UNIFORM APPROXIMATION OF BAYES SOLUTIONS AND POSTERIORS: FREQUENTISTLY VALID BAYES INFERENCE

by

Saurabh Mukhopadhyay and Anirban DasGupta
Duke University Purdue University

Technical Report #93-12C

Department of Statistics
Purdue University

February 1993

◇

UNIFORM APPROXIMATION OF BAYES SOLUTIONS AND POSTERIORS: FREQUENTISTLY VALID BAYES INFERENCE

By Saurabh Mukhopadhyay[1] and Anirban DasGupta [2]

Institute of Statistics and Decision Sciences, Duke University
Department of Statistics, Purdue University

The problem of deriving classically acceptable Bayesian estimation procedures is important for synthesis and reconciliation of the classical and Bayesian approaches to inference problems. In this work, we consider inference problems for location parameters. The idea is that if one can produce priors for which the posterior densities are uniformly close to the likelihood function, then the corresponding Bayesian inference should also be close to classical inference, at least for location parameters. We describe a large family of prior distributions meeting this goal.

Apart from obtaining approximations for the posterior density itself, we also derive uniform approximations to the Bayes rule and the posterior expected loss. We also demonstrate that for these priors, the sampling distributions of the Bayes rule and the classical unbiased estimate are close uniformly in the parameter and that all $100(1-\alpha)\,\%$ Bayesian HPD sets have classical coverage probabilities uniformly close to $1-\alpha$ as well. All of our results are nonasymptotic in nature.

1 Introduction

The study of the frequentist properties of Bayes procedures and development of Bayesian methods which are frequentistly justifiable have received scattered but significant attention from many researchers; Stein (1985) asked which priors would produce Bayes confidence sets that are frequentistly calibrated up to order $\frac{1}{n}$. Woodroofe (1976) discusses frequentist properties of Bayes methods in sequential problems; Rubin (1984) considers frequentistly valid Bayesian calculations through a series of examples, including model choice. DasGupta and Studden (1989), Hartigan (1966), Bickel and Yahav (1969), Welch and Peers (1963), Ghosh and Mukerjee (1991), Strawderman (1971), Strawderman and Cohen (1971), Brown (1971), Brown and Hwang (1982), Berger (1980), Berger and Robert (1990), Ghosh (1992), Zidek (1970), Efron and Morris (1971), DasGupta and Rubin (1986), Casella and Berger (1987), etc. are some of the other works on reconciliation of the classical and Bayesian approaches. While from a subjective Bayes point of view, frequentist properties of Bayes solutions are not of any direct interest, the general topic is of obvious theoretical interest, and also of importance in robustness studies; see Berger (1985). Furthermore, development of priors that guarantee frequentist validity in some strong mathematical sense is of clear interest to statisticians who are not necessarily Bayesians, but recognize the use and advantages of the Bayesian method as a tool.

In 1971, L. Brown conjectured that for estimating a multivariate normal mean using a squared error loss, a proper Bayes minimax estimator does not exist for 4 or less dimensions. This conjecture was proved in Strawderman (1971), who also settled the conjecture that for dimensions 5 or

[1]Research supported in part by the NSF Grant DMS 89-230-71 at Purdue University as supporting the thesis work of which this was a part.

[2]Research supported in part by NSF Grant DMS 89-23071

AMS 1980 *subject classifications.* Primary 62F15, 62F10; Secondary 62F25.

Key words and phrases. Bayes rule; HPD sets; coverage probability; location parameters; multivariate normal; point and set estimation; posterior; uniform convergence; risk function.

more such estimators do exist. It was, of course, recognized that while a Bayes minimax estimator would not exist in low dimensions, Bayes estimators with bounded risk functions surely exist in all dimensions. Indeed, if $\pi(\boldsymbol{\theta})$ is a prior density for a multivariate normal mean, then the corresponding Bayes rule has a bounded risk whenever the gradient of $\log \pi(\boldsymbol{\theta})$ $\left(\frac{\|\nabla \pi(\boldsymbol{\theta})\|}{\pi(\boldsymbol{\theta})}\right)$ is uniformly bounded by some finite number k. Such priors are typically thought of as heavy tailed priors in the literature. For instance, in one dimension, normal priors do not meet this requirement, but the Double Exponential and t priors do. The principal goal of this article is to establish the fact that this gradient condition on the prior density implies satisfactory frequentist behavior for the implied Bayes solution in a wide variety of problems, the point estimation problem mentioned above being just one of them. It is demonstrated that the general location parameter case can be handled with some conditions on the underlying density, and that in fact, if one starts with a prior density of the type described above and then scales it, then not only do the Bayes procedures for the scaled priors have good frequentist properties, they satisfy remarkably strong uniform approximation properties as well.

An illustrative example.

Consider the problem of finding a set estimate for a multivariate normal mean, when the normal distribution has a covariance matrix equal to the Identity. The common classical solution is the well known Hotelling set, namely the sphere of radius $\chi_\alpha(p)$ centered at the observation $\boldsymbol{X}$. If somebody were to use a $N(\mathbf{0}, I)$ prior for the unknown mean, then the corresponding Bayes HPD set would be centered at $c.\boldsymbol{X}$ for appropriate $0 < c < 1$. It is well known and clear that this set is not frequentistly justifiable in the sense that its frequentist coverage probability converges to 0 as the norm of $\boldsymbol{\theta}$ goes to ∞.

As an alternative, consider any prior density on the mean $\boldsymbol{\theta}$ which satisfies the stated gradient condition $\frac{\|\nabla \pi(\boldsymbol{\theta})\|}{\pi(\boldsymbol{\theta})} < k$, $k < \infty$. Now consider the scaled version $\frac{1}{c^p}\pi(\frac{\boldsymbol{\theta}}{c})$, $c > 0$. We prove that the corresponding $100(1-\alpha)\%$ Bayes HPD set satisfies the following two remarkable properties :

- **a.** The frequentist coverage probability of the Bayes HPD set converges uniformly to the nominal level $1-\alpha$, uniformly in $\boldsymbol{\theta}$, as the scale c converges to ∞; note this is not an asymptotic result in the usual sense. We are not letting the sample size go to infinity.
- **b.** There is a spherical band around the classical Hotelling sphere such that the $100(1-\alpha)\%$ Bayes HPD set lies inside this band, and moreover the width of this band is a number free of the observation $\boldsymbol{X}$, and converges to 0 as the scale c converges to infinity. Again, this is a fixed sample size result.

The joint implication of these two is that the Bayes HPD set has uniformly justifiable frequentist validity, and furthermore the HPD set itself is uniformly visually similar to the classical Hotelling set, and that both of these goals can be achieved simply by choosing a large value for the scale c.

