(\mu)|\mu \sim N(\overline{x}, \frac{1}{n})]}{E[\pi(\mu)|\mu \sim N(\overline{x}, \frac{1}{n})]} \\
&\quad (\; \textit{by assumption } \pi(\mu) \textit{ is decreasing and } |\mu| \textit{ is} \\
&\quad \textit{increasing in } |\mu|, \textit{ and so their covariance is} \leq 0) \\
&= \frac{b}{n} + \frac{a}{n}E[|\mu||\mu \sim N(\overline{x}, \frac{1}{n})] \\
&\leq \frac{b}{n} + \frac{a}{n}(|\overline{x}| + \sqrt{\frac{2}{\pi n}}) \\
&= \frac{1}{n}(b + a|\overline{x}| + \sqrt{\frac{2}{\pi n}}) \qquad (4.4)
\end{aligned}
$$

From (4.4), it quickly follows that $E_\mu|R_1| \to 0$ by simply using the fact $\underline{1}'\Sigma\underline{1} = O(n^{1+2d})$.

Similarly, $E_\mu|R_2| \to 0$, and thus $E_\mu|\hat{\mu}_2 - \hat{\mu}_4| \to 0$. Convergence in probability follows as a consequence.

4.2 Brown Identity

In this section, the asymptotic efficiency of the estimates $\hat{\mu}_i,\ i = 1, 2, 3$ with respect to the exact Bayes estimate will be investigated. Naturally, the criterion for efficiency is Bayes risk. In our view, the most important result is that the asymptotic efficiency of $\hat{\mu}_2$, the estimate that completely ignores the dependence structure, is close to 99% for every d. This is not apriori evident. The main technical tool is the Brown Identity for Bayes risk, which we describe next.

Suppose X is distributed according to a location parameter density $f(x|\mu) = f(x - \mu)$ and μ has prior density $\pi(\mu) > 0$. Let $m(x)$ be the marginal density of X:

$$m(x) = \int_{-\infty}^{\infty} f(x - \mu)\pi(\mu)d\mu = \int_{-\infty}^{\infty} f(t)\pi(x - t)dt. \qquad (4.5)$$

For an absolutely continuous function g, we define

$$I(g) = \int_{-\infty}^{\infty} \frac{(g'(x))^2}{g(x)}dx, \qquad (4.6)$$

the familiar Fisher Information function. See Bickel(1981) and Huber(1981).

Lemma 5. If $\pi(\mu)$ is absolutely continuous and $m(x)$ can be differentiated under the integral sign, then

$$I(m) \leq I(\pi).$$

Suppose now X is a normal random variable, with mean μ and variance σ^2. $\hat{\mu}$ is the posterior mean of μ, $r(\pi) = E^{m(x)}E^{\mu|x}(\hat{\mu} - \mu)^2$ is the Bayes risk. Then $I(m)$ and $r(\pi)$ have the following interesting relationship.

Theorem 6. If $\pi(\mu)$ is as in Lemma 5, then

$$r(\pi) = \sigma^2 - \sigma^4 I(m).$$

This is implicit in Brown(1971); we call it the **Brown Identity**.

So one immediately gets the Bayes risk lower bound:

Corollary 3. Under the setup of Theorem 6,

$$r(\pi) \geq \sigma^2 - \sigma^4 I(\pi).$$

This will be used below.

4.3 Asymptotic Efficiency

For the stationary Gaussian ARIMA(0,d,0) process, $X_t, t \in Z$, as defined in section 1, we have the following theorem. Following Theorem 7 and its Corollary, we will have a discussion of the implications of these results. $e(i,j)$ below means asymptotic efficiency of $\hat{\mu}_i$ with respect to $\hat{\mu}_j$. Also, $\tau(d)$ will denote the function

$$\tau(d) = \frac{(1+2d)\Gamma(1+d)\Gamma(2-2d)}{\Gamma(1-d)}. \tag{4.7}$$

Theorem 7. If the conditions of Theorem 6 are satisfied, then in general

(a) $e(3,4) = 1$,

(b) If, furthermore, $\pi(\mu)$ is symmetric and unimodal and $|\frac{\pi'(\mu)}{\pi(\mu)}| \leq a|\mu| + b$ for some $a,\ b \geq 0$ and $\int |\mu|^2 \pi(\mu) d\mu < \infty$, then

$$e(1,2) = 1.$$

Corollary 4.

(a) Without any condition on the prior,

$$e(1,3) = \tau(d) \tag{4.8}$$

(b) If $\pi(\mu)$ is as in Lemma 5,

$$e(1,4) = \tau(d) \tag{4.9}$$

(c) If furthermore the conditions in part (b) of Theorem 6 hold, then

$$e(2,4) = e(2,3) = \tau(d) \tag{4.10}$$

Proof of Corollary 4:

(a)

$$r(\hat{\mu}_1) = Var(\overline{X}) = \frac{\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}}{n^2}$$

$$r(\hat{\mu}_3) = Var(BLUE) = \frac{1}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}$$

Now, by using Theorem 8.1 of Beran (1994) (see also Adenstedt 1974, Samarov and Taqqu 1988)

$$e(1,3) = \tau(d)$$

Part (b) and (c) follow from Theorem 7.

