A NEW APPROACH TO DEFAULT PRIORS AND ROBUST BAYES METHODOLOGY*

by

Mohan Delampady
Purdue University and
Indian Statistical Institute

Anirban DasGupta
Purdue University

George Casella
Cornell University

Herman Rubin
Purdue University

William E. Strawderman
Rutgers University

Technical Report #99-27

Department of Statistics
Purdue University
West Lafayette, IN USA

December 1999

* Research was supported by the National Science Foundation and the National Security Agency.

◇

A NEW APPROACH TO DEFAULT PRIORS AND ROBUST BAYES METHODOLOGY

by

Mohan Delampady
Purdue University and
Indian Statistical Institute

Anirban DasGupta
Purdue University

George Casella
Cornell University

Herman Rubin
Purdue University

William E. Strawderman
Rutgers University

Abstract

Following the developments in DasGupta et al (1999), in this article we propose and explore a new method for construction of proper default priors, and a method to select one Bayes estimate from a family for actual use. The results are based on asymptotic expansions of certain marginal correlations.

The answers that emerge have nice general structures. The default prior methodology finally amounts to minimization of Fisher information. As a consequence, for any location parameter problem, the Bickel prior works out as the default prior if the location parameter is bounded. The selected Bayes estimate, on the other hand, corresponds to "Gaussian tilting" of an initial reference prior.

The calculations are illustrated with examples and computation.

1. Introduction

A radically different way of looking at the Pearson correlation coefficient in statistical methodology is detailed in DasGupta et al. (1999) by projecting it as a binding theme to connect together various approaches to statistical inference. It is also shown there that some of the properties of the Pearson correlation coefficient can lead to useful and substantial developments in mathematical statistics, particularly Bayesian statistics. This last theme is further developed in this article.

We begin with an observable X which is distributed as $f(x|\theta)$. The parameter θ is distributed according to a prior π, and hence there is then a joint probability distribution which we shall call P. As in our previous article, all the developments here also follow from consideration of the Pearson correlation coefficient between two functions $g_1(X, \theta)$ and $g_2(X, \theta)$ under the probability distribution P. However, most of our derivations here involve asymptotic arguments which can be rather intractable in general, and hence we shall mostly restrict our attention to random samples from location families. The exact setup required is outlined later in this section.

In Section 2, we use the correlation coefficient to address an important problem in robust Bayesian analysis. The overwhelming bulk of the work in robust Bayesian analysis has dealt with sensitivity with respect to the choice of the prior. However, the problem of selecting a specific Bayes estimate is also an important one. Indeed, classical robustness flourished and was taken seriously because specific procedures such as M estimates were developed. We first select a reference prior π; this reference prior need not belong to the family Γ of priors in consideration. However, usually one will choose π from inside Γ. For a general prior ν in Γ, we then consider the correlation $\rho_P(\theta, \delta_\nu)$ as a criterion for selecting a specific estimate δ_ν; ρ_P is calculated under the chosen reference prior π. We maximize an accurate approximation $\hat{\rho}_P(\theta, \delta_\nu)$ over δ_ν. The specific chosen estimate δ_ν is Bayes with respect to a prior density of the form $\nu(\theta) = \text{constant} \cdot \pi(\theta) \cdot e^{-\frac{1}{2\tau^2}(\theta-\mu)^2}$. There are several interesting things about this. First, the generality of the form; one always gets a Gaussian factor. Second, $\nu(\theta)$ has the following interpretation: presumably one will start with a flat reference density $\pi(\theta)$ due to robustness concerns. The final prior $\nu(\theta)$ is formally just the posterior density of θ when a Gaussian observation has been obtained and θ has the prior $\pi(\theta)$. By starting with a flat prior π and ultimately settling for a "formal posterior" as ν, one will pull in the tails but it will still be a more conservative choice than a straight Gaussian prior. We have examples then illustrating these results. Note that, in other contexts, Bayesians have been talking about such a "tilting" of an initial prior by collecting a pilot sample; see Perez (1998) for example. It is interesting that we see this tilting emerge in a purely theoretical way in our results here.

Next, in Section 3, we apply the correlation criterion to outline a new method for construction of default priors. Default prior Bayesian analysis has been a very active area of research for a considerable time. After the initial classic contributions of Jeffreys and Laplace, the recent renewed interest has much to do with objective Bayesian inference and the realization that default prior Bayes methods often provide satisfactory frequentist properties. See Berger (1986), Efron (1986), Stein (1982), among many. Conventional default priors in use tend to be improper; thus, nice frequentist properties like admissibility have to be often established case by case. We develop here an outline for construction of proper default priors. The method suggested is general, but we have worked it out here in detail only for a location parameter.

The method we suggest is as follows. Many Bayesians take the view that post-data opinion about a parameter should be reported simply in terms of a posterior density. On the other hand, there is another clear candidate for such a summary, namely the likelihood function. Just as one can try to minimize an appropriate distance between the two summaries, we suggest maximizing the correlation between them under the joint probability measure P.

Now, the exact correlation, of course, is not something that one can work with. So we provide an appropriate expansion for the correlation, and maximize the appropriate term of this expansion. The expansion is very technical and is presented in the appendix. It is remarkable that at the end the maximization based on this expansion corresponds to minimization of the Fisher information of the prior in the chosen family of proper priors. Minimization of Fisher information is a well known variational problem that has arisen in other statistical problems; see Bickel (1981), Bickel and Collins (1983), Huber (1964, 1974), Levit (1979, 1980), Kagan, Linnik and Rao (1973), and Brown (1971). We find this ultimate reduction of our approach to minimization of Fisher information quite interesting. As a result of this reduction, the Bickel prior (1981) is now seen to have the asymptotic correlation maximization property; compare this with the asymptotic entropy maximization by Jeffrey priors (Clarke and Barron (1994)).

The asymptotic expansions we needed are substantially more intense than what is necessary in other problems (e.g., Ghosh, Sinha and Joshi (1982)) because we need expansions to more terms for our results. The derivations thus require more smoothness assumptions on the likelihood function and the prior. Exact finite sample implementation of our approach was not pursued in this article, except we have shown that our formulation leads to the uniform prior in the Binomial case for every finite sample size n.

1.1 The Set-up

The following general notation will be used in the sequel: π and ν will denote prior densities for the parameter θ, m the marginal of X, and E_θ will denote conditional expectation given θ. Cov_P, Var_P, and ρ_P will, respectively, denote the covariance, variance and correlation under the joint distribution P, whereas Cov_π and Var_π will denote the covariance and variance under the distribution π on θ.

We consider i.i.d. observations $X_1, X_2, \ldots$ from a location parameter density $f(x|\theta) = e^{-h(x-\theta)}$; we assume that h is seven times continuously differentiable, and $h^{(6)}$ and $h^{(7)}$, are bounded. The following notation will be used:

$$
\begin{aligned}
\mathcal{L}(\theta, x) &= \sum_{i=1}^{n} \log f(x_i|\theta), \quad \hat{\theta} = \text{ mle of } \theta, \\
\mathcal{L}^{(i)} &= \frac{d^i}{d\theta^i}\mathcal{L}(\theta, x)|_{\theta=\hat{\theta}}, \\
\ell_i &= E_\theta\left(\frac{d^i}{d\theta^i}\log f(X_1|\theta)\right), \\
\sigma^2 &= \ell_2, \\
\hat{\sigma}^2 &= -\left(\frac{\mathcal{L}^{(2)}}{n}\right)^{-1}, \\
w_2 &= \mathrm{var}_p(h''(X))
\end{aligned}
\tag{1.1}
$$

We also assume that $\sigma^2 > 0$, $\ell_i = 0$ for odd i, and $\mathcal{L}^{(1)} = 0$, i.e., the mle $\hat{\theta}$ solves the likelihood equation.

By elementary calculations, one can see that $\mathcal{L}^{(3)} = O(n^{\frac{1}{2}})$, $E_p(W_n) = O(n^{-1})$ and $EP(W_n^2) = w_2 + O(n^{-1})$, where $W_n = \sqrt{n}\left(\frac{\mathcal{L}^{(2)}}{n} + \frac{1}{\sigma^2}\right)$. The normal location model with known variance fits into our set-up trivially. It can be checked that some other standard location models such as Student's t and Logistic also fit into our set-up.

Regarding the prior distributions under consideration in this article we make the following assumption.

Assumption A. The reference prior density π and every density ν in Γ is five times continuously differentiable a.e., with bounded fourth and fifth derivatives, and $E_\nu(\theta) = 0$, $E_\nu(\theta^2) < \infty$.

2. Selecting a Bayes Estimate

Robust Bayesian analysis has almost exclusively concentrated on sensitivity of the Bayes estimate and other posterior quantities to the choice of the prior. There is an extensive literature on this now; see the review article Berger (1994). Far less has been done in the direction of presenting methods for choosing a Bayes estimate from a collection specified by a family of priors; see Zen and DasGupta (1993) for some results on this question. We present below a method for selecting a particular Bayes estimate from a collection for the location parameter case, by using the correlation criterion. The notation and the final result are as follows: let Γ be a specified family of priors. Let π be a special prior, a reference. Let ν be a generic element of Γ with δ_ν as the corresponding Bayes estimate. The criterion for selecting a special δ_ν is to maximize over all ν the correlation $\rho_{_P}(\theta, \delta_\nu)$, under the reference prior π. The specific prior ν which does this is in general of the form $\nu^*(\theta) = \text{constant } \pi(\theta) \cdot e^{-\frac{1}{2\tau^2}(\theta-\mu)^2}$. Here μ, τ can be flexibly chosen so that ν^* belongs to Γ. If no such ν^* belongs to Γ, our method fails. But in many common examples, one will be able to find ν^* of the above form in the class Γ. Note the interesting general form of ν^*. If the reference prior π is flat, the selected prior ν^* still has normal tails. The selected prior ν^* is exactly a normal if and only if the reference π is already normal. We will see examples after the following presentation.

2.2. A Useful Approximation to $\rho_{_P}(\theta, \delta_\nu)$

The problem we wish to address cannot be solved in closed form (and possibly in any form) if we work with the exact correlation $\rho_{_P}(\theta, \delta_\nu)$. Instead, we again present an approximation $\hat{\rho}_{_P}(\theta, \delta_\nu)$. We have found in some test cases that the approximation is accurate; it is an asymptotic approximation, but can be highly accurate even for $n = 3$. The derivation of the approximation is intensely technical. So we shall break it up into little steps at a time and we will present only the gist.

We now present the approximation $\hat{\rho}_{_P}(\theta, \delta_\nu)$ which we shall maximize over ν. For this, first, we need an expansion for the Bayes estimate $\delta_\nu(x)$ as a function of the MLE $\hat{\theta}$ and $x - \hat{\theta}.1$ (here 1 stands for the vector with each element as 1).