A variety of such results are proved in this article in various contexts. The principal achievement is that we demonstrate that the gradient condition which was only known to produce satisfactory solutions in the point estimation problem, is now mathematically demonstrated to be a completely unifying condition. We also illustrate many of the results numerically for better understanding. Many of the proofs are technically complex, and sometimes look intimidating. Consequently, we have deferred a number of them to an appendix. Section 2 introduces some notations used extensively in the article; section 3 considers the problem of a doubly uniform approximation of the posterior density. Section 4 treats the set estimation problem, and in Section 5 we go into uniform approximations of other quantities, including the posterior expected loss and Bayes point estimators, and the sampling distribution as a whole of the Bayes estimator. We would like to remind the reader that approximations and expansions of the posterior are studied in many papers, notably Meeden and Isaacson (1977), Ghosh, Sinha and Joshi (1982), Johnson (1967), Umbach (1978), Walker (1969), Woodroofe (1989), etc. However, the uniformity aspect that we focus on was not the goal in these articles. Also see Ibragimov and Khas'minsky (1972, 1981), Kadane and Chuang (1978),and Lehmann (1983) in this connection. Our results now leave open the exciting possibility of strong expansions of an array of Bayesian quantities, including the posterior, the Bayes rule, and the posterior loss. Such uniform expansions would be valuable on many grounds.

2 Notations and definitions

Consider the general inference and decision problem for a location vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)'$ based on an observable $\boldsymbol{X} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_p)'$, where the likelihood function of $\boldsymbol{\theta}$ given $\boldsymbol{X} = \boldsymbol{x}$ is of the form $f(\boldsymbol{x} - \boldsymbol{\theta})$. If we assume a prior on $\boldsymbol{\theta}$ with density $\pi(.)$, then the corresponding posterior density will be denoted by $\pi(\boldsymbol{\theta}|\boldsymbol{x})$. Throughout this article, we will denote the scaled prior density $\frac{1}{c^p}\pi(\frac{\boldsymbol{\theta}}{c})$ (for $c \geq 1$) by $\pi_c(\boldsymbol{\theta})$ and the corresponding posterior density by $\pi_c(\boldsymbol{\theta}|\boldsymbol{x})$. We denote the uniform prior on p–dimensional Euclidean space, $\boldsymbol{R}^p$, by π^0 ($\pi^0(\boldsymbol{\theta}) \equiv 1$). We denote the *moment generating function* (MGF) of $\boldsymbol{X}$ and $\boldsymbol{X}^\star = (|\boldsymbol{X}_1|, \cdots, |\boldsymbol{X}_p|)'$ at some $\boldsymbol{t} \in \boldsymbol{R}^p$ under $\boldsymbol{\theta} = \boldsymbol{0}$ by $M_f(\boldsymbol{t})$ and $M_f^\star(\boldsymbol{t})$ respectively. We use the notation $B_p(\boldsymbol{x}, r)$ to denote the closed p–ball in $\boldsymbol{R}^p$ of radius $r\,(> 0)$ and center at $\boldsymbol{x} \in \boldsymbol{R}^p$. We denote the volume and surface area of $B_p(\boldsymbol{x}, r)$ by $V_p(r)$ and $V_p'(r)$ respectively, (for $p = 1$, $V_p(r) = 2r$ and $V_p'(r) = 2$). To denote a nondecreasing (or nonincreasing) function of real variable, $g(x)$, we write $g(x) \uparrow x$ (or $g(x) \downarrow x$). We denote a real valued sequence a_n increasing in the limit to a by $a_n \nearrow a$. Similarly, $b_n \searrow b$ means a real valued sequence b_n decreases in the limit to b.

In Sections 4 and 5, we only consider normal density as our likelihood function and use the notation $\varphi_p(\boldsymbol{z})$ for p–dimensional standard normal density. For $p = 1$, we use $\varphi(z)$ instead of $\varphi_1(z)$. For prior π and likelihood function $\varphi_p(\boldsymbol{x} - \boldsymbol{\theta})$, the $100(1-\alpha)\%$ *highest posterior density* (HPD) set for observation $\boldsymbol{X} = \boldsymbol{x}$, denoted by $S_\pi(\boldsymbol{x})$, is given by

$$S_\pi(\boldsymbol{x}) \;=\; \Big\{\boldsymbol{\theta} \,:\, \pi(\boldsymbol{\theta}|\boldsymbol{x}) \geq k_\alpha(\pi, \boldsymbol{x})\Big\},$$

where

$$k_\alpha(\pi, \boldsymbol{x}) \;=\; \sup\left\{k \,:\, \int_{\pi(\boldsymbol{\theta}|\boldsymbol{x}) \geq k} \pi(\boldsymbol{\theta}|\boldsymbol{x})\, d\boldsymbol{\theta} \geq 1 - \alpha\right\}.$$

For uniform prior, π^0, $k_\alpha(\pi^0, \boldsymbol{x})$ does not depend on $\boldsymbol{x}$ and we simply denote it by k_α^0; also, $S_{\pi^0}(\boldsymbol{x})$ simply becomes $B_p(\boldsymbol{x}, \chi_\alpha(p))$, where $\chi_\alpha^2(p)$ is the upper α–th quantile of chi–square distribution with p degrees of freedom. (Notice that $\chi_\alpha(p)$ is also the solution of $\varphi^p(z) = k_\alpha^0$). The Hausdorff distance (see Dugundji (1975)) of two HPD sets $S_\pi(\boldsymbol{x})$ and $S_{\pi^0}(\boldsymbol{x})$ is denoted by $H_\pi(\boldsymbol{x})$ and we denote $\sup_{\boldsymbol{x} \in \boldsymbol{R}^p} H_\pi(\boldsymbol{x})$ by $H_\pi^\star$. The frequentist coverage probability of the HPD set $S_\pi(\boldsymbol{x})$ under $\boldsymbol{\theta}$, dented by $\Psi_\pi(\boldsymbol{\theta})$, is given by

$$\Psi_\pi(\boldsymbol{\theta}) = \int_{\boldsymbol{R}^p} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta}) I_{\{S_\pi(\boldsymbol{x})\ni\boldsymbol{\theta}\}}(\boldsymbol{x})\, d\boldsymbol{x},$$

where $I_A(\boldsymbol{x})$ denotes the indicator function of any set $A \subseteq \boldsymbol{R}^p$. The *confidence coefficient* (see Lehmann (1983)) of the HPD sets $S_\pi(.)$ is therefore given by the infimum of the classical coverage probabilities, $\Psi_\pi(\boldsymbol{\theta})$, over all $\boldsymbol{\theta} \in \boldsymbol{R}^p$. We denote it by $\Psi_\pi^\star$.

3 Main results

Consider a prior density $\pi(.)$ satisfying

$$\text{(1)} \qquad \frac{\|\nabla\, \pi(\boldsymbol{\theta})\|}{\pi(\boldsymbol{\theta})} \le k \qquad \forall\, \boldsymbol{\theta} \in \boldsymbol{R}^p,\ (p \ge 1),$$

for some fixed and finite $k\,(>0)$, where $\nabla\pi(.)$ denotes the gradient vector of $\pi(.)$ and $\|.\|$ denotes usual Euclidean distance in $\boldsymbol{R}^p$. The prior belief about $\boldsymbol{\theta}$ is expressed in terms of a scaled prior density $\pi_c(\boldsymbol{\theta})$ (for some $c \ge 1$), where $\pi(.)$ satisfies the gradient condition (1).