Proof of Theorem 7:

(a) Since $Y = [\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}]^{-1}\underline{\mathbf{1}}'\Sigma^{-1}\underline{X}$ is a sufficient statistic for μ, distributed as $N(\mu, [\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}]^1)$, by Corollary 3, we have

$$r(\hat{\mu}_4) = r(\pi) \geq \frac{1}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}} - \frac{I(\pi)}{(\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}})^2}.$$

Hence,

$$1 \geq e(3,4) = \lim_{n\to\infty} \frac{r(\hat{\mu}_4)}{r(\hat{\mu}_3)} \geq \lim_{n\to\infty}(1 - \frac{I(\pi)}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}) = 1.$$

(b) The proof closely follows the steps of Theorem 5 and so we will limit the details. First,

$$\hat{\mu}_2 = \overline{x} + g_n(\overline{x}),$$

where

$$g_n(\overline{x}) = \frac{1}{n}\frac{\int e^{-\frac{n(\mu-\overline{x})^2}{2}}\pi'(\mu)d\mu}{\int e^{-\frac{n(\mu-\overline{x})^2}{2}}\pi(\mu)d\mu}. \tag{4.11}$$

Hence,

$$\begin{aligned} EE[\hat{\mu}_2 - \mu]^2 &= EE[\overline{X} + g_n(\overline{X}) - \mu]^2 \\ &= EE[\overline{X} - \mu]^2 + EEg_n^2(\overline{X}) + 2EE(\overline{X} - \mu)g_n(\overline{X}) \\ &\leq \frac{\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}}{n^2} + E_m g_n^2(\overline{X}) + 2\sqrt{\frac{\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}}{n^2} E_m g_n^2(\overline{X})}, \end{aligned} \tag{4.12}$$

where E_m, as before, denotes marginal expectation.

Since $r(\hat{\mu}_1) = \frac{\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}}{n^2}$, if we can show

$$\lim_{n\to\infty} \frac{E_m g_n^2(\overline{X})}{\frac{\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}}{n^2}} = 0, \tag{4.13}$$

(4.12) will imply $e(1,2) \leq 1$. That $e(1,2) \geq 1$ is easy to see. Together, therefore, (4.13) will complete the proof of part (b) of Theorem 7, which we now prove.

Using (4.4), one gets

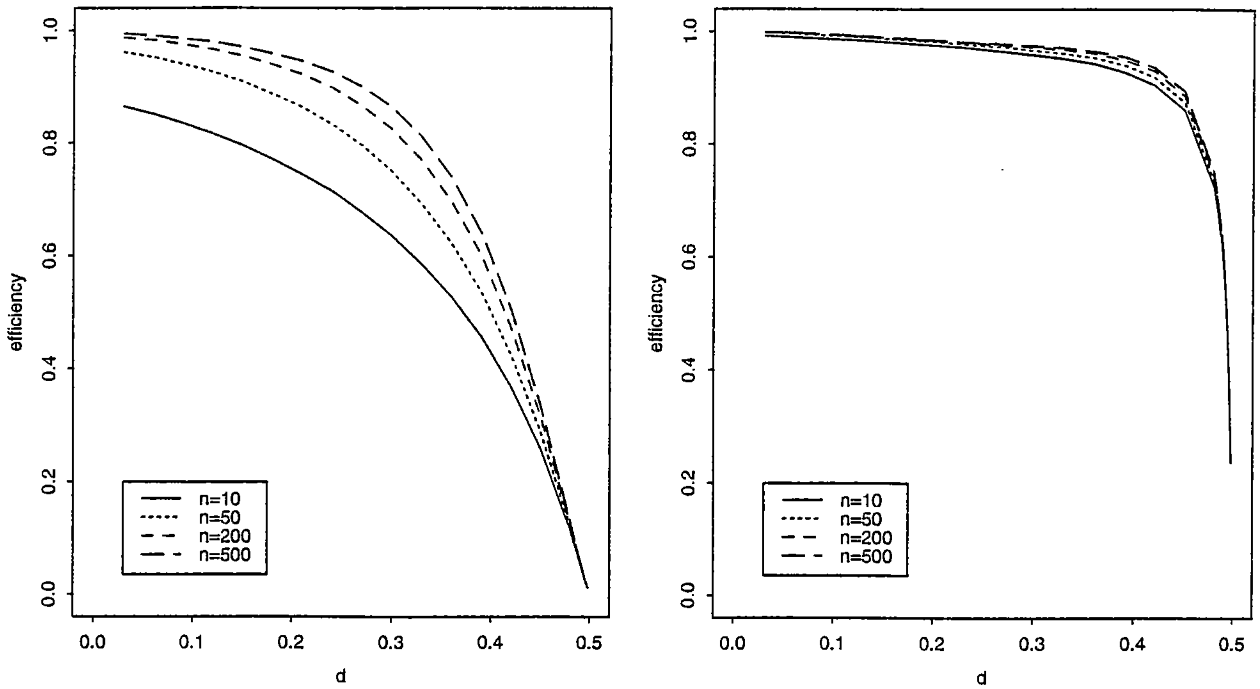
$$\begin{aligned} |g_n(\overline{x})| &\leq \frac{1}{n}(b + a|\overline{x}| + \sqrt{\frac{2}{\pi n}}) \\ \Rightarrow E_m|g_n^2(\overline{X})| &\leq \frac{3b^2}{n^2} + \frac{3a^2}{n^2} E_m|\overline{X}|^2 + \frac{6}{\pi n^3} \\ &\leq \frac{3b^2}{n^2} + \frac{3a^2}{n^2}\{E_\pi(\mu^2) + \frac{\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}}{n^2}\} + \frac{6}{\pi n^3}. \end{aligned} \tag{4.14}$$