Proposition 1. The Bayes estimate $\delta_\nu(x)$ satisfies

$$\delta_\nu(x) = \hat{\theta} + \frac{\sigma^2}{n}\frac{\nu'(\hat{\theta})}{\nu(\hat{\theta})} + \frac{\sigma^4}{2n^{3/2}}\left(\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 2\frac{\nu'(\hat{\theta})}{\nu(\hat{\theta})}W_n\right)$$

$$+\frac{\sigma^4}{2n^2}\left(\frac{\nu'''(\hat{\theta})}{\nu(\hat{\theta})}-\frac{\nu'(\hat{\theta})\nu''(\hat{\theta})}{\nu^2(\hat{\theta})}+\frac{\nu'(\hat{\theta})}{\nu(\hat{\theta})}\sigma^2\left(2W_n^2+\frac{\mathcal{L}^{(4)}}{n}\right)+\sigma^2\left(\frac{\mathcal{L}^{(5)}}{4n}+2\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\right)$$
$$+O(n^{-5/2}), \tag{2.1}$$

uniformly in x.

Proof: Details of this long derivation are deferred to Appendix. We would like to note here that expansions for the posterior mean of this nature can be found in Ghosh (1994), Ghosh, Sinha and Joshi (1982), Johnson (1970) and Lindley (1961). However, they are only accurate up to $O(n^{-3/2})$, whereas we need an approximation which is good to $O(n^{-5/2})$ (and, this of course requires a lot more work).

Using Proposition 1, we have the following asymptotic expressions for the covariance of δ_ν with θ and the variance of δ_ν. Proofs of Propositions 2 and 3 are again algebraically very intensive. However, we have outlined the proof of Proposition 2 in the Appendix since many other proofs (including that of Proposition 3) are similar.

For notational convenience, the following notation is used:

$$\left.\begin{array}{l}
a=\int\theta\pi(\theta)\frac{\nu'(\theta)}{\nu(\theta)}d\theta,\quad b=\int\theta\pi(\theta)\frac{\nu^{(3)}(\theta)}{\nu(\theta)}d\theta,\\
c=\int\theta\pi(\theta)\frac{\nu'(\theta)\nu''(\theta)}{\nu^2(\theta)}d\theta,\quad d=\int\theta\pi''(\theta)\frac{\nu'(\theta)}{\nu(\theta)}d\theta, f=\int\pi'(\theta)\frac{\nu'(\theta)}{\nu(\theta)}d\theta,\\
g=\int\pi(\theta)\left(\frac{\nu'(\theta)}{\nu(\theta)}\right)^2d\theta-\left(\int\pi(\theta)\frac{\nu'(\theta)}{\nu(\theta)}d\theta\right)^2
\end{array}\right\} \tag{2.2}$$

Also, recall that $w_2=\mathbf{Var}_P(h''(X))$ and $l_4=\mathbf{E}_P(h^{(4)}(X))$. In addition, $k(\pi)$, $k_1(\pi)$ and $K(\pi)$ used below are constants which can depend on π but not on ν.

Proposition 2.

$$\begin{aligned}\mathrm{Cov}_P(\theta,\delta_\nu)=V(\pi)+\frac{\sigma^2}{n}a+\frac{\sigma^4}{8n}(3w_2+l_4)+\frac{\sigma^4}{2n^2}(b-c+d+2f)\\
+\frac{5\sigma^6}{8n^2}(3w_2+l_4)a+\frac{\sigma^6}{4n^2}k(\pi)+O(n^{-\frac{5}{2}}).\end{aligned} \tag{2.3}$$

Proposition 3.

$$\begin{aligned}\mathrm{Var}_P(\delta_\nu)=V(\pi)+\frac{\sigma^2}{n}(1+2a)+\frac{\sigma^4}{8n}V(\pi)(3w_2+l_4)+\frac{\sigma^4}{n^2}(b-c+d+g)\\
+\frac{5\sigma^6}{4n^2}(3w_2+l_4)a+\frac{\sigma^6}{4n^2}k_1(\pi)+O(n^{-\frac{5}{2}}).\end{aligned} \tag{2.4}$$

These two results now lead to the desired approximation to $\rho_P(\theta, \delta_\nu)$ stated below.

Proposition 4.

$$\rho_P(\theta, \delta_\nu) = \hat{\rho}_P(\theta, \delta_\nu) + O(n^{-\frac{5}{2}}),$$

where

$$\hat{\rho}_P(\theta, \delta_\nu) = 1 - \frac{\sigma^2}{2n}\left(\frac{\sigma^2}{8}(3w_2 + l_4) - \frac{1}{V(\pi)}\right) + \frac{\sigma^4}{n^2}\left\{(f - \frac{g}{2})\frac{1}{V(\pi)} + (a + \frac{a^2}{2})\frac{1}{V^2(\pi)}\right\}$$
$$+ \frac{\sigma^4}{n^2V^2(\pi)}\left(\frac{3}{8} + \frac{\sigma^2}{8}(3w_2 + l_4)(2 - \frac{\sigma^2}{8}(3w_2 + l_4)) + K(\pi)\right). \qquad (2.5)$$

Proof: Use the definition of $\rho_P(\theta, \delta_\nu)$ and substitute the expressions (2.3) and (2.4) given above. (2.5) will then follow on simple algebra. Details are given in Appendix.

We can now state the result describing the particular selected Bayes estimate $\delta_\nu(X)$.

Proposition 5. The estimate $\delta_\nu(X)$ maximizing $\hat{\rho}_P(\theta, \delta_\nu)$ is Bayes with respect to the prior density

$$\nu(\theta) = ce^{-\frac{1}{2\tau^2}(\theta-\mu)^2}\pi(\theta), \qquad (2.6)$$

where μ, τ^2 are arbitrary and c is a normalizing constant.

Remark. $\nu(\theta)$ has the following nice interpretation. Start with a reference prior $\pi(\theta)$. From this construct the posterior density when an observation from the $N(\theta, \tau^2)$ distribution becomes available. Take this as the final prior ν and use the corresponding Bayes estimate δ_ν. Now, what is the effect of such a prior? Presumably, one will start with a rather flat prior π as the reference due to robustness concerns. The normal factor in the expression for $\nu(\theta)$ will pull in the prior compared to the reference π and δ_ν will provide greater shrinkage than δ_π. This will be further amplified in examples that follow. We now sketch a proof of Proposition 5.

Proof: We will give the proof for the case $V(\pi) = \text{Var}_\pi(\theta) = 1$. A minor modification works for $V(\pi) \neq 1$.

Step 1. From (2.5), we would like to maximize $f - \frac{g}{2} + a + \frac{a^2}{2}$ and hence minimize $g - 2f - 2a - a^2 = \text{Var}_\pi(\frac{\nu'(\theta)}{\nu(\theta)}) - 2\,\text{Cov}_\pi(\theta + \frac{\pi'(\theta)}{\pi(\theta)}, \frac{\nu'(\theta)}{\nu(\theta)}) - \text{Cov}^2_\pi(\theta, \frac{\nu'(\theta)}{\nu(\theta)})$.

Step 2. Using the fact that $V(\pi) = 1$, observe that $\text{Cov}_\pi(\theta, \theta + \frac{\pi'(\theta)}{\pi(\theta)}) = 0$.

Step 3. By using Step 2, write

$$\begin{aligned} &\text{Var}_\pi(\frac{\nu'(\theta)}{\nu(\theta)}) - 2\ \text{Cov}_\pi\left(\theta + \frac{\pi'(\theta)}{\pi(\theta)}, \frac{\nu'(\theta)}{\nu(\theta)}\right) - \text{Cov}^2_\pi\left(\theta, \frac{\nu'(\theta)}{\nu(\theta)}\right) \\ &= \text{Var}_\pi\left\{\frac{\nu'(\theta)}{\nu(\theta)} - \left(\theta + \frac{\pi'(\theta)}{\pi(\theta)}\right)\right\} - \text{Cov}^2_\pi\left\{\theta, \frac{\nu'(\theta)}{\nu(\theta)} - \left(\theta + \frac{\pi'(\theta)}{\pi(\theta)}\right)\right\} \\ &- \text{Var}_\pi\left(\theta + \frac{\pi'(\theta)}{\pi(\theta)}\right) \end{aligned} \tag{2.7}$$

Step 4. Since $\text{Var}_\pi(\theta) = 1$, by Schwartz's inequality,

$$\text{Var}_\pi\left\{\frac{\nu'(\theta)}{\nu(\theta)} - \left(\theta + \frac{\pi'(\theta)}{\pi(\theta)}\right)\right\} - \text{Cov}^2_\pi\left\{\theta, \frac{\nu'(\theta)}{\nu(\theta)} - \left(\theta + \frac{\pi'(\theta)}{\pi(\theta)}\right)\right\} \\ \geq 0,$$

with equality if

$$\frac{\nu'(\theta)}{\nu(\theta)} - \theta - \frac{\pi'(\theta)}{\pi(\theta)} = a + b\theta. \tag{2.8}$$

Step 5. The solutions of the differential equation (2.8) are of the form (2.6).

2.3. Investigation of the Selected Bayes Estimate and Examples:

Suppose the reference prior density is $\pi(\theta) = \frac{1}{2}e^{-|\theta|}$, a good middle ground between sharp and flat priors. Further suppose that the family Γ under consideration contains only symmetric priors and so $\nu(\theta)$ is of the form $\nu(\theta) = ce^{-|\theta|}e^{-\frac{\theta^2}{2\tau^2}}$. There is a value of τ (approximately 7.52) that gives the largest variance among all priors of this form. For our first illustration, this is the specific $\nu(\theta)$ we use.

Example 1. Let us take $X \sim N(\theta, 1)$. Under both $\pi(\theta)$ and $\nu(\theta)$, the marginal density of X can be found in closed form and hence the Bayes estimates $\delta_\pi(X)$ and $\delta_\nu(X)$ can also be found in closed form by using the familiar identity (Brown (1985)):

$$\delta_\nu(x) = x + \frac{\sigma^2}{n}\frac{m'_\nu(x)}{m_\nu(x)},$$

where m_ν is the marginal density. Some selected values are reported below; as commented before, δ_ν results in a bit more shrinkage than δ_π.

X	0	.5	1	1.5	2	3	5	8	10	15
$\delta_\pi(x)$	0	.241	.503	.806	1.161	2.026	4	7	9	14
$\delta_\nu(x)$	0	.238	.497	.795	1.144	1.992	3.931	6.878	8.844	13.758

Example 2. Consider $X \sim \text{Logistic}(\theta, \sigma)$, with known σ and having density

$$f(x|\theta) = \frac{1}{\sigma}\exp(-\frac{(x-\theta)}{\sigma})\left[1+\exp(-\frac{(x-\theta)}{\sigma})\right]^{-2}.$$

In this case it is not possible to obtain in closed form either the marginals or the Bayes estimates, but both $\delta_\pi(x)$ and $\delta_\nu(x)$ can be easily computed for any given x.

Again, some selected values are reported below for two different values of σ, 0.5 and 1. As before, δ_ν receives a little more shrinkage than δ_π. Further, $\sigma = 1$ results in much more shrinkage than $\sigma = 0.5$, with the values for $\sigma = 0.5$ being closer to those in Example 1.