We use the following two different methods or criteria for measuring the uniform closeness of posterior densities $\pi_c(\boldsymbol{\theta}|\boldsymbol{x})$ to the likelihood function $f(\boldsymbol{x}-\boldsymbol{\theta})$:

(I)

$$\sup_{\boldsymbol{x}} \int |\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - f(\boldsymbol{x}-\boldsymbol{\theta})|\, d\boldsymbol{\theta};$$

(II)

$$\sup_{\boldsymbol{x}} \sup_{\boldsymbol{\theta}} |\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - f(\boldsymbol{x}-\boldsymbol{\theta})|.$$

In Theorems 3.1 and 3.2, we obtain approximations to the quantities described in (I) and (II). First, we need the following key lemma (see Appendix for proof):

<u>Lemma</u> **3.1** *If the MGF of* $\boldsymbol{X}$ *exists (in a neighborhood of* $\mathbf{0}$*), then for any prior sequence* $\{\pi_c(.)\}$ *with* $\pi(.)$ *satisfying (1),*

$$\text{(2)} \qquad c\,|\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - f(\boldsymbol{x}-\boldsymbol{\theta})| \le \left(d_1 + d_2 G(\boldsymbol{x}-\boldsymbol{\theta}) + d_3 e^{G(\boldsymbol{x}-\boldsymbol{\theta})\frac{k}{c_0}}\right) f(\boldsymbol{x}-\boldsymbol{\theta})$$

for all $c \ge c_0 (\ge 1)$*; here* $c_0, d_1, d_2,$ *and* d_3 *are appropriate constants, and* $G(\boldsymbol{z}) = \sum_{i=1}^{p} |z_i|$.

Thus, if we have a prior π satisfying (1), and if the MGF of $\boldsymbol{X}$ exists, the convergence in L_1 norm of the posterior to the likelihood function is immediate from the above lemma. With some minor additional conditions, we can also obtain doubly uniform convergence (in observation and the parameter). The following theorems formally summarize these results.

<u>Theorem</u> **3.1** *If the MGF of* $\boldsymbol{X}$ *exists and* π *satisfies (1), then*

$$\text{(3)} \qquad \sup_{\boldsymbol{x}} \int_{\boldsymbol{R}^p} |\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - f(\boldsymbol{x}-\boldsymbol{\theta})|\, d\boldsymbol{\theta} = \mathcal{O}(\frac{1}{c}) \qquad (\forall\, c \ge c_0).$$

Proof: By using Lemma 3.1, when the MGF of $\boldsymbol{X}$ exists then, for any prior sequence $\{\pi_c(.)\}$ with $\pi(.)$ satisfying (1), we get

$$
\begin{aligned}
c\int |\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - f(\boldsymbol{x}-\boldsymbol{\theta})|\, d\boldsymbol{\theta} &\leq d_1 + d_2\int G(\boldsymbol{\eta})f(\boldsymbol{\eta})\, d\boldsymbol{\eta} + d_3\int e^{G(\boldsymbol{\eta})\frac{k}{c_0}} f(\boldsymbol{\eta})\, d\boldsymbol{\eta} \\
&= M_0 \text{ (say)}, \quad (\forall c \geq c_0). \qquad (4)
\end{aligned}
$$

Also, from the proof of the above lemma (in Appendix), we can see that, $\int e^{G(\boldsymbol{\eta})\frac{k}{c_0}} f(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ exists and is finite. Hence M_0 is a finite constant (free of $\boldsymbol{x}$ and c) for any $c \geq c_0$. ♣

<u>Theorem</u> **3.2** *In addition to the conditions of the above theorem, if* $e^{\frac{G(\boldsymbol{z})}{c_1}} f(\boldsymbol{z})$ *is also uniformly bounded for some* $0 < c_1 < \infty$, *then*

$$
\sup_{\boldsymbol{x}} \sup_{\boldsymbol{\theta}} |\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - f(\boldsymbol{x}-\boldsymbol{\theta})| = \mathcal{O}(\frac{1}{c}) \qquad (\forall\, c \geq c_0 \vee kc_1) \qquad (5)
$$

Proof: Since $e^{\frac{G(\boldsymbol{z})}{c_1}} f(\boldsymbol{z})$ is uniformly bounded, for any $c \geq kc_1$ there exists a positive constant d_4 such that

$$
f(\boldsymbol{z}) \leq e^{G(\boldsymbol{z})\frac{k}{c}} f(\boldsymbol{z}) \leq e^{\frac{G(\boldsymbol{z})}{c_1}} f(\boldsymbol{z}) \leq d_4 \qquad \forall\, \boldsymbol{z} \in \boldsymbol{R}^p. \qquad (6)
$$

Also,

$$
G(\boldsymbol{z})f(\boldsymbol{z}) \leq c_1 e^{\frac{G(\boldsymbol{z})}{c_1}} f(\boldsymbol{z}) \leq c_1 d_4 \qquad \forall\, \boldsymbol{z} \in \boldsymbol{R}^p \quad (\text{since } e^x \geq x \text{ (say) for } x \geq 0)\,. \qquad (7)
$$

Applying (6) and (7) to (2), we obtain for $c \geq c_0 \vee kc_1$

$$
\begin{aligned}
c\,|\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - f(\boldsymbol{x}-\boldsymbol{\theta})| &\leq d_1 d_4 + d_2 c_1 d_4 + d_3 d_4 \\
&= M_1 \text{ (say) } < \infty \qquad (8)
\end{aligned}
$$

where M_1 is free of $\boldsymbol{\theta}$, $\boldsymbol{x}$ and c. The result follows. ♣

Numerical illustration: $\pi_c(\boldsymbol{\theta}|\boldsymbol{x})$ and $f(\boldsymbol{x}-\boldsymbol{\theta})$ are the posterior densities with respect to the priors π_c and π^0; if we denote the corresponding measures by $P_c(\cdot|\boldsymbol{x})$ and $P^0(\cdot|\boldsymbol{x})$ respectively, then

$$
\begin{aligned}
\sup_A \left|P_c(A|\boldsymbol{x}) - P^0(A|\boldsymbol{x})\right| &= \frac{1}{2}\Omega_{\pi_c}(\boldsymbol{x}) \\
\text{where} \qquad \Omega_{\pi_c}(\boldsymbol{x}) &\stackrel{\text{def}}{=} \int |\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - f(\boldsymbol{x}-\boldsymbol{\theta})| d\boldsymbol{\theta}.
\end{aligned}
$$

This fact enhances the interest in the quantity $\Omega_{\pi_c}(\boldsymbol{x})$. So, we will consider a particular one dimensional example to actually compute and plot these values; we also compute $\sup_x \Omega_{\pi_c}(x)$, which we will denote by $\Omega^*_{\pi_c}$.