Hence,

$$\begin{aligned} \frac{E_m g_n^2(\overline{X})}{\frac{\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}}{n^2}} &\leq \frac{3b^2}{\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}} + \frac{3a^2 E_\pi \mu^2}{\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}} + \frac{3a^2}{n^2} + \frac{6}{\pi n \underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}} \\ &\to 0 \quad as \quad n \to \infty, \end{aligned} \tag{4.15}$$

completing the proof.

Discussion. Theorem 7 says that asymptotically, the sample mean $\overline{X}$, the independence case estimate $\hat{\mu}_2$, and the $BLUE$ $\hat{\mu}_3$, are all safe proxies to the exact Bayes estimate $\hat{\mu}_4$ for most priors. **Figure 1** describes the situation in finite samples for the double exponential prior $\frac{1}{2\tau}e^{-\frac{|\mu|}{\tau}}$, which has variance $2\tau^2$. Note the interesting **Gibbs Phenomenon**: at any given n, there is a drop in efficiency at $d \approx \frac{1}{2}$; but for any given d, this is cured as $n \to \infty$.



$\tau^2 = 0.5$ $\qquad$ $\tau^2 = 12.5$

Figure 1: $e_n(\hat{\mu}_1, \hat{\mu}_4)$ as a function of d for $\frac{1}{2\tau}e^{-\frac{|\mu|}{\tau}}$ prior

5 Interval Estimation of μ

5.1 Introduction

In this section, we will show that asymptotically the easily computed interval $BLUE \pm z_{\alpha/2}s.d.(BLUE)$ has the nominal $1-\alpha$ posterior probability for a general prior on μ. This is certainly a positive fact, as computation of the exact HPD set, which may not be even an interval, is by all means a formidable task. We will also show another rather interesting phenomenon: **the interval $\overline{X} \pm z_{\alpha/2}s.d.(\overline{X})$ behaves randomly**, i.e., its posterior probability p_n converges in law to a random variable p. We will explicity give the density of p and we will see that p acts essentially like a point mass at the nominal level $1-\alpha$, a reassuring property.

5.2 Three Simple Intervals

If the prior is general rather than normal, the HPD set can be very complex. We hope to find some simple intervals instead of the exact HPD set with asymptotically nice properties.

Consider the following three intervals

(a) $I_1 : Y \pm [\underline{1}'\Sigma^{-1}\underline{1}]^{-1/2} z_{\alpha/2}$, where Y denotes the *BLUE*;

(b) $I_2 : \overline{X} \pm n^{-1/2} z_{\alpha/2}$;

(c) $I_3 : \overline{X} \pm \frac{[\underline{1}'\Sigma\underline{1}]^{1/2}}{n} z_{\alpha/2}$;

Remark 1. Note that all three intervals are of the form $T' \pm b_n z_{\alpha/2}$ for appropriate choices of T' and b_n; for I_1, $T' = Y$, $b_n = [\underline{1}'\Sigma^{-1}\underline{1}]^{-1/2}$; for I_2, $T' = \overline{X}$, $b_n = n^{-1/2}$; and for I_3, $T' = \overline{X}$, $b_n = \frac{[\underline{1}'\Sigma\underline{1}]^{1/2}}{n}$.

Remark 2. Because $\frac{\underline{1}'\Sigma\underline{1}}{n^{1+2d}} \sim \frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}$, we can equivalently use

$$I_3' : \overline{X} \pm (\frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)})^{\frac{1}{2}} n^{d-\frac{1}{2}} z_{\alpha/2}$$

instead of I_3.

5.3 Asymptotic Posterior Probability

Let μ_0 be a fixed given value of μ. We will study the behavior of these three intervals under P_{μ_0}. First we state a general theorem which we then apply to the intervals I_1, I_2 and I_3.