$\sigma = 0.5$

X	0	.5	1	1.5	2	3	5	8	10	15
$\delta_\pi(x)$	0	.276	.580	.922	1.301	2.147	4.030	7.002	9	14
$\delta_\nu(x)$	0	.274	.574	.912	1.285	2.113	3.948	6.834	8.771	13.607

$\sigma = 1.0$

X	0	.5	1	1.5	2	3	5	8	10	15
$\delta_\pi(x)$	0	.166	.336	.513	.698	1.095	1.970	3.386	4.357	6.820
$\delta_\nu(x)$	0	.163	.329	.501	.680	1.059	1.864	3.038	3.722	4.938

From the numerical tables above it seems that in these cases when the reference prior is a double exponential, δ_π and δ_ν behave similarly. For another choice of the reference prior, this need not be the case.

To see this, consider the reference prior density $\pi(\theta) \propto (\frac{1}{1+\theta^2/3})^{-2}$, density of the Student's t_3 prior which is a flat prior. Suppose again that the family Γ under consideration contains only symmetric priors and so $\nu(\theta)$ is of the form $\nu(\theta) = c(\frac{1}{1+\theta^2/3})^{-2}e^{-\frac{\theta^2}{2\tau^2}}$.

Example 3. Now consider $X \sim \text{Cauchy}(\theta, \sigma)$, with known σ and having density

$$f(x|\theta) = \frac{1}{(\sigma\pi)\left(1+(\frac{x-\theta}{\sigma})^2\right)}.$$

Some selected values are reported below for $\sigma = .2$.

X	0	.5	1	1.5	2	3	5	8	10	15
$\delta_\pi(x)$	0	.349	.687	1.002	1.284	1.735	2.197	2.216	2.065	1.645
$\delta_\nu(x)$	0	.348	.683	.993	1.267	1.685	1.976	1.60	1.15	.481

Note that, for small and moderate values of x, δ_π and δ_ν behave similarly, whereas for large values δ_ν results in much more shrinkage than δ_π. But we would expect this because the penultimate ν has normal tails, whereas the reference π has very flat tails.

3. Further Potential for Practical Uses

3.1. Selecting a Default Prior

Extensive literature exists on default prior Bayesian analysis; the literature includes much general theory and methods and applications of these to specific problems. There seems to be widely different opinions regarding appropriate definitions of default priors. We will say that a prior chosen from a specified class by a specified (and hopefully reasonable) selection rule is a default prior. Assessment of such a default prior is a separate issue and we will not address that here. Our intention is to show a potential use of our correlation approach in this important problem. In addition, as we shall show, our proposal has a very distinct connection to the Fisher information. These are interesting consequences of our general approach and the Bickel prior arises as special from this development. Among the literature on default priors, particularly pertinent to our discussion are Cifarelli and Regazzini (1987), Clarke and Wasserman (1993), Ghosh and Mukerjee (1992), Datta and Ghosh (1995), and Kass and Wasserman (1996).

The approach we take is the following. A likelihood based method will summarize the post-data opinion about θ by the likelihood function $f(x|\theta)$; a Bayesian method based on a given prior $\pi(\theta)$ will use the posterior density $\pi(\theta|x)$. Minimizing a suitable distance between these two summaries is a well accepted approach for construction of default priors. We are proposing, instead, maximization of the correlation between $f(x|\theta)$ and $\pi(\theta|x)$ in the joint probability space.

Before we derive the results of this section for the location parameter case, let us see an important case as an illustrative example for our suggested approach. This example will show that the general approach we are suggesting has the potential for producing common sense default priors.

Example 4. Suppose $X \sim$ Bin (n, θ) and we wish to estimate θ. In the literature, various priors have been suggested as default priors for θ, the uniform and the Jeffreys prior included; see Berger (1986). It is well known that if $\theta \sim \mathcal{U}[0,1]$, then, curiously, X

has a marginal uniform distribution too. Thus, $\pi(\theta|x) = (n+1)f(x|\theta)$, for all x and all θ. We therefore have the curious result that the correlation between $f(x|\theta)$ and $\pi(\theta|x)$ in the joint probability space is 1 if θ has a uniform prior. A fortiori, the uniform prior is the default prior for θ according to our criterion just as long as the class of priors entertained includes the uniform prior. Thus, in the important binomial case, our general approach leads to a credible default prior. This is encouraging.

For ease of exposition, we shall only state the main results here. Major steps involved in the proofs will be outlined in the appendix. Proofs of many of these steps are similar, and hence details of only some are given in the appendix. The set-up required in this section is similar, but somewhat weaker than what is stated in Section 1.1. Specifically, it is enough to assume that the likelihood function is continuously differentiable five times with a bounded fifth derivative. With regard to the prior densities also, we can weaken Assumption A, and work with the following Assumption B.

Assumption B. The class Γ of prior densities under consideration consists of prior densities π which are three times continuously differentiable a.e., with bounded third derivative, and $E_\pi(\theta) = 0$, $E_\pi(\theta^2) < \infty$.

3.2. An Expansion for Correlation

It is not possible to derive any analytical results by working with the exact marginal correlation $\rho_P(f(x|\theta), \pi(\theta|x))$. We will present an expansion for the marginal correlation; in this expansion, the leading term is 1, and the second term is $-\frac{a_1\sigma^2}{n} \cdot I(\pi) + \frac{a_2}{n}$, where $I(\pi)$ is the Fisher Information of the prior π, $a_1 > 0$ is an absolute constant and a_2 depends only on f. Therefore, according to our correlation criterion, we propose to maximize this second term as a rule for selecting a default prior. This is formally similar to certain results in Clarke and Barron (1994) and Clarke and Wasserman (1993). As indicated in the Introduction, we thus end up minimizing the Fisher Information of the prior, a well known approach which has been adopted by other authors for different reasons altogether. This connection of our default prior methodology to minimization of Fisher information is interesting.

Proposition 6. Under Assumption B,

$$\rho_P(f(x|\theta), \pi(\theta|x)) = 1 - \frac{\sigma^2}{2n(c_1 - c_3^2)} c_2 I(\pi) - \frac{\sigma^4}{8n(c_1 - c_3^2)} \{c_1 w_2 + (c_1 - c_6) l_4\} + o(n^{-1}), \tag{3.1}$$

where $I(\pi)$ is the Fisher Information functional, and c_1, c_2, c_3, c_4, c_5 and c_6 are the following constants:

$$\left.\begin{array}{l} c_1 = \int_{-\infty}^{\infty} \phi^3(z)dz = \frac{1}{2\pi\sqrt{3}} \\ c_2 = \int_{-\infty}^{\infty} z^2\phi^3(z)dz = \frac{1}{6\pi\sqrt{3}} \\ c_3 = \int_{-\infty}^{\infty} \phi^2(z)dz = \frac{1}{2\sqrt{\pi}} \\ c_4 = \int_{-\infty}^{\infty} z^2\phi^2(z)dz = \frac{1}{4\sqrt{\pi}} \\ c_5 = \int_{-\infty}^{\infty} z^4\phi^2(z)dz = \frac{3}{8\sqrt{\pi}} \\ c_6 = \int_{-\infty}^{\infty} z^4\phi^3(z)dz = \frac{1}{6\sqrt{3}\pi} \end{array}\right\} . \tag{3.2}$$

Remark: The constants c_4 and c_5 appear in the derivation but ultimately disappear. Since $c_1 - c_3^2 = .01231 > 0$, we would want to minimize $I(\pi)$ in appropriate families Γ. A formal derivation of the expansion is outlined in the appendix.

3.3. Illustrations

Consider a general location parameter model (with parameter θ) which fits into our set-up. Then we have the following results.

a. Suppose $|\theta| \leq 1$. Then from Bickel (1981) or Huber (1974), the following prior density achieves the minimum Fisher Information in the class of all priors (i.e., now compactly supported on $[-1, 1]$ since this is the parameter space).

$$\pi(\theta) = \begin{cases} \cos^2(\frac{\pi}{2}\theta), & \text{if } |\theta| \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the Bickel prior is the default prior under our correlation criterion.

b. Instead of constraining the parameter space to $[-1, 1]$, we could obviously also look at this prior as the solution in the unconstrained problem with Γ consisting of only priors which are compactly supported on $[-1, 1]$.

c. Fix $\tau > 0$, and consider

$$\Gamma = \left\{ \pi : \pi \text{ is symmetric about 0 and } \int_{-\infty}^{\infty} \theta^2 \pi(\theta)\, d\theta = \tau^2 \right\}.$$

The prior which achieves the minimum Fisher information in this class is $N(0, \tau^2)$. (See Kagan et. al. (1973).) This class, however, is somewhat restricted since it excludes very heavy tailed priors such as Cauchy and t_2.

d. Now take

$$\Gamma = \left\{ \pi : \pi(\theta) = \int \phi(\theta e^{-\gamma}) \mu(d\gamma), \mu \text{ an arbitrary probability measure on } [-\infty, \infty] \right\}.$$

This is a useful collection of priors in robustness since it is the class of scale mixtures of normal priors, thus including heavy tailed distributions like Student's t (Cauchy being a special case) and double exponential. Unfortunately, it is the case that (see Bickel and Collins, 1983) the infimum Fisher information for this class is 0.

e. Bickel and Collins (1983) then modify this class and consider instead the ϵ-contamination class (in some other context),

$$\Gamma = \left\{ \pi : \pi(\theta) = \epsilon\phi(\theta) + (1-\epsilon) \int_{[-\infty,\infty]} \phi(\theta e^{-\gamma}) \mu(d\gamma), \mu \text{ arbitrary} \right\}.$$

They show that the π which minimizes the Fisher information in this class is given by π^* where

$$\pi^*(\theta) = \epsilon\phi(\theta) + (1-\epsilon) \sum_{i=1}^{\infty} p_i \sigma_i^{-1} \phi(\sigma_i^{-1}\theta),$$

with $0 < \sigma_i < \infty$, $0 < p_i < 1$, $\sum_{i=1}^{\infty} p_i = 1$. However, identifying the p_i and σ_i is a numerically challenging problem.