Let X follow $N(\theta, 1)$ and let θ have a scaled Cauchy prior given by

$$
\mathcal{C}_c(\theta) = \frac{1}{c}\mathcal{C}(\frac{\theta}{c}), \qquad \mathcal{C}(\theta) = \frac{1}{\pi}\frac{1}{1+\theta^2}, \ -\infty < \theta < \infty.
$$

It is straightforward to check that if the prior is symmetric around zero and the likelihood function is symmetric around the observation x, then the function $\Omega_{\pi_c}(.)$ is also symmetric (around zero). Therefore, we plot $\Omega_{\mathcal{C}_c}(x)$ as a function of $x(\geq 0)$ for different values of $c(\geq 1)$ (see Fig 1). In Table 1, we show corresponding values of $\Omega^{\star}_{\mathcal{C}_c}$.

In the following two sections, we apply our results to estimating multivariate normal means by using any prior family $\{\pi_c\}$ satisfying gradient condition (1) and obtain closed form approximations of the Bayesian credible sets, and various Bayesian quantities arising from squared error and linear loss functions.

4 Applications to set estimation of normal means

Set estimation is one of the most common statistical problems and has a long history. The literature is too huge to list here; see Lehman (1986). We consider set estimation of multivariate normal means here. While the particular sampling distribution considered here is of interest in its own right, the techniques used to obtain the results can also be useful for some other location families.

Our goal in this section is to show that for a prior π satisfying (1), the implied HPD sets will be close to the Hotelling confidence set uniformly over the observation. Notice that such credible sets will, therefore, have an inherent Bayesian robustness with respect to the particular choice of any prior satisfying (1). Also notice that, as we mentioned in the introduction,the strong uniform approximations we are about to prove do not hold for any scaled normal priors.

Let $\boldsymbol{X}$ follow a p–variate normal distribution with unknown mean vector $\boldsymbol{\theta}$ and covariance matrix I (identity matrix of order p). Then the standard Hotelling confidence set will simply be $S_{\pi^0}(\boldsymbol{x})$.

As described before, our objective, in one hand, is to obtain classically acceptable Bayesian credible sets, and on the other hand, is to obtain robust Bayesian credible sets each being uniformly close to the Hotelling confidence set. To meet the objectives, we consider the following two formulations:

(i) To meet our first objective, we consider the confidence coefficient $\Psi^{\star}_{\pi}$ to assess if the HPD set $S_{\pi}(.)$ is classically justifiable. The asymptotic version of this problem has been studied before; see Stein (1985), Welch and Peers (1963), Hartigan (1966), Ghosh and Mukerjee (1991). Our results are for fixed sample size and therefore in some sense have a more direct bearing on practice. Observe that, by construction, $\Psi_{\pi^0}(\boldsymbol{\theta}) \equiv 1-\alpha$. So, for a given π_c, the implied HPD credible sets are classically justifiable if $\Psi^{\star}_{\pi_c}$ is close to the nominal level $1-\alpha$. Indeed, we shall prove that $\Psi^{\star}_{\pi}$ is $1-\alpha+\mathcal{O}(\frac{1}{c})$.

(ii) We will also give an uniform approximation of the HPD set $S_{\pi_c}(\boldsymbol{x})$ itself; the exact result is given in (10). Roughly, the assertion in (10) says that one can find a small spherical band around the classical set in which the Bayes solution lies and that the width of this band is a small number, free of the observation $\boldsymbol{x}$. We find this remarkable.

We start with the following lemma (see Appendix for proof) which is essential for proving the results.

Lemma **4.1** *For any fixed $\varepsilon > 0$ and a fixed prior density $\pi(.)$ of $\boldsymbol{\theta}$ satisfying (1), let us denote*

$$S^{\varepsilon}_{\pi}(\boldsymbol{x}) \;=\; \left\{\boldsymbol{\theta} : \pi(\boldsymbol{\theta}|\boldsymbol{x}) \geq k^0_{\alpha} - \varepsilon\right\}. \tag{9}$$

Then there exist fixed positive constants L_1 and U_1, such that

$$S_{\pi_c}^{-\frac{L_1}{c}}(\boldsymbol{x}) \subseteq S_{\pi_c}(\boldsymbol{x}) \subseteq S_{\pi_c}^{\frac{U_1}{c}}(\boldsymbol{x}), \quad \forall\, \boldsymbol{x} \in \boldsymbol{R}^p,\ \forall\, c \geq c_2,$$

for appropriate $c_2\,(\geq 1)$.

<u>Theorem</u> **4.1** *Assume $\pi(.)$ satisfies the gradient condition (1). Then there exist universal positive constants $c^\star(\geq 1)$, $U(>0)$, and $L(>0)$ such that*

$$B_p\left(\boldsymbol{x}, \chi_\alpha(p) - \frac{L}{c}\right) \subseteq S_{\pi_c}(\boldsymbol{x}) \subseteq B_p\left(\boldsymbol{x}, \chi_\alpha(p) + \frac{U}{c}\right) \quad \forall\, \boldsymbol{x} \in \boldsymbol{R}^p, \tag{10}$$

whenever $c \geq c^\star$.

Proof: We will only prove the right inclusion part; the other part is exactly similar. In order to prove the right inclusion, first in Step I, we show that the HPD set $S_{\pi_c}(\boldsymbol{x})$ is contained in a closed ball with center at $\boldsymbol{x}$ and of radius $\chi_\alpha(p) + \zeta_c$ for some appropriate positive constant ζ_c. Then, in Step II, we prove our claim by showing that ζ_c is in fact $\mathcal{O}(\frac{1}{c})$.

Step I: By Theorem 3.2 (or equivalently, (8)) and the above lemma, we know that whenever $c \geq c_0 \vee kc_1 \vee c_2 (= c_3,\ say)$ one has that for all $\boldsymbol{x} \in \boldsymbol{R}^p$,

$$\begin{aligned}
S_{\pi_c}(\boldsymbol{x}) &\subseteq S_{\pi_c}^{\frac{U_1}{c}}(\boldsymbol{x}) = \left\{\boldsymbol{\theta} \,:\, \pi_c(\boldsymbol{\theta}|\boldsymbol{x}) \geq k_\alpha^0 - \frac{U_1}{c}\right\} \\
&\subseteq \left\{\boldsymbol{\theta} \,:\, \varphi_p(\boldsymbol{x} - \boldsymbol{\theta}) \geq k_\alpha^0 - \frac{U_1}{c} - \frac{M_1}{c}\right\} \qquad \text{[by (8)]} \\
&= \{\boldsymbol{\theta} \,:\, ||\boldsymbol{\theta} - \boldsymbol{x}|| \leq \chi_\alpha(p) + \zeta_c\} \\
&= B_p(\boldsymbol{x}, \chi_\alpha(p) + \zeta_c)
\end{aligned} \tag{11}$$

where ζ_c is given by

$$\begin{aligned}
&\varphi^p(\chi_\alpha(p) + \zeta_c) = k_\alpha^0 - \frac{M_1 + U_1}{c} \\
\Rightarrow\ &\zeta_c\, p\,(\chi_\alpha(p) + \zeta_c^\star)\varphi^p(\chi_\alpha(p) + \zeta_c^\star) = \frac{M_1 + U_1}{c}, \text{ where } 0 < \zeta_c^\star < \zeta_c
\end{aligned} \tag{12}$$

[by Taylor expansion and using the fact that $\varphi^p(\chi_\alpha(p)) = k_\alpha^0$].