Theorem 8. Consider the stationary ARIMA(0,d,0) process as in section 1. Suppose μ_0 is an interior point of Θ, and $\pi(\mu)$ is continuous and positive at μ_0. $\{b_n\}$ is a sequence such that $b_n \to 0$ as $n \to \infty$. T' is an estimator of μ such that $Y - T' \to 0$ *a.s.* (P_{μ_0}). Then the following hold:

(1) If $\lim_{n\to\infty} b_n \sqrt{\underline{1}'\Sigma^{-1}\underline{1}} = c > 0$, then

$$P(\mu \in [Y - b_n z_{\alpha/2}, Y + b_n z_{\alpha/2}] | \underline{X}) \xrightarrow{P_{\mu_0}} \Phi(c z_{\alpha/2}) - \Phi(-c z_{\alpha/2}),$$

(2) If $\lim_{n\to\infty} b_n \sqrt{\underline{1}'\Sigma^{-1}\underline{1}} = 0$, then

$$P(\mu \in [T' - b_n z_{\alpha/2}, T' + b_n z_{\alpha/2}] | \underline{X}) \xrightarrow{P_{\mu_0}} 0,$$

(3) If $\lim_{n\to\infty} b_n \sqrt{\underline{1}'\Sigma^{-1}\underline{1}} = c > 0$, and $\sqrt{\underline{1}'\Sigma^{-1}\underline{1}}(T' - Y) \xrightarrow{\mathcal{L}} w$, a random variable, then

$$P(\mu \in [T' - b_n z_{\alpha/2}, T' + b_n z_{\alpha/2}] | \underline{X}) \xrightarrow{\mathcal{L}} \Phi(w + c z_{\alpha/2}) - \Phi(w - c z_{\alpha/2}).$$

In the above, Φ denotes the $N(0,1)$ CDF.

Remark. Part(1) is a special case of a theorem in Heyde & Johnstone(1979).

Corollary 5.

(a) $P(\mu \in I_1|\underline{X}) \xrightarrow{P_{\mu_0}} 1-\alpha$,

(b) $P(\mu \in I_2|\underline{X}) \xrightarrow{P_{\mu_0}} 0$,

(c) $P(\mu \in I_3|\underline{X}) \xrightarrow{\mathcal{L}} \Phi(\sqrt{C^2-1}\ \ Z + Cz_{\alpha/2}) - \Phi(\sqrt{C^2-1}\ \ Z - Cz_{\alpha/2})$, where $C^2 = C^2(d) = \frac{1}{\tau(d)}$, $\tau(d)$ was defined in (4.7). Z is a standard normal variable.

Proof of Corollary 5:

(a) Here $b_n = \frac{1}{\sqrt{\underline{1}'\Sigma^{-1}\underline{1}}}$, and so $c=1$; by part(1) of Theorem 8,

$$P(\mu \in I_1|\underline{X}) \xrightarrow{P_{\mu_0}} \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 1-\alpha.$$

(b) Here $b_n = n^{-1/2}$, and so $P(\mu \in I_2|\underline{X}) \xrightarrow{P_{\mu_0}} 0$ from part(2) of Theorem 8.

(c) Here $b_n = \frac{[\underline{1}'\Sigma\underline{1}]^{1/2}}{n}$, and so $c = C = \sqrt{\frac{1}{\tau(d)}}$; in addition, $T' = \overline{X}, \sqrt{\underline{1}'\Sigma^{-1}\underline{1}}(T'-Y) \xrightarrow{\mathcal{L}} N(0, C^2-1)$; so by part(3) of Theorem 8,

$$P(\mu \in I_3|\underline{X}) \xrightarrow{\mathcal{L}} \Phi(\sqrt{C^2-1}\ \ Z + Cz_{\alpha/2}) - \Phi(\sqrt{C^2-1}\ \ Z - Cz_{\alpha/2}).$$

Part (c) of Corollary 5 says that the posterior probability of I_3 goes to a random variable

$$p = \Phi(\sqrt{C^2-1}\ \ Z + Cz_{\alpha/2}) - \Phi(\sqrt{C^2-1}\ \ Z - Cz_{\alpha/2}). \tag{5.1}$$

Straightforward calculations using Jacobians give the following density function for p:

$$f(p) = \frac{2}{\sqrt{2\pi(C^2-1)}} \frac{e^{-(g^{-1}(p))^2/2(C^2-1)}}{\phi(g^{-1}(p) - Cz_{\alpha/2}) - \phi(g^{-1}(p) + Cz_{\alpha/2})} \tag{5.2}$$

and

$$\overline{F}(\beta) = P(p \geq \beta) = \Phi(\frac{g^{-1}(\beta)}{\sqrt{C^2-1}}) - \Phi(-\frac{g^{-1}(\beta)}{\sqrt{C^2-1}}), \tag{5.3}$$

where

$$g(x) = \Phi(x + Cz_{\alpha/2}) - \Phi(x - Cz_{\alpha/2}). \tag{5.4}$$

Example 2. Consider the nominal level of 95%, i.e., $\alpha = .05$. The following table is quite reassuring. It says that although the interval $\overline{X} \pm 1.96s.d.(\overline{X})$ behaves randomly in theory, its posterior probability is likely to be very good asymptotically.