4. Appendix

Proof of Proposition 1. If $\nu(\theta)$ is any prior density, then the Bayes estimator of θ with respect to this prior is

$$\begin{aligned}
\delta_\nu(x) &= \frac{\int \theta \exp(\mathcal{L}(\theta, x)) \nu(\theta)\, d\theta}{\int \exp(\mathcal{L}(\theta, x)) \nu(\theta)\, d\theta} \\
&= \hat{\theta} + \frac{\int (\theta - \hat{\theta}) \exp(\mathcal{L}(\theta, x)) \nu(\theta)\, d\theta}{\int \exp(\mathcal{L}(\theta, x)) \nu(\theta)\, d\theta} \\
&= \hat{\theta} + R_n(x), \text{ say.}
\end{aligned}$$

Now note that ν is continuously differentiable 4 times, and that the 4th derivative is bounded. Expand $\mathcal{L}$ in a Taylor Series around $\hat{\theta}$, and note that $\mathcal{L}^{(6)}(\theta, x)/n$ is uniformly bounded. Then, using (1.1), letting $z = \sqrt{n}(\theta - \hat{\theta})/\sigma$, and letting ϕ denote the standard normal density, we obtain

$$R_n(x) = \frac{\int (\theta - \hat{\theta}) \exp(\mathcal{L}(\theta, x))\nu(\theta)\, d\theta}{\int \exp(\mathcal{L}(\theta, x))\nu(\theta)\, d\theta}$$

$$= \frac{\int (\theta - \hat{\theta}) \exp(-\frac{n}{2\hat{\sigma}^2}(\theta - \hat{\theta})^2 + (\theta - \hat{\theta})^3 \frac{\mathcal{L}^{(3)}}{3!} + \cdots + (\theta - \hat{\theta})^6 \frac{\mathcal{L}^{(6)}(\theta^*, x)}{6!})\nu(\theta)\, d\theta}{\int \exp(-\frac{n}{2\hat{\sigma}^2}(\theta - \hat{\theta})^2 + (\theta - \hat{\theta})^3 \frac{\mathcal{L}^{(3)}}{3!} + \cdots + (\theta - \hat{\theta})^5 \frac{\mathcal{L}^{(5)}}{5!} + (\theta - \hat{\theta})^6 \frac{\mathcal{L}^{(6)}(\theta^*, x)}{6!})\nu(\theta)\, d\theta}$$

$$= \Big[\int \frac{\sigma}{\sqrt{n}} z \exp(-\frac{z^2}{2}(1 - \frac{\sigma^2}{\sqrt{n}} W_n))$$
$$\times \left\{1 + \frac{\sigma^3}{6n} z^3 \frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{\sigma^4}{24n} z^4 \frac{\mathcal{L}^{(4)}}{n} + \frac{\sigma^5}{120 n^{3/2}} z^5 \frac{\mathcal{L}^{(5)}}{n} + O(n^{-2})\right\}$$
$$\times \left\{\nu(\hat{\theta}) + \frac{\sigma}{\sqrt{n}} z \nu'(\hat{\theta}) + \frac{\sigma^2}{2n} z^2 \nu''(\hat{\theta}) + \frac{\sigma^3}{6n^{3/2}} z^3 \nu^{(3)}(\hat{\theta}) + O(n^{-2})\right\} dz\Big]$$
$$/\Big[\int \exp(-\frac{z^2}{2}(1 - \frac{\sigma^2}{\sqrt{n}} W_n)) \left\{1 + \frac{\sigma^3}{6n} z^3 \frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{\sigma^4}{24n} z^4 \frac{\mathcal{L}^{(4)}}{n} + \frac{\sigma^5}{120 n^{3/2}} z^5 \frac{\mathcal{L}^{(5)}}{n} + O(n^{-2})\right\}$$
$$\times \left\{\nu(\hat{\theta}) + \frac{\sigma}{\sqrt{n}} z \nu'(\hat{\theta}) + \frac{\sigma^2}{2n} z^2 \nu''(\hat{\theta}) + \frac{\sigma^3}{6n^{3/2}} z^3 \nu^{(3)}(\hat{\theta}) + O(n^{-2})\right\} dz\Big]$$
$$= \frac{\sigma}{\sqrt{n}} \Big[\int z\phi(z) \left\{1 + \frac{\sigma^2}{2\sqrt{n}} W_n z^2 + \frac{\sigma^4}{8n} W_n^2 z^4 + \frac{\sigma^6}{48 n^{3/2}} W_n^3 z^6 + O(n^{-2})\right\}$$
$$\times \left\{1 + \frac{\sigma^3}{6n} z^3 \frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{\sigma^4}{24n} z^4 \mathcal{L}^{(4)} + \frac{\sigma^5}{120 n^{3/2}} z^5 \frac{\mathcal{L}^{(5)}}{n} + O(n^{-2})\right\}$$
$$\times \left\{\nu(\hat{\theta}) + \frac{\sigma}{\sqrt{n}} z \nu'(\hat{\theta}) + \frac{\sigma^2}{2n} z^2 \nu''(\hat{\theta}) + \frac{\sigma^3}{6n^{3/2}} z^3 \nu^{(3)}(\hat{\theta}) + O(n^{-2})\right\} dz\Big]$$
$$/\Big[\int \phi(z) \left\{1 + \frac{\sigma^2}{2\sqrt{n}} W_n z^2 + \frac{\sigma^4}{8n} W_n^2 z^4 + \frac{\sigma^6}{48 n^{3/2}} W_n^3 z^6 + O(n^{-2})\right\}$$
$$\times \left\{1 + \frac{\sigma^3}{6n} z^3 \frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{\sigma^4}{24n^2} z^4 \mathcal{L}^{(4)} + \frac{\sigma^5}{120 n^{3/2}} z^5 \frac{\mathcal{L}^{(5)}}{n} + O(n^{-2})\right\}$$
$$\times \left\{\nu(\hat{\theta}) + \frac{\sigma}{\sqrt{n}} z \nu'(\hat{\theta}) + \frac{\sigma^2}{2n} z^2 \nu''(\hat{\theta}) \frac{\sigma^3}{6n^{3/2}} z^3 \nu^{(3)}(\hat{\theta}) + O(n^{-2})\right\} dz\Big]$$
$$= \frac{\sigma}{\sqrt{n}} \frac{Num(x)}{Denom(x)},$$

Num and *Denom* denoting the numerator and denominator respectively. These can be evaluated by the expressions,

$$
\begin{aligned}
Num(x) = \int \phi(z)\Big\{ & z\nu(\hat\theta) + \frac{\sigma}{\sqrt{n}}\left[z^2\nu^{'}(\hat\theta) + z^3\frac{\sigma\nu(\hat\theta)W_n}{2}\right] \\
&+ \frac{\sigma^2}{n}\left[z^3\frac{\nu^{''}(\hat\theta)}{2} + z^4\frac{\sigma}{2}\big(\frac{\nu(\hat\theta)}{3}\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \nu^{'}(\hat\theta)W_n\big) + z^5\frac{\sigma^2\nu(\hat\theta)}{8}\big(\frac{\mathcal{L}^{(4)}}{3n} + W_n^2\big)\right] \\
&+ \frac{\sigma^3}{n^{3/2}}\Big[z^4\frac{\nu^{'''}(\hat\theta)}{6} + z^5\sigma\big(\frac{\nu^{''}(\hat\theta)}{4}W_n + \frac{\nu^{'}(\hat\theta)}{6}\frac{\mathcal{L}^{(3)}}{\sqrt{n}}\big) \\
&+ z^6\frac{\sigma^2}{4}\big(\frac{\nu^{'}(\hat\theta)W_n^2}{2} + \frac{\nu^{'}(\hat\theta)}{6}\frac{\mathcal{L}^{(4)}}{n} + \frac{\nu(\hat\theta)}{30}\frac{\mathcal{L}^{(5)}}{n}\big) + \frac{\nu(\hat\theta)}{3}\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\big) \\
&+ z^7\nu(\hat\theta)\frac{\sigma^3}{48}W_n\big(W_n^2 + \frac{\mathcal{L}^{(4)}}{n}\big)\Big] + O(n^{-2})\Big\}\, dz \\
= \frac{\sigma}{\sqrt{n}}\Big[& \nu^{'}(\hat\theta) + \frac{\sigma^2}{2\sqrt{n}}\left(\nu(\hat\theta)\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 3\nu^{'}(\hat\theta)W_n\right) \\
&+ \frac{\sigma^2}{n}\left(\frac{\nu^{'''}(\hat\theta)}{2} + \frac{15\sigma^2}{8}\nu^{'}(\hat\theta)\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right) + \frac{\sigma^2}{4}\nu(\hat\theta)\left(\frac{\mathcal{L}^{(5)}}{2n} + 5\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\right)\Big] + O(n^{-3/2}),
\end{aligned}
$$

$$
\begin{aligned}
Denom(x) = \int \phi(z)\Big\{ & \nu(\hat\theta) + \frac{\sigma}{\sqrt{n}}\left[z\nu^{'}(\hat\theta) + z^2\frac{\sigma\nu(\hat\theta)W_n}{2}\right] \\
&+ \frac{\sigma^2}{n}\left[z^2\frac{\nu^{''}(\hat\theta)}{2} + z^3\big(\frac{\sigma\nu(\hat\theta)}{6}\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{\sigma\nu^{'}(\hat\theta)W_n}{2}\big) + z^4\frac{\sigma^2\nu(\hat\theta)}{8}\big(\frac{\mathcal{L}^{(4)}}{3n} + W_n^2\big)\right] \\
&+ \frac{\sigma^3}{n^{3/2}}\Big[z^3\frac{\nu^{'''}(\hat\theta)}{6} + z^4\sigma\big(\frac{\nu^{''}(\hat\theta)}{4}W_n + \frac{\nu^{'}(\hat\theta)}{6}\frac{\mathcal{L}^{(3)}}{\sqrt{n}} \\
&+ z^5\frac{\sigma^2}{4}\big(\nu^{'}(\hat\theta)W_n^2 + \frac{\nu^{'}(\hat\theta)}{6}\frac{\mathcal{L}^{(4)}}{n} + \frac{\nu(\hat\theta)}{30}\frac{\mathcal{L}^{(5)}}{n}\big) + \frac{\nu(\hat\theta)}{3}\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n \\
&+ z^6\nu(\hat\theta)\frac{\sigma^3}{48}W_n\big(W_n^2 + \frac{\mathcal{L}^{(4)}}{n}\big)\Big] + O(n^{-2})\Big\}\, dz \\
= \nu(\hat\theta) & + \frac{\sigma^2}{2\sqrt{n}}\nu(\hat\theta)W_n + \frac{\sigma^2}{n}\left(\frac{\nu^{''}(\hat\theta)}{2} + \frac{3\sigma^2\nu(\hat\theta)}{8}\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right)\right) + O(n^{-3/2}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
R_n(x) =& \frac{\sigma^2}{n}\Big\{\nu'(\hat\theta) + \frac{\sigma^2}{\sqrt{n}}\left(\frac{\nu(\hat\theta)}{2}\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{3\nu'(\hat\theta)W_n}{2}\right) + \frac{\sigma^2}{n}\Big(\frac{\nu'''(\hat\theta)}{2} + \frac{15\nu'(\hat\theta)\sigma^2}{8}\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right) \\
&+ \frac{\nu(\hat\theta)\sigma^2}{4}\left(\frac{\mathcal{L}^{(5)}}{2n} + 5\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\Big) + O(n^{-3/2})\Big\} \\
&\times \left\{\nu(\hat\theta) + \frac{\sigma^2}{2\sqrt{n}}\nu(\hat\theta)W_n + \frac{\sigma^2}{2n}\left(\nu''(\hat\theta) + \frac{3\sigma^2\nu(\hat\theta)}{4}\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right)\right) + O(n^{-3/2})\right\}^{-1} \\
=& \frac{\sigma^2}{n\nu(\hat\theta)}\Big\{\nu'(\hat\theta) + \frac{\sigma^2}{2\sqrt{n}}\left(\nu(\hat\theta)\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 3\nu'(\hat\theta)W_n\right) + \frac{\sigma^2}{n}\Big(\frac{\nu'''(\hat\theta)}{2} + \frac{15\nu'(\hat\theta)\sigma^2}{8}\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right) \\
&+ \frac{\nu(\hat\theta)\sigma^2}{4}\left(\frac{\mathcal{L}^{(5)}}{2n} + 5\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\Big)\Big\} \\
&\times \left[1 + \frac{\sigma^2}{2\sqrt{n}}W_n + \frac{\sigma^2}{2n}\left(\frac{\nu''(\hat\theta)}{\nu(\hat\theta)} + \frac{3\sigma^2}{4}\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right)\right)\right]^{-1} + O(n^{-5/2}) \\
=& \frac{\sigma^2}{n}\Big\{\frac{\nu'(\hat\theta)}{\nu(\hat\theta)} + \frac{\sigma^2}{2\sqrt{n}}\left(\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 3\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}W_n\right) + \frac{\sigma^2}{n}\Big(\frac{\nu'''(\hat\theta)}{2\nu(\hat\theta)} + \frac{15\sigma^2}{8}\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right) \\
&+ \frac{\sigma^2}{4}\left(\frac{\mathcal{L}^{(5)}}{2n} + 5\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\Big)\Big\} \\
&\times \left\{1 - \frac{\sigma^2}{2\sqrt{n}}W_n - \frac{\sigma^2}{2n}\left(\frac{\nu''(\hat\theta)}{\nu(\hat\theta)} + \frac{3\sigma^2}{4}W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right) + \frac{\sigma^4}{4n}W_n^2\right\} + O(n^{-5/2}) \\
=& \frac{\sigma^2}{n}\frac{\nu'(\hat\theta)}{\nu(\hat\theta)} + \frac{\sigma^4}{2n^{3/2}}\left(\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 2\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}W_n\right) \\
&+ \frac{\sigma^4}{n^2}\left(\frac{\nu'''(\hat\theta)}{2\nu(\hat\theta)} + \frac{15\sigma^2}{8}\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right) + \frac{\sigma^2}{4}\left(\frac{\mathcal{L}^{(5)}}{2n} + 5\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\right) \\
&- \frac{\sigma^6}{4n^2}W_n\left(\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 3\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}W_n\right) - \frac{\sigma^4}{2n^2}\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}\left(\frac{\nu''(\hat\theta)}{\nu(\hat\theta)} + \frac{3\sigma^2}{4}\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right)\right) \\
&+ \frac{\sigma^6}{4n^2}\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}W_n^2 + O(n^{-5/2}) \\
=& \frac{\sigma^2}{n}\frac{\nu'(\hat\theta)}{\nu(\hat\theta)} + \frac{\sigma^4}{2n^{3/2}}\left(\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 2\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}W_n\right) + \frac{\sigma^4}{2n^2}\Big(\frac{\nu'''(\hat\theta)}{\nu(\hat\theta)} - \frac{\nu'(\hat\theta)\nu''(\hat\theta)}{\nu^2(\hat\theta)} \\
&+ \frac{\nu'(\hat\theta)}{\nu(\hat\theta)}\sigma^2\left(2W_n^2 + \frac{\mathcal{L}^{(4)}}{n}\right) + \sigma^2\left(\frac{\mathcal{L}^{(5)}}{4n} + 2\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\Big) + O(n^{-5/2}).
\end{aligned}
$$