Step II: Since $\varphi^p(z)$ is a *strictly* decreasing function of $z > 0$, we can choose c_4 large enough, so that

$$\varphi^p(\chi_\alpha(p)) - \varphi^p(\chi_\alpha(p) + 1) > \frac{M_1 + U_1}{c}.$$

Hence, for $c \geq c_4$ we always have $\zeta_c \leq 1$, and whence

$$\begin{aligned}
\inf_{0<\zeta_c^\star<\zeta_c} (\chi_\alpha(p) + \zeta_c^\star)\varphi^p(\chi_\alpha(p) + \zeta_c^\star) &\geq \inf_{0<\zeta_c^\star<1} (\chi_\alpha(p) + \zeta_c^\star)\varphi^p(\chi_\alpha(p) + \zeta_c^\star) \\
&= \min\{\chi_\alpha(p)\varphi^p(\chi_\alpha(p)), (\chi_\alpha(p) + 1)\varphi^p(\chi_\alpha(p) + 1)\} \\
&= M_2 \text{ (say)}.
\end{aligned}$$

So, from (12) and above, it follows that for all $c \geq c_3 \vee c_4$ we have

$$\begin{aligned}\frac{M_1+U_1}{c} &= \zeta_c\, p\,(\chi_\alpha(p)+\zeta_c^\star)\varphi^p(\chi_\alpha(p)+\zeta_c^\star) \geq \zeta_c p M_2 \\ \Rightarrow \zeta_c &\leq \frac{M_1+U_1}{p\,M_2\,c} = \frac{U}{c} \quad \left[\text{by writing } U = \frac{M_1+U_1}{p\,M_2}\right],\end{aligned}$$

and hence from (11), we have

$$\begin{aligned} S_{\pi_c}(\boldsymbol{x}) &\subseteq B_p(\boldsymbol{x}, \chi_\alpha(p)+\zeta_c) \\ &\subseteq B_p\left(\boldsymbol{x}, \chi_\alpha(p)+\frac{U}{c}\right) \quad \forall\, \boldsymbol{x} \in \boldsymbol{R}^p,\ \forall\, c \geq c_3 \vee c_4, \end{aligned} \tag{13}$$

as claimed. ♣

<u>*Corollary*</u> **4.1** *$\Psi^\star_{\pi_c} = 1-\alpha+\mathcal{O}(\frac{1}{c})$ for $c \geq c^\star$.*

Proof: By definition

$$\Psi_{\pi_c}(\boldsymbol{\theta}) = \int_{\boldsymbol{R}^p} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta}) I_{\{S_\pi(\boldsymbol{x})\ni\boldsymbol{\theta}\}}(\boldsymbol{x})\, d\boldsymbol{x}.$$

So, using (10), we get for $c \geq c^\star$ that

$$\begin{aligned}\Psi_{\pi_c}(\boldsymbol{\theta}) &\geq \int_{\boldsymbol{R}^p} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta}) I_{\{B_p(\boldsymbol{x},\chi_\alpha(p)-\frac{L}{c})\ni\boldsymbol{\theta}\}}(\boldsymbol{x})\, d\boldsymbol{x} \\ &= 1-\alpha-\int_{\chi_\alpha(p)-\frac{L}{c}\leq\|\boldsymbol{x}-\boldsymbol{\theta}\|\leq\chi_\alpha(p)} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})\, d\boldsymbol{x} \\ &= 1-\alpha-\int_{\chi_\alpha(p)-\frac{L}{c}\leq\|\boldsymbol{x}\|\leq\chi_\alpha(p)} \varphi_p(\boldsymbol{x})\, d\boldsymbol{x} \\ &\geq 1-\alpha-\varphi^p(\chi_\alpha(p))\int_{\chi_\alpha(p)-\frac{L}{c}\leq\|\boldsymbol{x}\|\leq\chi_\alpha(p)} d\boldsymbol{x} \qquad [\text{since } \varphi_p(\boldsymbol{x}) \downarrow \|\boldsymbol{x}\|] \\ &= 1-\alpha-k_\alpha^0\left[V_p(\chi_\alpha(p))-V_p(\chi_\alpha(p)-\frac{L}{c})\right] \\ &\geq 1-\alpha-k_\alpha^0\frac{L}{c}V_p'(\chi_\alpha(p))\end{aligned}$$

(by using Taylor expansion and the fact that $V_p(r)\uparrow r$). Thus, writing $\overline{L} = k_\alpha^0\, L\, V_p'(\chi_\alpha(p))$, we have

$$\begin{aligned}\Psi_{\pi_c}(\boldsymbol{\theta}) &\geq 1-\alpha-\frac{\overline{L}}{c} \quad \forall\boldsymbol{\theta}\in\boldsymbol{R}^p \\ \Rightarrow \Psi^\star_{\pi_c} &\geq 1-\alpha-\frac{\overline{L}}{c}\end{aligned}$$

whenever $c \geq c^\star$. The result follows. ♣

In the following corollary we establish that the Hausdorff distance between the HPD set and the Hotelling confidence set is small uniformly over the observation. Notice that a small Lebesgue measure of the symmetric difference of those two sets, which is an alternative measure to determine their closeness, does not guarantee a visual similarity of the two sets; however, a small Hausdorff distance generally does. The uniform approximation result of Theorem 4.1 will imply similar results for many other distances, Hausdorff distance is particularly nice.

Corollary **4.2** $\sup_{\boldsymbol{x}} H_{\pi_c}(\boldsymbol{x}) = \mathcal{O}(\frac{1}{c})$ *for* $c \geq c^\star$.

Proof: From (10), it immediately follow that for $c \geq c^\star$

$$0 \leq H_{\pi_c}(\boldsymbol{x}) \leq \frac{U}{c} \vee \frac{L}{c} \leq \frac{U+L}{c}, \ \forall \boldsymbol{x} \in \boldsymbol{R}^p.$$

♣

Examples and numerical illustrations: We consider the construction of 95 % HPD credible sets (i.e., $\alpha = 0.05$). The numerical examples we consider here, using scaled Cauchy priors, illustrate the convergences described in Corollaries 4.2 and 4.1. We use the same notations for scaled Cauchy priors ($\mathcal{C}_c$) as were used in Section 3.