Table 6. Values of $\overline{F}(\beta)$

β	d	
	0.25	0.49
0.90	1	1
0.91	0.999996	1
0.92	0.999995	1
0.93	0.999252	1
0.94	0.987336	1
0.95	0.683821	0.687326

Proof of Theorem 8: The posterior density of μ is

$$\pi(\mu|\underline{X}) = \pi(\mu|y) = \sqrt{\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{2\pi}} e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}(\mu-y)^2}{2}} \pi(\mu)/m(y),$$

where

$$m(y) = \int_\Theta \sqrt{\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{2\pi}} e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}(\mu-y)^2}{2}} \pi(\mu)d\mu.$$

Therefore the posterior probability of the general interval $T' \pm b_n z_{\alpha/2}$ is

$$\int_{T'-b_n z_{\alpha/2}}^{T'+b_n z_{\alpha/2}} \sqrt{\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{2\pi}} e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}(\mu-y)^2}{2}} \pi(\mu)/m(y). \tag{5.5}$$

By standard arguments, $m(y) \xrightarrow{P_{\mu_0}} \pi(\mu_0)$. So we will only sketch how to handle the numerator

$$S_1 = \int_{T'-b_n z_{\alpha/2}}^{T'+b_n z_{\alpha/2}} \sqrt{\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{2\pi}} e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}(\mu-y)^2}{2}} \pi(\mu)d\mu. \tag{5.6}$$

First construct a neighborhood $N_\delta = (\mu_0 - \delta, \mu_0 + \delta)$ such that given $\epsilon > 0$,

$$(1-\epsilon)\pi(\mu_0) < \pi(\mu) < (1+\epsilon)\pi(\mu_0),$$

for $\mu \in N_\delta$. Next, since $Y - T'$, $Y - \mu_0 \to 0$ $a.s.$, and $b_n \to 0$, one can say that for all large n, with probability 1,

$$(T' - b_n z_{\alpha/2}, T' + b_n z_{\alpha/2}) \subset N_\delta. \tag{5.7}$$

(5.6) and (5.7) imply

$$(1-\epsilon)\pi(\mu_0)\int_{T'-b_n z_{\alpha/2}}^{T'+b_n z_{\alpha/2}} \sqrt{\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{2\pi}} e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}(\mu-y)^2}{2}} d\mu < S_1 < (1+\epsilon)\pi(\mu_0)\int_{T'-b_n z_{\alpha/2}}^{T'+b_n z_{\alpha/2}} \sqrt{\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{2\pi}} e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}(\mu-y)^2}{2}} d\mu, \quad (5.8)$$

We specialize to only case (3) of our theorem. This is the case when $b_n\sqrt{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}} \to c$ and $\sqrt{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}(T'-Y) \to w$, a random variable.

Now,

$$\begin{aligned}
& \int_{T'-b_n z_{\alpha/2}}^{T'+b_n z_{\alpha/2}} \sqrt{\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{2\pi}} e^{-\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}(\mu-y)^2}{2}} d\mu \\
= \;& \Phi[\sqrt{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}(T'-y+b_n z_{\alpha/2})] - \Phi[\sqrt{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}(T'-y-b_n z_{\alpha/2})] \\
\xrightarrow{\mathcal{L}} \;& \Phi(w+cz_{\alpha/2}) - \Phi(w-cz_{\alpha/2}), \quad \textit{by Slutsky's theorem.}
\end{aligned}$$

Thus, as $\epsilon > 0$ is arbitrary,

$$S_1 \xrightarrow{\mathcal{L}} [\Phi(w+cz_{\alpha/2}) - \Phi(w-cz_{\alpha/2})]\pi(\mu_0),$$

and combining this with (5.5), case (3) of the theorem follows. The other two cases are omitted.

Table 7(next page) gives simulated posterior probability of the interval centered at the $BLUE$ for $n = 20$ and $n = 50$. The numbers are very encouraging.

6 Estimation of σ^2

6.1 Introduction

In this section, the important variance estimation problem will be discussed. Unlike the case of point estimation of μ, the everyday estimate s^2 can no longer be safely used; however, the UMVUE can be. As before, the criterion is Bayes risk. The principal tool is a Bayes risk lower bound given in Brown and Gajek(1990); we will call it the **Brown-Gajek lower bound**. First we give the exact setup.