This yields the desired approximation for δ_ν.

Proof of Proposition 2. Note that

$$
\begin{aligned}
&\mathbf{Cov}_P(\theta, \delta_\nu) \\
&= \int\int \theta\delta_\nu(x)\exp(\mathcal{L}(\theta,x))\pi(\theta)\,d\theta\,dx \\
&= \int\int \theta\Big\{\hat\theta + \frac{\sigma^2}{n}\frac{\nu'(\hat\theta)}{\nu(\hat\theta)} + \frac{\sigma^4}{2n^{3/2}}\left(\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 2\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}W_n\right) + \frac{\sigma^4}{2n^2}\Big(\frac{\nu'''(\hat\theta)}{\nu(\hat\theta)} - \frac{\nu'(\hat\theta)\nu''(\hat\theta)}{\nu^2(\hat\theta)} \\
&+ \frac{\nu'(\hat\theta)}{\nu(\hat\theta)}\sigma^2\left(2W_n^2 + \frac{\mathcal{L}^{(4)}}{n}\right) + \sigma^2\left(\frac{\mathcal{L}^{(5)}}{4n} + 2\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\Big)\Big\}\exp(\mathcal{L}(\theta,x))\pi(\theta)\,d\theta\,dx + O(n^{-5/2}),
\end{aligned}
$$

where all the $\mathcal{L}^{(i)}$ above are actually $\mathcal{L}^{(i)}(\hat\theta, x)$. Proceeding as in the derivation of R_n above, and assuming that π is also 5 times continuously differentiable and that $\pi^{(5)}$ is bounded, we get,

$$
\begin{aligned}
&\mathbf{Cov}_P(\theta, \delta_\nu) \\
&= \frac{\sigma}{\sqrt{n}}\int\int(\hat\theta + \frac{\sigma}{\sqrt{n}}z)\Big\{\hat\theta + \frac{\sigma^2}{n}\frac{\nu'(\hat\theta)}{\nu(\hat\theta)} + \frac{\sigma^4}{2n^{3/2}}\left(\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 2\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}W_n\right) \\
&+ \frac{\sigma^4}{2n^2}\Big(\frac{\nu'''(\hat\theta)}{\nu(\hat\theta)} - \frac{\nu'(\hat\theta)\nu''(\hat\theta)}{\nu^2(\hat\theta)} + \frac{\nu'(\hat\theta)}{\nu(\hat\theta)}\sigma^2\left(2W_n^2 + \frac{\mathcal{L}^{(4)}}{n}\right)\Big) + \sigma^2\left(\frac{\mathcal{L}^{(5)}}{4n} + 2\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\Big\} \\
&\qquad \times \exp(\mathcal{L}(\hat\theta,x))\exp(-\frac{z^2}{2}(1 - \frac{\sigma^2}{\sqrt{n}}W_n)) \\
&\times\left\{1 + \frac{\sigma^3}{6n}z^3\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{\sigma^4}{24n}z^4\frac{\mathcal{L}^{(4)}}{n} + \frac{\sigma^5}{120n^{3/2}}z^5\frac{\mathcal{L}^{(5)}}{n} + \frac{\sigma^6}{720n^2}z^6\frac{\mathcal{L}^{(6)}}{n}\right\} \\
&\times\left\{\pi(\hat\theta) + \frac{\sigma}{\sqrt{n}}z\pi'(\hat\theta) + \frac{\sigma^2}{2n}z^2\pi''(\hat\theta) + \frac{\sigma^3}{6n^{3/2}}z^3\pi^{(3)}(\hat\theta) + \frac{\sigma^4}{24n^2}z^4\pi^{(4)}(\hat\theta)\right\}dz\,dx + O(n^{-5/2}) \\
&= \frac{\sigma}{\sqrt{n}}\sqrt{2\pi}\int\exp(\mathcal{L}(\hat\theta,x))\Big\{\hat\theta + \frac{\sigma^2}{n}\frac{\nu'(\hat\theta)}{\nu(\hat\theta)} + \frac{\sigma^4}{2n^{3/2}}\left(\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 2\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}W_n\right) \\
&\qquad + \frac{\sigma^4}{2n^2}\Big(\frac{\nu'''(\hat\theta)}{\nu(\hat\theta)} - \frac{\nu'(\hat\theta)\nu''(\hat\theta)}{\nu^2(\hat\theta)} + \frac{\nu'(\hat\theta)}{\nu(\hat\theta)}\sigma^2\left(2W_n^2 + \frac{\mathcal{L}^{(4)}}{n}\right)\Big) + \sigma^2\left(\frac{\mathcal{L}^{(5)}}{4n} + 2\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\Big\}
\end{aligned}
$$

$$\Big(\int \phi(z)\Big\{\hat\theta\pi(\hat\theta) + \frac{\sigma}{\sqrt{n}}(z(\pi(\hat\theta)+\hat\theta\pi(\hat\theta)) + z^2\frac{\sigma}{2}W_n\hat\theta\pi(\hat\theta)) + \frac{\sigma^2}{n}(z^2(\frac{\hat\theta\pi''(\hat\theta)}{2}+\pi'(\hat\theta))$$

$$+ z^3\frac{\sigma}{2}(\frac{\hat\theta\pi(\hat\theta)}{3}\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \hat\theta\pi'(\hat\theta))W_n + \pi(\hat\theta))W_n) + z^4\frac{\sigma^2}{8}\hat\theta\pi(\hat\theta)(W_n^2+\frac{\mathcal{L}^{(4)}}{3n})$$

$$+ \frac{\sigma^3}{2n^{3/2}}(z^3(\pi''(\hat\theta) + \frac{\hat\theta\pi'''(\hat\theta)}{3} + z^4\sigma((\pi(\hat\theta)+\hat\theta\pi'(\hat\theta))\frac{\mathcal{L}^{(3)}}{3\sqrt{n}} + W_n(\pi'(\hat\theta)+\frac{\hat\theta\pi''(\hat\theta)}{2})$$

$$+ z^5\frac{\sigma^2}{4}(W_n^2(\pi(\hat\theta)+\hat\theta\pi'(\hat\theta)) + \frac{\hat\theta\pi(\hat\theta)}{3}(\frac{\mathcal{L}^{(5)}}{5n} + \frac{2\mathcal{L}^{(3)}}{\sqrt{n}}W_n) + \frac{\mathcal{L}^{(4)}}{3n}(\pi(\hat\theta)+\hat\theta\pi'(\hat\theta)))$$

$$+ z^6\frac{\sigma^3}{24}\hat\theta\pi(\hat\theta)W_n(W_n^2+\frac{\mathcal{L}^{(4)}}{n})) + \frac{\sigma^4}{2n^2}(z^3\frac{\pi'''}{3} + z^4\frac{\hat\theta\pi^{(4)}(\hat\theta)}{12}$$

$$+ z^5\sigma(\frac{\mathcal{L}^{(3)}}{3\sqrt{n}}(\pi'(\hat\theta)+\frac{\hat\theta\pi''(\hat\theta)}{2}) + \frac{W_n}{2}(\frac{\hat\theta\pi'''(\hat\theta)}{3}+\pi''(\hat\theta)) + \frac{\sigma}{60}\pi(\hat\theta)\frac{\mathcal{L}^{(5)}}{n})$$

$$+ z^6\frac{\sigma^2}{2}(W_n^2(\pi'(\hat\theta)+\frac{\hat\theta\pi''(\hat\theta)}{2}) + W_n\frac{\mathcal{L}^{(3)}}{3\sqrt{n}}(\pi(\hat\theta)+\hat\theta\pi'(\hat\theta)) + \frac{\mathcal{L}^{(4)}}{6n}(\pi'(\hat\theta)+\frac{\hat\theta\pi''(\hat\theta)}{2})$$

$$+ \frac{\hat\theta}{30}(\pi'(\hat\theta)\frac{\mathcal{L}^{(5)}}{n} + \pi(\hat\theta)\frac{\mathcal{L}^{(6)}}{6n}))$$