In Figure 2, we plot $\Psi_{\mathcal{C}_c}(\theta)$ as a function of θ for different values of c. We can see that for each fixed c, the plot of $\Psi_{\mathcal{C}_c}(\theta)$ is symmetric around zero which actually follows from the symmetry of normal and Cauchy densities. Also notice that, for each c, $\Psi_{\mathcal{C}_c}(\theta)$ attains its supremum at $\theta = 0$ where it always exceeds the nominal level; then, as θ increases, it decreases below the nominal level and attains the infimum ($\Psi^\star_{\mathcal{C}_c}$) for some moderate value of θ (depending on c). Thereafter, $\Psi_{\mathcal{C}_c}(\theta)$ increases in the limit to the nominal level. We compute $\Psi^\star_{\mathcal{C}_c}$ for different values of c (see Table 3). We also compute $H^\star_{\mathcal{C}_c}$ for different values of $c\,(\geq 1)$ (see Table 2).

5 Further applications

Intuitively, the doubly uniform convergence of the posterior to the likelihood function, established in Section 3, should imply a broad range of further approximation results. For instance, one would expect that for many loss functions, the Bayes rule for the prior π_c should be uniformly close to the uniform prior Bayes rule. Two concrete and common losses are addressed here, although, generalization seems certainly possible. We consider a univariate normal random variable $X \sim N(\theta, 1)$. The two losses we consider are squared error loss and the linear loss of the form

$$L(\theta, a) = \begin{cases} K_0(\theta - a) & \text{if } \theta - a \geq 0 \\ K_1(a - \theta) & \text{if } \theta - a < 0 \end{cases} \tag{14}$$

for some arbitrary but fixed positive constants K_0 and K_1. The quantities of interest are the Bayes rule, the posterior expected loss, the Bayes risk, the risk function, and the frequentist distribution of the Bayes rule.

5.1 Approximation of the Bayes rule:

Here, we will show that if $\delta_c(x)$ denotes the Bayes rule implied by the prior $\pi_c(\theta)$, and $\delta^0(x)$ denotes the generalized Bayes rule implied by the uniform prior $\pi^0(\theta) \equiv 1$, then both for squared error and the linear loss function,

$$\sup_x |\delta_c(x) - \delta^0(x)| = \mathcal{O}(\frac{1}{c}).$$

For squared error loss, the following theorem is immediate from the variational formulas of Brown and Hwang (1982).

Theorem **5.1** *For any $c \geq 1$,*

$$\left| \int \theta \, \pi_c(\theta|x) \, d\theta - x \right| \leq \frac{k}{c} \quad \forall x \in \boldsymbol{R}. \tag{15}$$

provided $\pi(\theta)$ satisfies (1), i.e., $\left|\frac{\pi'(\theta)}{\pi(\theta)}\right| \leq k < \infty$.

Next, considering the linear loss of the form (14), we will show that similar approximations for the implied Bayes rule are also available here. First notice that (see Section 4.4 in Berger (1985)), the Bayes rule is any $\frac{K_0}{K_0+K_1}$-th fractile of the posterior distribution. So, an approximation of the Bayes rule is equivalent to an approximation of this fractile. For a fixed α, $(0 < \alpha < 1)$, let $q_c(x)$ denote the α–th fractile of the posterior density $\pi_c(\theta|x)$, and let $q^0(x)$ denote the α–th fractile of the normal density with mean at x and variance one. Then we have the following approximation of $q_c(x)$.

Theorem **5.2**

$$\sup_x \left| q_c(x) - q^0(x) \right| = \mathcal{O}(\frac{1}{c}),$$

provided $\pi(\theta)$ satisfies the gradient condition (1).

Proof: From Theorem 3.1, we know that there exists $c_0 \geq 1$ such that

$$\sup_x \int_{\boldsymbol{R}} |\pi_c(\theta|x) - \varphi(\theta - x)| \, d\theta \leq \frac{M_0}{c}, \quad \forall c \geq c_0, \tag{16}$$

for some positive constant M_0. So, for $c \geq c_0$, we have

$$\begin{aligned} \left| \alpha - \int_{-\infty}^{q_c(x)} \varphi(\theta - x) \, d\theta \right| &= \left| \int_{-\infty}^{q_c(x)} \pi_c(\theta|x) \, d\theta - \int_{-\infty}^{q_c(x)} \varphi(\theta - x) \, d\theta \right| \\ &\leq \int_{-\infty}^{q_c(x)} |\pi_c(\theta|x) - \varphi(\theta - x)| \, d\theta \\ &\leq \int_{\boldsymbol{R}} |\pi_c(\theta|x) - \varphi(\theta - x)| \, d\theta \\ &\leq \frac{M_0}{c} \quad \forall x \in \boldsymbol{R}, \text{ [by (16)]}. \end{aligned} \tag{17}$$

Notice that $q^0(x) = x + q^0(0)$. Fix any $\varepsilon(> 0)$, however small; then there exists $c_6(\varepsilon)(\geq 1)$ such that for all $c \geq c_6(\varepsilon)$,

$$\left.\begin{array}{rrcl} & \int_{-\infty}^{q^0(x)+\varepsilon} \varphi(\theta - x)\, d\theta & = & \int_{-\infty}^{q^0(0)+\varepsilon} \varphi(\theta)\, d\theta > \alpha + \frac{M_0}{c} \\ \text{and} & \int_{-\infty}^{q^0(x)-\varepsilon} \varphi(\theta - x)\, d\theta & = & \int_{-\infty}^{q^0(0)-\varepsilon} \varphi(\theta)\, d\theta < \alpha - \frac{M_0}{c} \end{array}\right\} \quad \forall c \geq c_6(\varepsilon).$$

Thus combining (17) and the above, we have that for $c \geq c_0 \vee c_6(\varepsilon)$,

$$\sup_x |q_c(x) - q^0(x)| \leq \varepsilon.$$

Hence, for $c \geq c_0 \vee c_6(\varepsilon)$, we get

$$\begin{aligned} \left| \alpha - \int_{-\infty}^{q_c(x)} \varphi(\theta - x)\, d\theta \right| &= \left| \int_{q^0(x)}^{q_c(x)} \varphi(\theta - x)\, d\theta \right| \\ &= \left| \int_{q^0(x)}^{q^0(x)+(q_c(x)-q^0(x))} \varphi(\theta - x)\, d\theta \right| \\ &= \left| \int_{q^0(0)}^{q^0(0)+(q_c(x)-q^0(x))} \varphi(\theta)\, d\theta \right| \quad [\text{since } q^0(x) = x + q^0(0)] \\ &\geq \left| q_c(x) - q^0(x) \right| \min\left\{ \varphi(q^0(0) - \varepsilon), \varphi(q^0(0) + \varepsilon) \right\}, \qquad (18) \\ &\qquad [\text{since } \sup_x |q_c(x) - q^0(x)| \leq \varepsilon]. \end{aligned}$$

(19)

The assertion of the theorem now clearly follows from (17) and (18). ♣

5.2 Approximations of posterior expected losses and Bayes risks:

From a Bayesian decision theoretic viewpoint, apart from the Bayes rule, other quantities of interest are the posterior expected loss and the Bayes risk. In this section we will obtain an approximation of the posterior expected loss uniformly over the observation x and hence an approximation of the same order will also hold for the Bayes risk. We will therefore make no further mention of the approximation of Bayes risks.