Table 7. Simulated Posterior Probability of I_1 under $\frac{1}{2}e^{-|\mu|}$ prior (nominal $\alpha = .05$)

n	d	I_1	$Prob$
20	0.01	(-0.17,0.73)	0.9594
		(0.17,1.07)	0.9448
		(-0.85,0.05)	0.9537
	0.25	(-0.79,1.11)	0.9762
		(-0.39,1.52)	0.9673
		(-0.83,1.07)	0.9765
	0.49	(-5.49,9.81)	0.9998
		(-11.42,3.88)	0.9984
		(-7.72,7.59)	0.9999
50	0.01	(-0.21,0.37)	0.9594
		(-0.66,-0.08)	0.9483
		(-0.26,0.32)	0.9601
	0.25	(-0.92,0.59)	0.9718
		(-0.80,0.71)	0.9727
		(-1.08,0.43)	0.9688
	0.49	(-10.53,4.64)	0.9993
		(-11.37,3.80)	0.9983
		(-5.44,9.73)	0.9996

6.2 Exact Setup

For the stationary ARIMA(0,d,0) process defined in section 1, one has $\underline{X} \sim N(\mu\mathbf{1}, \sigma^2\Sigma)$, and suppose both μ and σ^2 are unknown. We define

$$T_1 = [\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}]^{-1}\underline{\mathbf{1}}'\Sigma^{-1}\underline{X} \tag{6.1}$$

$$T_2 = \frac{1}{n-1}(\underline{X}'\Sigma^{-1}\underline{X} - \underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}T_1^2) \tag{6.2}$$

(T_1, T_2) is a sufficient statistic for (μ, σ^2). Furthermore, T_1 and T_2 are respectively the UMVUE of μ and σ^2. If we assume $d = 0$, T_1 and T_2 become $\overline{X}$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

For notational convenience, let $\theta = \sigma^2$. In this section, we use the natural invariant loss

$$L(a, \theta) = \frac{1}{\theta^2}(a - \theta)^2 \tag{6.3}$$

as a loss function for estimation of θ. Under this loss, the Bayes estimate for $\theta = \sigma^2$ equals

$$\hat{\theta} = \frac{E(\frac{1}{\theta}|\underline{X})}{E(\frac{1}{\theta^2}|\underline{X})}. \tag{6.4}$$

First let us see an illustrative example.

6.3 An Anticipatory Example

Suppose the prior distribution for μ and θ are the natural conjugate priors

$$\begin{aligned} \mu|\theta &\sim N(0,\theta) \\ \theta &\sim Inverse\ \ Gamma(\alpha,\beta) \end{aligned} \tag{6.5}$$

(see Berger(1986)). Then the Bayes estimate of θ is

$$\hat{\theta} = [\frac{1}{\beta} + \frac{\underline{X}'\Sigma^{-1}\underline{X}}{2} - \frac{(\underline{\mathbf{1}}'\Sigma^{-1}\underline{X})^2}{2(n+1)}]/(\alpha + \frac{n}{2} + 1). \tag{6.6}$$

On calculation, β drops out and the Bayes risk is

$$r(\pi) = \frac{2}{2\alpha + n + 2}. \tag{6.7}$$

To get the asymptotic efficiency of T_2 and s^2 relative to the correct Bayes estimate in (6.6), we use the following expressions obtained on simple calculations (see Seber(1977)).

$$r(s^2) = \frac{2}{(n-1)^2} tr[(\Sigma H)^2] + [\frac{1}{n-1} tr(\Sigma H) - 1]^2, \tag{6.8}$$

where $H = I - \frac{J}{n}$, $J = \underline{\mathbf{1}}\underline{\mathbf{1}}'$, I is the identity matrix, and

$$r(T_2) = \frac{2}{n-1}. \tag{6.9}$$

However,

$$tr(\Sigma H) = n\frac{\Gamma(1-2d)}{\Gamma^2(1-d)} - \frac{1}{n}\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}}, \tag{6.10}$$

and hence

$$[\frac{1}{n-1} tr(\Sigma H) - 1]^2 \to [\frac{\Gamma(1-2d)}{\Gamma^2(1-d)} - 1]^2. \tag{6.11}$$

Also,

$$\begin{aligned} (\Sigma H)^2 &= \Sigma^2 - \frac{1}{n}\Sigma^2\underline{\mathbf{1}}\underline{\mathbf{1}}' - \frac{1}{n}\Sigma\underline{\mathbf{1}}\underline{\mathbf{1}}'\Sigma + \frac{1}{n^2}(\Sigma\underline{\mathbf{1}}\underline{\mathbf{1}}')^2 \\ \Rightarrow tr[(\Sigma H)^2] &= \sum_i\sum_j \gamma(i,j)^2 - \frac{2}{n}\underline{\mathbf{1}}'\Sigma^2\underline{\mathbf{1}} + \frac{1}{n^2}(\underline{\mathbf{1}}'\Sigma\underline{\mathbf{1}})^2, \end{aligned} \tag{6.12}$$

and hence

$$\frac{1}{(n-1)^2} tr[(\Sigma H)^2] \to 0. \tag{6.13}$$

(6.8),(6.11) and (6.13) imply that even when $n \to \infty$, the Bayes risk of s^2 will not converge to 0; indeed,

$$r(s^2) \to [\frac{\Gamma(1-2d)}{\Gamma^2(1-d)} - 1]^2. \tag{6.14}$$

By (6.7), (6.9) and (6.14), we get

$$\begin{aligned} e(T_2, \hat{\theta}) &= 1, \\ e(s^2, \hat{\theta}) &= 0. \end{aligned}$$

Thus **the UMVUE is asymptotically efficient, but s^2 is very bad**. This is extended to more general priors by use of the Brown-Gajek lower bound in the following subsection.