$$+ z^7\frac{\sigma^2}{24}(+W_n^3(\pi(\hat\theta)+\hat\theta\pi'(\hat\theta)) + \sigma\hat\theta W_n(\pi(\hat\theta)\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n + \pi(\hat\theta)\frac{\mathcal{L}^{(5)}}{5n} + \pi'(\hat\theta)\frac{\mathcal{L}^{(4)}}{n})$$

$$+ z^8\frac{\sigma^4}{192}\hat\theta\pi(\hat\theta)W_n^2(\frac{W_n^2}{2}+\frac{\mathcal{L}^{(4)}}{n}))\Big\}\,dz\Big)\,dx + O(n^{-5/2})$$

$$= \frac{\sigma}{\sqrt{n}}\sqrt{2\pi}\int \exp(\mathcal{L}(\hat\theta,x))\Big\{\hat\theta + \frac{\sigma^2}{n}\frac{\nu'(\hat\theta)}{\nu(\hat\theta)} + \frac{\sigma^4}{2n^{3/2}}\left(\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 2\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}W_n\right)$$

$$+ \frac{\sigma^4}{2n^2}\Big(\frac{\nu'''(\hat\theta)}{\nu(\hat\theta)} - \frac{\nu'(\hat\theta)\nu''(\hat\theta)}{\nu^2(\hat\theta)} + \frac{\nu'(\hat\theta)}{\nu(\hat\theta)}\sigma^2\left(2W_n^2+\frac{\mathcal{L}^{(4)}}{n}\right)\Big) + \sigma^2\left(\frac{\mathcal{L}^{(5)}}{4n} + 2\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\Big\}$$

$$\Big(\hat\theta\pi(\hat\theta) + \frac{\sigma^2}{2\sqrt{n}}\hat\theta\pi(\hat\theta)W_n + \frac{\sigma^2}{n}(\frac{\hat\theta\pi''(\hat\theta)}{2} + \pi'(\hat\theta) + \frac{\sigma^2}{8}\hat\theta\pi(\hat\theta)(3W_n^2+\frac{\mathcal{L}^{(4)}}{n}))$$

$$+ \frac{\sigma^4}{2n^{3/2}}(3((\pi(\hat\theta)+\hat\theta\pi'(\hat\theta))\frac{\mathcal{L}^{(3)}}{3\sqrt{n}} + W_n(\pi'(\hat\theta)+\frac{\hat\theta\pi''(\hat\theta)}{2}) + 15\frac{\sigma^2}{24}\hat\theta\pi(\hat\theta)W_n(W_n^2+\frac{\mathcal{L}^{(4)}}{n}))$$

$$+ \frac{\sigma^4}{8n^2}(\hat\theta\pi^{(4)}(\hat\theta) + 15\sigma^2(W_n^2(\pi'(\hat\theta)+\hat\theta\pi''(\hat\theta)) + 2W_n\frac{\mathcal{L}^{(3)}}{3\sqrt{n}}(\pi(\hat\theta)+\hat\theta\pi'(\hat\theta))$$

$$+ \frac{\mathcal{L}^{(4)}}{3n}(\pi'(\hat\theta)+\frac{\hat\theta\pi''(\hat\theta)}{2}) + \frac{\hat\theta}{15}(\pi'(\hat\theta)\frac{\mathcal{L}^{(5)}}{n} + \pi(\hat\theta)\frac{\mathcal{L}^{(6)}}{6n}))$$

$$+ 105\frac{\sigma^4}{48}\hat\theta\pi(\hat\theta)W_n^2(W_n^2+2\frac{\mathcal{L}^{(4)}}{n}))\Big)\,dx + O(n^{-5/2})$$

$$
\begin{aligned}
&= \frac{\sigma}{\sqrt{n}}\sqrt{2\pi}\int \exp(\mathcal{L}(\hat\theta,x))\Big\{\hat\theta^2\pi(\hat\theta) + \frac{\sigma^2}{2\sqrt{n}}\hat\theta^2\pi(\hat\theta)W_n \\
&\quad + \frac{\sigma^2}{n}\left(\hat\theta\pi(\hat\theta)\frac{\nu'(\hat\theta)}{\nu(\hat\theta)} + \frac{\hat\theta^2\pi''(\hat\theta)}{2} + \hat\theta\pi'(\hat\theta) + \frac{\sigma^2}{8}\hat\theta^2\pi(\hat\theta)(3W_n^2 + \frac{\mathcal{L}^{(4)}}{n})\right) \\
&\quad + \frac{\sigma^4}{2n^{3/2}}\Big(\hat\theta\pi'(\hat\theta)\left(2\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 3\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}W_n\right) + \hat\theta^2\pi'(\hat\theta)\frac{\mathcal{L}^{(3)}}{\sqrt{n}} \\
&\quad + \hat\theta^2\pi(\hat\theta)\frac{5\sigma^2}{8}W_n(W_n^2 + \frac{\mathcal{L}^{(4)}}{n}) + 3\sigma\hat\theta(\pi'(\hat\theta) + \frac{\hat\theta\pi''(\hat\theta)}{2})W_n\Big) \\
&\quad + \frac{\sigma^4}{2n^2}\Big(\frac{\hat\theta^2\pi^{(4)}(\hat\theta)}{4} + \hat\theta\pi(\hat\theta)\left(\frac{\nu'''(\hat\theta)}{\nu(\hat\theta)} - \frac{\nu'(\hat\theta)\nu''(\hat\theta)}{\nu^2(\hat\theta)}\right) + \hat\theta\pi''(\hat\theta)\frac{\nu'(\hat\theta)}{\nu(\hat\theta)} + 2\pi'(\hat\theta)\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}\Big) \\
&\quad + \frac{\sigma^6}{4n^2}\Big(\hat\theta\pi(\hat\theta)\left(\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}(\frac{13}{2}W_n^2 + 5\frac{\mathcal{L}^{(4)}}{2n}) + (\frac{\mathcal{L}^{(5)}}{2n} + 5W_n\frac{\mathcal{L}^{(3)}}{\sqrt{n}})\right) \\
&\quad + \hat\theta^2\pi(\hat\theta)\left(\frac{105\sigma^2}{96}W_n^2(W_n^2 + 2\frac{\mathcal{L}^{(4)}}{n}) + \frac{\mathcal{L}^{(6)}}{12n}\right) + \hat\theta^2\pi'(\hat\theta)(\frac{\mathcal{L}^{(5)}}{30n} + 5W_n\frac{\mathcal{L}^{(3)}}{\sqrt{n}}) \\
&\quad + \hat\theta^2\pi''(\hat\theta)\frac{15}{2}(W_n^2 + \frac{\mathcal{L}^{(4)}}{6n}) + \hat\theta\pi(\hat\theta)5W_n\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \hat\theta\pi'(\hat\theta)\frac{15}{2}\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right)\Big)\,dx \\
&\qquad + O(n^{-5/2}) \\
&= V(\pi) + \frac{\sigma^2}{n}a + \frac{\sigma^4}{8n}(3w_2 + l_4) + \frac{\sigma^4}{2n^2}(b - c + d + 2f) + \frac{\sigma^6}{8n^2}(15w_2 + 5l_4)a \\
&\qquad + \frac{\sigma^6}{4n^2}k(\pi) + O(n^{-5/2}),
\end{aligned}
$$

where $k(\pi)$ is some constant which depends on π but not on ν.

Also, the final step involving integration (w.r.t. x) in the derivation of $\mathbf{Cov}_P(\theta, \delta_\nu)$ is done as follows. Consider a typical term that needs to be integrated. The integration is of the form

$$
\frac{\sigma}{\sqrt{n}}\sqrt{2\pi}\int \exp(\mathcal{L}(\hat\theta,x))p_1(\hat\theta)p_2\left(\mathcal{L}^{(2)}(\hat\theta,x), \mathcal{L}^{(3)}(\hat\theta,x), \mathcal{L}^{(4)}(\hat\theta,x), \mathcal{L}^{(5)}(\hat\theta,x))\right)\,dx
$$

for some p_1 and p_2. Note that $\mathcal{L}(\hat\theta,x) = \sum_{i=1}^n h(x_i - \hat\theta)$ and $\mathcal{L}^{(i)}(\hat\theta,x) = \sum_{i=1}^n h^{(i)}(x_i - \hat\theta)$, $2 \le i \le 5$, and hence the integrand is a function of $\hat\theta$ and $x - \hat\theta 1$. We now assume that there exists a one-one mapping of x with $(x_1 - x_n, \ldots, x_{n-1} - x_n, \hat\theta)$. (This is clearly the case if the likelihood equation, $\mathcal{L}'(\theta, x) = 0$, has a unique root for all x.) It can be checked that the Jacobian depends only on $(x_1 - x_n, \dots, x_{n-1} - x_n)$. Integrating with respect to $\hat{\theta}$ and then with respect to $(x_1 - x_n, \dots, x_{n-1} - x_n)$ yields,

$$
\begin{aligned}
&\frac{\sigma}{\sqrt{n}}\sqrt{2\pi}\int \exp(\mathcal{L}(\hat{\theta}, x))p_1(\hat{\theta})p_2\left(\mathcal{L}^{(2)}(\hat{\theta}, x), \mathcal{L}^{(3)}(\hat{\theta}, x), \mathcal{L}^{(4)}(\hat{\theta}, x), \mathcal{L}^{(5)}(\hat{\theta}, x)\right)\right)\, dx \\
&= \int p_2\left(a_2(u), a_3(u), a_4(u), a_5(u)\right)\right) \left(\int p_1(\hat{\theta})\, d\hat{\theta}\right) q(u)\, du,
\end{aligned}
$$

for some a_i's and where q can be recognized as the marginal density of $(X_1 - X_n, \dots, X_{n-1} - X_n)$.