We denote the posterior expected loss of an action a for an observation $X = x$ under prior π by $\rho(\pi(\theta|x), a)$. Then denoting the posterior expected loss at x for prior π_c and π^0 by $\rho_{\pi_c}(x) = \rho(\pi_c(\theta|x), \delta_c(x))$ and $\rho_{\pi^0}(x) = \rho(\pi^0(\theta|x), \delta^0(x))$ respectively, we obtain uniform approximation of $\rho_{\pi_c}(x)$ for both squared error and the linear loss function in the following theorem.

<u>*Theorem*</u> **5.3** $\sup_x |\rho_{\pi_c}(x) - \rho_{\pi^0}(x)| = \mathcal{O}(\frac{1}{c})$.

Proof: Let us denote $\xi(z) = (d_1 + d_2|z| + d_3 e^{|z|\frac{k}{c_0}})\varphi(z)$, where d_i's are positive constants as in Lemma 3.1. Then, using Lemma 3.1, we have that there exists $c_0(\geq 1)$ such that whenever $c \geq c_0$,

$$|\pi_c(\theta|x) - \varphi(\theta - x)| \leq \frac{1}{c}\xi(\theta - x), \quad \forall\, \theta \in \boldsymbol{R},\ \forall\, x \in \boldsymbol{R}. \tag{20}$$

Thus for any loss function $L(\theta, a)$, we have

$$\begin{aligned}
|\rho_{\pi_c}(x) - \rho_{\pi^0}(x)| &\leq \left|\rho_{\pi_c}(x) - \rho(\pi^0(\theta|x), \delta_c(x))\right| + \left|\rho(\pi^0(\theta|x), \delta_c(x)) - \rho_{\pi^0}(x)\right| \\
&\leq \int_{\boldsymbol{R}} L(\theta, \delta_c(x))\, |\pi_c(\theta|x) - \varphi(\theta - x)|\, d\theta \\
&\quad + \left|\int_{\boldsymbol{R}} \left\{L(\theta, \delta_c(x)) - L(\theta, \delta^0(x))\right\}\varphi(\theta - x)\, d\theta\right| \\
&\leq \frac{1}{c}\int_{\boldsymbol{R}} L(\theta, \delta_c(x))\xi(\theta - x)\, d\theta \\
&\quad + \left|\int_{\boldsymbol{R}} \left\{L(\theta, \delta_c(x)) - L(\theta, \delta^0(x))\right\}\varphi(\theta - x)\, d\theta\right| \\
&\qquad \text{[by (20), for all } c \geq c_0] \\
&\leq \frac{1}{c}\int_{\boldsymbol{R}} L(\theta, \delta^0(x))\xi(\theta - x)\, d\theta \\
&\quad + \frac{1}{c}\left|\int_{\boldsymbol{R}} \left\{L(\theta, \delta_c(x)) - L(\theta, \delta^0(x))\right\}\xi(\theta - x)\, d\theta\right| \\
&\quad + \left|\int_{\boldsymbol{R}} \left\{L(\theta, \delta_c(x)) - L(\theta, \delta^0(x))\right\}\varphi(\theta - x)\, d\theta\right| \qquad (21)
\end{aligned}$$

Now we will consider the squared error loss and the linear loss case by case and obtain the required result.

Case I: $L(\theta, a) = (\theta - a)^2$:

Notice that here $\delta^0(x) = x$ and hence

$$\begin{aligned}
L(\theta, \delta_c(x)) - L(\theta, \delta^0(x)) &= (\delta_c(x) - x)^2 + 2(\theta - x)(x - \delta_c(x)) \\
\Rightarrow \int_{\boldsymbol{R}} \left\{L(\theta, \delta_c(x)) - L(\theta, \delta^0(x))\right\}\xi(\theta - x)\, d\theta &= (\delta_c(x) - x)^2 \int_{\boldsymbol{R}} \xi(\theta - x)\, d\theta \\
&\qquad \text{[since } \xi(z) \text{ is a symmetric function]} \\
\text{and } \int_{\boldsymbol{R}} \left\{L(\theta, \delta_c(x)) - L(\theta, \delta^0(x))\right\}\varphi(\theta - x)\, d\theta &= (\delta_c(x) - x)^2.
\end{aligned}$$

Also notice that

$$\begin{aligned}
\int_{\boldsymbol{R}} (\theta - x)^2 \xi(\theta - x)\, d\theta &= \int_{\boldsymbol{R}} t^2 \xi(t)\, dt < \infty \\
\text{and } (\delta_c(x) - x)^2 &= \mathcal{O}\left(\frac{1}{c^2}\right), \quad \text{[by (15)]}
\end{aligned}$$

Hence from (21) we see that whenever $c \geq c_0$,

$$\sup_x |\rho_{\pi_c}(x) - \rho_{\pi^0}(x)| \leq \frac{1}{c}\int_{\boldsymbol{R}} t^2\xi(t) + \mathcal{O}(\frac{1}{c^2}).$$

This completes the proof for the squared error loss function.

Case II: $L(\theta, a)$ is linear given by (14):
In this case it is easy to see that

$$\Big|L(\theta,\delta_c(x)) - L(\theta,\delta^0(x))\Big| \leq (K_0 \vee K_1)|\delta_c(x) - \delta^0(x)|; \tag{22}$$

moreover, since $\delta_c(x) = q_c(x)$ here, and $\delta^0(x) = q^0(x) = x + q^0(0)$, we get from (21) that for $c \geq c_0$,

$$\begin{aligned}
|\rho_{\pi_c}(x) - \rho_{\pi^0}(x)| &\leq \frac{1}{c}\Bigg[\int_{\theta \geq x + q^0(0)} K_0(\theta - x - q^0(0))\xi(\theta - x)\,d\theta \\
&\quad + \int_{\theta < x+q^0(0)} K_1(x + q^0(0) - \theta)\xi(\theta - x)\,d\theta\Bigg] \\
&\quad + \int_{\boldsymbol{R}} \Big|L(\theta,\delta_c(x)) - L(\theta,\delta^0(x))\Big|\left\{\varphi(\theta - x) + \frac{1}{c}\xi(\theta - x)\right\}d\theta \\
&\leq \frac{1}{c}\left[\int_{t \geq q^0(0)} K_0(t - q^0(0))\xi(t)\,dt + \int_{t<q^0(0)} K_1(q^0(0) - t)\xi(t)\,dt\right] \\
&\quad + (K_0 \vee K_1)|\delta_c(x) - \delta^0(x)| \int_{\boldsymbol{R}} \left(\varphi(t) + \frac{1}{c}\xi(t)\right)\,dt \\
&\qquad \text{(on using (22)).}
\end{aligned}$$

By Theorem 5.2, $\sup_x |\delta_c(x) - \delta^0(x)| = \mathcal{O}(\frac{1}{c})$ and therefore

$$\sup_x |\rho_{\pi_c}(x) - \rho_{\pi^0}(x)| = \mathcal{O}(\frac{1}{c}).$$

This proves the result for the linear loss.