6.4 Brown-Gajek Lower Bound

Theorem 9. Suppose a density function $f(x|\theta)$ satisfies the regularity conditions which are required for the Cramer-Rao inequality. For the loss function $w(\theta)(a-\theta)^2$, where $w(\theta)$ and the prior density function $\pi(\theta)$ are absolutely continuous, the Bayes risk satisfies

$$r(\pi) \geq \frac{C^2}{C+D},$$

where

$$\begin{aligned} C &= \int \frac{w(\theta)\pi(\theta)}{I(\theta)} d\theta, \\ D &= \int [(\frac{w(\theta)\pi(\theta)}{I(\theta)})']^2 / [w(\theta)\pi(\theta)] d\theta, \end{aligned} \tag{6.15}$$

and

$$I(\theta) = E_\theta[(\frac{\partial}{\partial\theta} log f(x|\theta))^2].$$

6.5 Asymptotic Efficiency

Theorem 10. For the model in section 6.2, suppose that given θ, μ has the normal prior $N(0, \theta)$, and θ has a prior with an absolutely continuous prior density $\pi(\theta)$, satisfying

$$k = \int \frac{(\theta\pi'(\theta))^2}{\pi(\theta)} d\theta < \infty.$$

Then, under the loss function (6.3), the following hold:

$$\begin{aligned} e(T_2, \hat{\theta}) &= 1, \\ e(s^2, \hat{\theta}) &= 0. \end{aligned}$$

Here $\hat{\theta}$ denotes the correct Bayes estimate under the assumed prior.

Remark 1. It will be aesthetically pleasing and also useful to be able to remove the assumption of a conditional normal prior on μ. However, we were unable to do so.

Remark 2. The integrability condition $\int \frac{(\theta\pi'(\theta))^2}{\pi(\theta)} d\theta < \infty$ is satisfied by all gamma and some other common priors.

Proof: The key simplification is the following:

If

$$T_1|\mu, \theta \sim N(\mu, \frac{\theta}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}),$$

and

$$\mu|\theta \sim N(0, \theta),$$

then

$$T_1|\theta \sim N(0, \theta(1 + \frac{1}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}})). \tag{6.16}$$

For convenience, we use the following notation:

$$\begin{aligned} f_1(.|\mu, \theta) &: \textit{ the density of } N(\mu, \frac{\theta}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}) \textit{ distribution} \\ f_2(.|\theta) &: \textit{ the density of } Gamma(\frac{n-1}{2}, \frac{2\theta}{n-1}) \textit{ distribution} \\ f_3(.|\theta) &: \textit{ the density of } N(0, \theta(1 + \frac{1}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}})) \textit{ distribution} \\ f(., .|\theta) &: f_3(.|\theta) f_2(.|\theta) \end{aligned} \tag{6.17}$$

Then

$$\begin{aligned} \hat{\theta} &= \frac{\int\int \theta^{-1} f_1(t_1|\mu, \theta) f_2(t_2|\theta)\pi(\mu|\theta)\pi(\theta) d\mu d\theta}{\int\int \theta^{-2} f_1(t_1|\mu, \theta) f_2(t_2|\theta)\pi(\mu|\theta)\pi(\theta) d\mu d\theta} \\ &= \frac{\int \theta^{-1} f_3(t_1|\theta) f_2(t_2|\theta)\pi(\theta) d\theta}{\int \theta^{-2} f_3(t_1|\theta) f_2(t_2|\theta)\pi(\theta) d\theta} \\ &= \frac{\int \theta^{-1} f(t_1, t_2|\theta)\pi(\theta) d\theta}{\int \theta^{-2} f(t_1, t_2|\theta)\pi(\theta) d\theta} \end{aligned} \tag{6.18}$$

By a tedious but direct calculation, in which $\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}$ terms cancel out,

$$I(\theta) = -E[\frac{\partial^2}{\partial\theta^2} log f(t_1, t_2|\theta)] = \frac{n}{2\theta^2}. \tag{6.19}$$

Furthermore, in expression (6.15),

$$\begin{aligned} w(\theta) &= \theta^{-2}, \\ C &= \int \frac{\theta^{-2}\pi(\theta)}{I(\theta)} d\theta = \frac{2}{n}, \\ D &= \int \{[(\frac{\theta^{-2}\pi(\theta)}{I(\theta)})']^2/[\theta^{-2}\pi(\theta)\} d\theta \\ &= \frac{4}{n^2}\int \frac{(\theta\pi'(\theta))^2}{\pi(\theta)} d\theta = \frac{4}{n^2}k. \end{aligned} \tag{6.20}$$

Therefore, by the lower bound of Theorem 9,

$$r(\pi) \geq \frac{(\frac{2}{n})^2}{\frac{2}{n} + \frac{4}{n^2}k} = \frac{\frac{2}{n}}{1 + \frac{2}{n}k}. \tag{6.21}$$

Now, the formulae (6.8) and (6.9) for $r(s^2)$ and $r(T_2)$ are in fact generally valid and so, from (6.21),

$$1 \geq e(T_2, \theta) = \lim_{n\to\infty} \frac{r(\pi)}{r(T_2)} \geq \lim_{n\to\infty} \frac{\frac{2}{n}}{1 + \frac{2k}{n}} \frac{n-1}{2} = 1;$$

also, plainly,

$$e(s^2, T_2) = 0,$$

and hence

$$e(s^2, \hat{\theta}) = e(s^2, T_2)e(T_2, \hat{\theta}) = 0.$$

This completes the proof of the theorem.