Proof of Proposition 4.

$$
\begin{aligned}
\rho_P(\theta, \delta_\nu) = &\Big[V(\pi) + \frac{\sigma^2}{n}a + \frac{\sigma^4}{8n}V(\pi)(3w_2 + l_4) + \frac{\sigma^4}{2n^2}(b - c + d + 2f) \\
&+ \frac{5\sigma^6}{8n^2}(3w_2 + l_4)a + \frac{\sigma^6}{4n^2}k(\pi)\Big] \\
\times &\Big[V(\pi)\Big\{V(\pi) + 2\frac{\sigma^2}{n}a + \frac{\sigma^2}{n} + \frac{\sigma^4}{8n}V(\pi)(3w_2 + l_4) + \frac{\sigma^4}{n^2}(b - c + d + g) \\
&+ \frac{5\sigma^6}{4n^2}(3w_2 + l_4)a + \frac{\sigma^6}{4n^2}k_1(\pi)\Big\}\Big]^{-1/2} + O(n^{-5/2})
\end{aligned}
$$

$$
\begin{aligned}
= &\Big[1 + \frac{1}{V(\pi)}\Big\{\frac{\sigma^2}{n}\left(a + \frac{\sigma^2}{8}V(\pi)(3w_2 + l_4)\right) \\
&+ \frac{\sigma^4}{2n^2}\left(b - c + d + 2f + \frac{5\sigma^2}{4}(3w_2 + l_4)a\right) + \frac{\sigma^6}{4n^2}k(\pi)\Big\}\Big] \\
\times &\Big[1 - \frac{1}{2V(\pi)}\Big\{\frac{\sigma^2}{n}\left(1 + 2a + \frac{\sigma^2}{8}V(\pi)(3w_2 + l_4)\right) \\
&+ \frac{\sigma^4}{n^2}\left(b - c + d + g + \frac{5\sigma^2}{4}(3w_2 + l_4)a\right) + \frac{\sigma^6}{4n^2}k_1(\pi)\Big\} \\
&+ \frac{3}{8V^2(\pi)}\Big\{\frac{\sigma^2}{n}\left(1 + 2a + \frac{\sigma^4}{8}V(\pi)(3w_2 + l_4)\right) \\
&+ \frac{\sigma^4}{n^2}\left(b - c + d + g + \frac{5\sigma^2}{4}(3w_2 + l_4)a\right) + \frac{\sigma^6}{4n^2}k_1(\pi)\Big\}^2 + O(n^{-5/2}) \\
= &\,1 + \frac{\sigma^2}{2n}\left(\frac{\sigma^2}{8}(3w_2 + l_4) - \frac{1}{V(\pi)}\right) + \frac{\sigma^4}{n^2V(\pi)}\left(f - \frac{g}{2} + \frac{1}{V(\pi)}(a + \frac{a^2}{2})\right) \\
&+ \frac{\sigma^4}{n^2V^2(\pi)}\left(\frac{3}{8} + \frac{\sigma^2}{8}(3w_2 + l_4)(2 - \frac{\sigma^2}{8}(3w_2 + l_4)) + K(\pi)\right) + O(n^{-5/2}).
\end{aligned}
$$

Next, we give an outline of the derivation of expansion (3.1).

Step 1. The marginal correlation $\rho_P(f(x|\theta), \pi(\theta|x))$ equals

$$\rho_P(f(x|\theta), \pi(\theta|x)) = \frac{E_P(f(x|\theta)\pi(\theta|x)) - E_P(f(x|\theta))E_P(\pi(\theta|x))}{\sqrt{\{E_P(f^2(x|\theta)) - E_P^2(f(x|\theta))\}\{E_P(\pi^2(\theta|x)) - E_P^2(\pi(\theta|x))\}}} \qquad (4.1)$$

Step 2. A derivation similar to that of $\mathrm{Cov}_P(\theta, \delta_\nu)$ yields,

$$E_P(f(x|\theta)) = \int\int f^2(x|\theta)\pi(\theta)d\theta dx$$

$$= \int\int \exp(2\mathcal{L}(\theta, x))\pi(\theta)d\theta dx$$

$$= \int\int \frac{\sigma}{\sqrt{n}} \exp(2[\mathcal{L}(\hat{\theta}, x) - \frac{z^2}{2}(1 - \frac{\sigma^2}{\sqrt{n}}W_n)])$$

$$\times \left\{1 + \frac{2\sigma^3}{6n}z^3\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{2\sigma^4}{24n}z^4\frac{\mathcal{L}^{(4)}}{n} + O(n^{-3/2})\right\}$$

$$\times \left\{\pi(\hat{\theta}) + \frac{\sigma}{\sqrt{n}}z\pi'(\hat{\theta}) + \frac{\sigma^2}{2n}z^2\pi''(\hat{\theta}) + +O(n^{-3/2})\right\} dz\, dx$$

$$= \int\int \frac{\sigma}{\sqrt{n}} \exp(2\mathcal{L}(\hat{\theta}, x) - z^2) \left\{1 + \frac{\sigma^2}{\sqrt{n}}W_n z^2 + \frac{\sigma^4}{2n}W_n^2 z^4 + O(n^{-3/2})\right\}$$

$$\times \left\{1 + \frac{2\sigma^3}{6n}z^3\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{2\sigma^4}{24n}z^4\frac{\mathcal{L}^{(4)}}{n} + O(n^{-3/2})\right\}$$

$$\times \left\{\pi(\hat{\theta}) + \frac{\sigma}{\sqrt{n}}z\pi'(\hat{\theta}) + \frac{\sigma^2}{2n}z^2\pi''(\hat{\theta}) + +O(n^{-3/2})\right\} dz\, dx$$

$$= \int\int \frac{\sigma}{\sqrt{n}} \exp(2\mathcal{L}(\hat{\theta}, x) - z^2)\Bigg\{1 + \frac{1}{\sqrt{n}}\left(z^2\sigma^2 W_n\pi(\hat{\theta}) + z\sigma\pi'(\hat{\theta})\right)$$

$$+ \frac{1}{n}\left(z^4\frac{\sigma^4}{2}W_n^2\pi(\hat{\theta}) + z^3\sigma^3 W_n\pi'(\hat{\theta}) + z^3\frac{2\sigma^3}{6}\frac{\mathcal{L}^{(3)}}{\sqrt{n}}\pi(\hat{\theta}) + z^4\frac{2\sigma^4}{24}\frac{\mathcal{L}^{(4)}}{n}\pi(\hat{\theta}) + z^2\frac{\sigma^2}{2}\pi''(\hat{\theta})\right)$$

$$+ O(n^{-3/2})\Bigg\} dz\, dx$$

$$= \frac{\sqrt{n}}{\sigma}\int \phi^2(z)\left\{1 + \frac{\sigma^2}{2n}\left(z^4\sigma^2(w_2 + \frac{l_4}{6}) + +z^2\int \pi''(u)\, du\right) + o(n^{-1})\right\}$$

$$= \frac{\sqrt{n}}{\sigma}\left\{c_3 + \frac{\sigma^2}{2n}\left(c_5\sigma^2(w_2 + \frac{l_4}{6}) + c_4\int \pi''(u)\, du\right) + o(n^{-1})\right\}$$

$$= \frac{\sqrt{n}}{\sigma}\left\{c_3 + \frac{\sigma^4}{2n}c_5(w_2 + \frac{l_4}{6}) + o(n^{-1})\right\} \qquad (4.2)$$

Step 3. The marginal density $m(x)$ admits the expansion

$$m(x) = \frac{\sigma}{\sqrt{n}}\exp(\mathcal{L}(\hat{\theta}, x))\sqrt{2\pi}\pi(\hat{\theta})\Big\{1 + \frac{\sigma^2}{2\sqrt{n}}W_n$$

$$+\frac{\sigma^2}{2n}\left(\frac{\pi''(\hat\theta)}{\pi(\hat\theta)}+\frac{\sigma^2}{4}\left(\frac{\mathcal{L}^{(4)}}{n}+3W_n^2\right)\right)+O(n^{-3/2})\Big\} \tag{4.3}$$

Proof. Proceeding as in Step 2,

$$\begin{aligned}
m_\pi(x) &= \int f(x|\theta)\pi(\theta)d\theta = \int \exp(\mathcal{L}(\theta,x))\pi(\theta)d\theta \\
&= \int\int \frac{\sigma}{\sqrt{n}}\exp(\mathcal{L}(\hat\theta,x)-\frac{z^2}{2}(1-\frac{\sigma^2}{\sqrt{n}}W_n)) \\
&\times\left\{1+\frac{\sigma^3}{6n}z^3\frac{\mathcal{L}^{(3)}}{\sqrt{n}}+\frac{\sigma^4}{24n}z^4\frac{\mathcal{L}^{(4)}}{n}+O(n^{-3/2})\right\} \\
&\times\left\{\pi(\hat\theta)+\frac{\sigma}{\sqrt{n}}z\pi'(\hat\theta)+\frac{\sigma^2}{2n}z^2\pi''(\hat\theta)++O(n^{-3/2})\right\}dz
\end{aligned}$$

$$\begin{aligned}
&= \int \frac{\sigma}{\sqrt{n}}\exp(\mathcal{L}(\hat\theta,x)-\frac{z^2}{2})\left\{1+\frac{\sigma^2}{2\sqrt{n}}W_nz^2+\frac{\sigma^4}{8n}W_n^2z^4+O(n^{-3/2})\right\} \\
&\times\left\{1+\frac{\sigma^3}{6n}z^3\frac{\mathcal{L}^{(3)}}{\sqrt{n}}+\frac{\sigma^4}{24n}z^4\frac{\mathcal{L}^{(4)}}{n}+O(n^{-3/2})\right\} \\
&\times\left\{\pi(\hat\theta)+\frac{\sigma}{\sqrt{n}}z\pi'(\hat\theta)+\frac{\sigma^2}{2n}z^2\pi''(\hat\theta)++O(n^{-3/2})\right\}dz \\
&= \frac{\sqrt{2\pi}\sigma}{\sqrt{n}}\exp(\mathcal{L}(\hat\theta,x)\int\phi(z)\Bigg\{\pi(\hat\theta)+\frac{1}{\sqrt{n}}\left(z^2\frac{\sigma^2W_n}{2}\pi(\hat\theta)+z\sigma\pi'(\hat\theta)\right) \\
&\quad+\frac{1}{n}\left(z^4\frac{\sigma^4}{8}W_n^2\pi(\hat\theta)+z^3\frac{\sigma^3W_n}{2}\pi'(\hat\theta)+z^3\frac{\sigma^3}{6}\frac{\mathcal{L}^{(3)}}{\sqrt{n}}\pi(\hat\theta)+z^4\frac{\sigma^4}{24}\frac{\mathcal{L}^{(4)}}{n}\pi(\hat\theta)+z^2\frac{\sigma^2}{2}\pi''(\hat\theta)\right) \\
&\qquad+O(n^{-3/2})\Bigg\}dz \\
&= \frac{\sqrt{2\pi}\sigma}{\sqrt{n}}\exp(\mathcal{L}(\hat\theta,x)\Bigg\{\pi(\hat\theta)+\frac{\sigma^2}{2\sqrt{n}}W_n\pi(\hat\theta)+\frac{\sigma^2}{2n}\left(\pi''(\hat\theta)+\frac{\sigma^2}{4}\left(\frac{\mathcal{L}^{(4)}}{n}+3W_n^2\right)\pi(\hat\theta)\right) \\
&\qquad+O(n^{-3/2})\Bigg\}.
\end{aligned}$$