References

[1] Adenstedt, R.K.(1974). On Large-Sample Estimation For The Mean of a Stationary Random Sequence. *Ann. Stat.* **2** 1095-1107.

[2] Beran, J.(1992). Statistical Methods for Data with Long-Range Dependence. *Statistical Science* **7** 404-427.

[3] Beran, J.(1994). *Statistics for Long-Memory Processes.* Chapman & Hall, New York.

[4] Berger, J.O.(1986). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York.

[5] Berger, J.O.(1996). Bayes Factors. (*for Encyclopaedia of Stat., S. Kotz, Ed.*)

[6] Bickel, P.J.(1981). Minimax Estimation of the Mean of a Normal Distribution When the Parameter Space Is Restricted. *Ann. Stat.* **9** 1301-1309.

[7] Billingsley, P.(1995). *Probability and Measure (3rd ed.).* John Wiley & Sons, New York.

[8] Brockwell, P.J. and Davis, R.A.(1991). *Time Series:Theory and Methods.* Springer-Verlag, New York.

[9] Brown, L.D.(1971). Admissible Estimators, Recurrent Diffusions, and Insoluble Boundary Value Problems. *Ann.Math.Stat.* **42** 855-903.

[10] Brown, L.D.(1986). *Fundamentals of Statistical Exponential Families.* Hayward, California.

[11] Brown, L.D. and Gajek, L.(1990). Information Inequalities for the Bayes Risk. *Ann. Stat.* **18** 1578-1594.

[12] Dahlhaus, R. (1995). Efficient location and regression estimation for long range dependent regression models. *Ann. Stat.* **23** 1029-1047.

[13] Dahlhaus, R. (1989). Efficient Parameter Estimation for Self-similar Processes. *Ann. Stat.* **17** 1749-1766.

[14] DasGupta, A.(1997). The Telegraph Equation, An Expectation Identity, and Applications. (*In Preparation*)

[15] Fox, R. and Taqqu, M.S. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Th. Rel. Fields* **74** 213-240.

[16] Geweke, J. and Porter-Hudak, S.(1983). The Estimation and Application of Long Memory Time Series Models. *J.Time Ser. Anal.* **4** 221-237.

[17] Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle's estimate. *Probab. Th. Rel. Fields* **86** 87-104.

[18] Heyde, C.C. and Johnstone, I.M.(1979). On Asymptotic Posterior Normality for Stochastic Processes. *J.R.Stat. Soc. B* **41** 184-189.

[19] Huber, P.J.(1981). *Robust Statistics.* John Wiley, New York.

[20] Jeffreys, H.(1961). *Theory of Probability (3rd ed.).* Oxford, U.K.: Oxford University Press.

[21] Karlin, S. and Taylor, H.M.(1975). *A First Course in Stochastic Processes (2ed ed.).* Academic Press, Inc.

[22] Kass, R.E. and Raftery, A.E.(1995). Bayes Factors and Model Uncertainty. *JASA* **90** 773-795.

[23] Kass, R.E. and Wasserman, L.(1995). A Reference Bayesian Test for Nested Hypotheses. *JASA* **90** 928-934.

[24] Koul, H.L.(1992). M-estimators in Linear Models with Long Range Dependent Errors. *Stat. and Prob. Lett.* **14** 153-164.

[25] Lebedev, N.N.(1965). *Special Functions and Their Applications.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

[26] Mathai, A.M.(1992). *Quadratic form in random variables: theory and application.* Dekker, New York.

[27] Provost, S.B.(1988). The exact density of a general linear combination of gamma variates. *Metron* **46** 61-64.

[28] Robinson, P.M.(1994). Time Series with Strong Dependence, In Advances in Econometrics. *Sixth World Congress (C.A.Sims,Ed.)* **1**, 47-95, Cambridge University Press.

[29] Samarov, A. and Taqqu, M.S.(1988). On the Efficiency of The Sample Mean In Long-Memory Noise. *J. Time Ser. Anal.* **9** 191-200.

[30] Seber, G.A.F. (1977). *Linear Regression Analysis.* John Wiley & Son, New York.

[31] Yajima, Y.(1985). On Estimation of Long-Memory Time Series Models. *Austral.J.Statist.* **27** 303-320.

[32] Yajima, Y.(1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Ann. Stat.* **19** 158-177.