Step 4. Using Step 3 and arguing as in Step 2, we obtain

$$\begin{aligned}
E_P(\pi(\theta|x)) &= \int\int \pi(\theta|x)f(x|\theta)\pi(\theta)d\theta dx \\
&= \int\int \frac{\exp(2\mathcal{L}(\theta,x))\pi^2(\theta)}{m(x)}d\theta dx
\end{aligned}$$

$$= \int\int \frac{\sigma}{\sqrt{n}} \exp(2[\mathcal{L}(\hat{\theta}, x) - \frac{z^2}{2}(1 - \frac{\sigma^2}{\sqrt{n}} W_n)]) \left\{1 + \frac{2\sigma^3}{6n} z^3 \frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{2\sigma^4}{24n} z^4 \frac{\mathcal{L}^{(4)}}{n} + O(n^{-3/2})\right\}$$

$$\times \left[\frac{\sqrt{2\pi}\sigma}{\sqrt{n}} \exp(\mathcal{L}(\hat{\theta}, x)\left\{\pi(\hat{\theta}) + \frac{\sigma^2}{2\sqrt{n}} W_n \pi(\hat{\theta}) + \frac{\sigma^2}{2n}\left(\pi''(\hat{\theta}) + \frac{\sigma^2}{4}\left(\frac{\mathcal{L}^{(4)}}{n} + 3W_n^2\right)\pi(\hat{\theta})\right)\right\}\right]^{-1}$$

$$\times \left[\pi(\hat{\theta}) + \frac{\sigma}{\sqrt{n}} z\pi'(\hat{\theta}) + \frac{\sigma^2}{2n} z^2 \pi''(\hat{\theta}) + +O(n^{-3/2})\right]^2 \, dz\, dx$$

$$= \int \exp(\mathcal{L}(\hat{\theta}, x))\left\{\pi(\hat{\theta}) + \frac{\sigma^2}{2\sqrt{n}} W_n \pi(\hat{\theta}) + \frac{\sigma^2}{2n}\left(\pi''(\hat{\theta}) + \frac{\sigma^2}{4}\left(\frac{\mathcal{L}^{(4)}}{n} + 3W_n^2\right)\pi(\hat{\theta})\right)\right\}^{-1}$$

$$\int \phi^2(z)\left\{\pi^2(\hat{\theta}) + \frac{1}{\sqrt{n}}(2\sigma z\pi'(\hat{\theta})\pi(\hat{\theta}) + \sigma^2 z^2 W_n \pi^2(\hat{\theta}))\right.$$

$$+ \frac{1}{n}\left(\sigma^2 z^2 (\pi'(\hat{\theta}))^2 + \sigma^2 z^2 \pi''(\hat{\theta})\pi(\hat{\theta}) + \frac{\sigma^4}{2} z^4 W_n^2 \pi^2(\hat{\theta}) + 2\sigma^3 z^3 W_n \pi'(\hat{\theta})\pi(\hat{\theta})\right.$$

$$\left.\left. + \left(\frac{2\sigma^3}{6} z^3 \frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{2\sigma^4}{24} z^4 \frac{\mathcal{L}^{(4)}}{n}\right)\pi^2(\hat{\theta})\right) + O(n^{-3/2})\right\} dz\, dx$$

$$= \frac{\sqrt{n}}{\sigma}\left\{c_3 + \frac{\sigma^2}{n} c_4 I(\pi) + \frac{\sigma^4}{2n}\left(w_2(c_5 - c_4 - \frac{c_3}{4}) + l_4(\frac{c_5}{6} - \frac{c_3}{4})\right) + o(n^{-1})\right\} \tag{4.4}$$

where $I(\pi)$ is the Fisher Information of π.

Step 5. Analogously,

$$E_P(f(x|\theta)\pi(\theta|x)) = \frac{n}{\sigma^2}\left(c_1 + c_2\frac{\sigma^2}{n} I(\pi) + \frac{\sigma^4}{4n}\left(w_2(\frac{9c_6}{2} - 3c_2 - \frac{c_1}{2}) + l_4(\frac{c_6}{2} - \frac{c_1}{2})\right)\right.$$
$$\left. + o(n^{-1})\right), \tag{4.5}$$

and so

$$\mathrm{Cov}_P(f(x|\theta, \pi(\theta|x))$$
$$= \frac{n}{\sigma^2}(c_1 - c_3^2) + (c_2 - c_3 c_4) I(\pi) + \frac{\sigma^2}{2}\left(w_2(\frac{9}{4}c_6 - \frac{3}{2}c_2 - \frac{c_1}{4} - 2c_3c_5 + c_3c_4 + \frac{c_3^2}{4})\right.$$
$$\left. + l_4(\frac{c_6}{2} - \frac{c_1}{2} - \frac{c_3 c_5}{3} + \frac{c_3^2}{4})\right) + o(1). \tag{4.6}$$

Step 6. Proceeding exactly as in Step 1, one obtains

$$E_P(f^2(x|\theta)) = \frac{n}{\sigma^2}\left(c_1 + \frac{\sigma^4}{8n} c_6 (9w_2 + l_4) + o(n^{-1})\right). \tag{4.7}$$

Step 7. Algebra similar to that in Step 4 yields,

$$E_P(\pi^2(\theta|x)) = \int\int \pi^2(\theta|x)f(x|\theta)\pi(\theta)\, d\theta dx = \int\int \frac{f^3(x|\theta)\pi^3(\theta)}{m^2(x)} d\theta dx$$
$$= \int \frac{1}{m^2(x)} \Bigg(\int \frac{\sigma}{\sqrt{n}} \exp(3[\mathcal{L}(\hat\theta, x) - \frac{z^2}{2}(1 - \frac{\sigma^2}{\sqrt{n}} W_n)]$$
$$\times \left\{ 1 + \frac{3\sigma^3}{6n} z^3 \frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{3\sigma^4}{24n} z^4 \frac{\mathcal{L}^{(4)}}{n} + O(n^{-3/2}) \right\}$$
$$\times \left[\pi(\hat\theta) + \frac{\sigma}{\sqrt{n}} z\pi'(\hat\theta) + \frac{\sigma^2}{2n} z^2 \pi''(\hat\theta) + +O(n^{-3/2}) \right]^3 dz\, dx$$
$$= \frac{n}{\sigma^2} \left(c_1 + 3c_2 \frac{\sigma^2}{n} I(\pi) + \frac{\sigma^4}{2n} \left(w_2(\frac{9}{4}c_6 - 3c_2) + l_4(\frac{c_6}{4} - \frac{c_1}{2}) \right) + o(n^{-1}) \right) \qquad (4.8)$$

Step 8. By (4.2), (4.7),

$$\mathrm{Var}_P(f(x|\theta)) = \frac{n}{\sigma^2}(c_1 - c_3^2) + \sigma^2 \left(w_2(\frac{9}{8}c_6 - c_3c_5) + l_4(\frac{c_6}{8} - \frac{c_3c_5}{6}) \right) + o(1), \qquad (4.9)$$

and by (4.4), (4.8),

$$\mathrm{Var}_P(\pi(\theta|x)) = \frac{n}{\sigma^2}(c_1 - c_3^2) + (3c_2 - 2c_3c_4)I(\pi) \qquad (4.10)$$
$$+ \frac{\sigma^2}{2} \left\{ w_2(\frac{9}{4}c_6 - 3c_2 - 2c_3c_5 + 2c_3c_4 + \frac{c_3^2}{2}) + l_4(\frac{c_6}{4} - \frac{c_1}{2} - \frac{c_3c_5}{3} + \frac{c_3^2}{2}) \right\} + o(1).$$

Step 9. Combining (4.6), (4.9), and (4.10), after several steps,

$$\rho_P(f(x|\theta), \pi(\theta|x)) = 1 - \frac{\sigma^2}{2n(c_1 - c_3^2)} c_2 I(\pi)) - \frac{\sigma^4}{8n(c_1 - c_3^2)} \{c_1 w_2 + (c_1 - c_6)l_4\}$$
$$+ o(n^{-1}), \qquad (4.11)$$

which concludes the description.

References

Berger, J. (1986). Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York.

Berger, J. (1994). An overview of robust Bayesian Analysis, Test, 3, 5–59.

Bickel, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted, *Ann. Stat.*, **9**, 1301–1309.

Bickel, P. J. and Collins, J. R. (1983). Minimizing Fisher information over mixtures of distributions, *Sankhya, A*, **45**, 1–19.

Brown, L. D. (1971). Admissibility, recurrent diffusions, and insoluble boundary value problems, *Ann. Stat.*, **42**, 855–903.

Brown, L. D. (1985). Fundamentals of Exponential Families, Institute of Mathematical Statistics Monograph, Hayward, California.

Cifarelli, D. M. and Regazzini, E. (1987). Priors for exponential families which maximize the association between past and future observations, In *Prob. and Bayesian Stat.*, Viertl, R. (Ed.), Plenum, London, 83–95.

Clarke, B. and Barron, A. (1994). Jeffrey's prior is asymptotically least favorable under entropy risk, *Jour. Stat. Planning and Inf.*, **41**, 37–60.

Clarke, B. and Wasserman, L. (1993). Noninformative priors and nuisance parameters, *Jour. Amer. Stat. Assoc.*, **88**, 1427–1432.

DasGupta, A., Casella, G., Delampady, M., Rubin, H. and Strawderman, W. E. (1998). The Pearson correlation coefficient as a useful tool in statistical theory, Tech Report #98–04, Dept. of Statistics, Purdue University.

Datta, G. S. and Ghosh, J. K. (1995). On priors providing frequentist validity for Bayesian inference, *Biometrika*, **82**, 37–45.

Efron, B. (1986). Why isn't everyone a Bayesian, *Amer. Statist.,* **40**, 1, 1–11.

Ghosh, J. K. (1994). Higher Order Asymptotics, Institute of Mathematical Statistics Monograph, Hayward, California.

Ghosh, J. K. and Mukerjee, R. (1992). Noninformative priors, In *Bayesian Stat.*, **4**, Bernardo, J. M. et al (Eds.), Oxford Univ. Press, 195–210.

Ghosh, J. K., Sinha, B. K. and Joshi, S. N. (1982). Expansions for posterior probability and integrated Bayes risk, In *Statistical Decision Theory and Related Topics*, **4**, Gupta,

S. S. and Berger, J. O. (Eds.), Academic Press, New York.

Huber, P. J. (1964). Robust estimation of a location parameter, *Ann. Math. Stat.*, **35**, 73–101.

Huber, P. J. (1974). Fisher information and spline interpolation, *Ann. Stat.*, **2**, 1029–1034.

Johnson, R. A. (1970). Asymptotic expansions associated with posterior distributions, *Ann. Math. Stat.*, **41**, 851–864.

Kagan, A. M., Linnik, Yu. V. and Rao, C. R. (1973). Characterization Problems in Mathematical Statistics, Jon Wiley, New York.

Kass, R. E. and Wasserman, L. (1996). Selection of prior distributions by formal rules, *Jour. Amer. Stat. Assoc.*, **91**, 1343–1370.

Lindley, D. V. (1961). The use of prior probability distributions in statistical inference and decision, In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, Univ. California Press, Berkeley.

Levit, B. (1979). On the theory of the asymptotic minimax property of second order, *Theo. Prob. Appl.*, **24**, 435–437.

Levit, B. (1980). Asymptotic minimax estimators of the second order, *Theo. Prob. Appl.*, **25**, 552–568.

Perez, J. (1998). Development of Expected Posterior Prior Distributions for Model Comparison, Ph.D. dissertation, Purdue University.

Stein, C. (1982). The coverage probabilities of confidence sets based on a prior distribution, Preprint.

Zen, Mei-Mei and DasGupta, A. (1993). Is robust Bayes real Bayes, *Stat. and Decision*, **11**, 37–60.