♣

5.3 Approximation of frequentist distributions of Bayes rules:

Let us denote

$$\begin{aligned}
F_{c,\theta}(t) &= P_{X|\theta}(\delta_c(X) \leq t) \\
\text{and } F^0_\theta(t) &= P_{X|\theta}(\delta^0(X) \leq t).
\end{aligned}$$

Then the following theorem holds:

<u>*Theorem*</u> **5.4** *For both the squared error and linear loss functions,*

$$\sup_\theta \sup_t \left|F_{c,\theta}(t) - F_\theta^0(t)\right| \to 0 \quad as\ c \to \infty$$

Proof: By using results from Section 5.1, there exists $c_7(\geq 1)$ such that

$$\sup_x |\delta_c(x) - \delta^0(x)| \;\leq\; \frac{M_3}{c}, \;\forall\, c \geq c_7$$

for some positive constant M_3. Hence, for $c \geq c_7$ we will have for all θ and t,

$$\begin{aligned} & P_{X|\theta}\left(\delta^0(X) \leq t - \frac{M_3}{c}\right) \;\leq\; P_{X|\theta}\left(\delta_c(X) \leq t\right) \;\leq\; P_{X|\theta}\left(\delta^0(X) \leq t + \frac{M_3}{c}\right) \\ \Leftrightarrow\quad & F_\theta^0\left(t - \frac{M_3}{c}\right) \;\leq\; F_{c,\theta}(t) \;\leq\; F_\theta^0\left(t + \frac{M_3}{c}\right) \end{aligned}$$

Now, notice that $F_\theta^0(t + \mathcal{O}(\frac{1}{c}))$ converges to $F_\theta^0(t)$ as $c \to \infty$ uniformly in θ and t. Hence the result follows. ♣

5.4 Approximation of risk functions:

In the following theorem, we obtain an uniform approximation for $R(\theta, \delta_c)$, again for both the squared error and the linear loss functions.

<u>*Theorem*</u> **5.5** $\sup_\theta |R(\theta,\delta_c) - R(\theta,\delta^0)| = \mathcal{O}(\frac{1}{c})$.

Proof: First notice that

$$\sup_\theta |R(\theta,\delta_c) - R(\theta,\delta^0)| \;\leq\; E_{X|\theta}\,|L(\theta,\delta_c(X)) - L(\theta,\delta^0(X))| \tag{23}$$

For the case of squared error loss,

$$\begin{aligned} E_{X|\theta}\,|L(\theta,\delta_c(X)) - L(\theta,\delta^0(X))| \;&\leq\; 2E_{X|\theta}\,[|\theta - X||X - \delta_c(X)|] \\ &\quad\; + E_{X|\theta}\,(X - \delta_c(X))^2 \\ &\leq\; 2\frac{k}{c}E_{X|\theta}\,|\theta - X| + \frac{k^2}{c^2}, \quad \text{[by (15)]}. \end{aligned} \tag{24}$$

In the case of a linear loss, we have for $c \geq c_7$,

$$\begin{aligned} E_{X|\theta}\,|L(\theta,\delta_c(X)) - L(\theta,\delta^0(X))| \;&\leq\; (K_0 \vee K_1)E_{X|\theta}\,|X - \delta_c(X)| \\ &\leq\; (K_0 \vee K_1)\frac{M_3}{c}. \quad \text{[by Theorem 5.2]} \end{aligned} \tag{25}$$

Combining (24) and (25), the theorem follows from (23). ♣

6 Summary

The main goal of this work was to demonstrate satisfactory frequentist behavior for the implied Bayes solution in a wide variety of problems; we described a large family of scaled prior distributions meeting this goal. Development of these priors guaranteed frequentist validity and at the same time we obtain strong robustness in Bayesian inference with respect to the particular form of the prior distribution.

Acknowledgement: The idea of this work arose in conversation with Joe Eaton and Bill Sudderth. Various ideas were discussed with Larry Brown, Jayanta Ghosh, Herman Rubin and Bill Strawderman. We are very glad to thank all of them.

7 Appendix

Proof of Lemma 3.1: We have to prove that if $M_f(\boldsymbol{t})$, the MGF of $\boldsymbol{X}$, exists and is finite for all $\boldsymbol{t}$ in an open neighbourhood of $\mathbf{0}$, then for any prior sequence $\{\pi_c(.)\}$ with $\pi(.)$ satisfying the gradient condition

$$\frac{||\nabla\pi(\boldsymbol{\theta})||}{\pi(\boldsymbol{\theta})} \leq k \qquad \forall\, \boldsymbol{\theta}\in \boldsymbol{R}^p, \tag{26}$$

there exist positive constants $c_0(\geq 1)\,, d_1, d_2$, and d_3 such that whenever $c \geq c_0$,

$$c\,|\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - f(\boldsymbol{x}-\boldsymbol{\theta})| \leq \left(d_1 + d_2 G(\boldsymbol{x}-\boldsymbol{\theta}) + d_3 e^{G(\boldsymbol{x}-\boldsymbol{\theta})\frac{k}{c_0}}\right) f(\boldsymbol{x}-\boldsymbol{\theta}) \tag{27}$$

where $G(\boldsymbol{z}) = \sum_{i=1}^{p} |z_i|$.

We prove (27) by using the following steps: first in Step I, we get bounds on posterior densities by obtaining bounds on prior densities and the fact that the MGF of $\boldsymbol{X}$ exists; then in Step II, by some applications of the *Dominated Convergence Theorem* (DCT) on the bounds obtained in Step I, we prove the result.

Step I: When $\pi(.)$ satisfies (26), it immediately follows that

$$\frac{||\nabla\pi_c(\boldsymbol{\theta})||}{\pi_c(\boldsymbol{\theta})} \leq \frac{k}{c} \quad \forall\, \boldsymbol{\theta}\in \boldsymbol{R}^p \tag{28}$$

Now, expanding $\log\pi_c(\boldsymbol{\theta})$ around $\boldsymbol{x}$, by using multivariate Taylor expansion, we get

$$\log\pi_c(\boldsymbol{\theta}) = \log\pi_c(\boldsymbol{x}) + (\boldsymbol{\theta}-\boldsymbol{x})'\frac{\nabla\pi_c(\boldsymbol{\theta}-\boldsymbol{x}^\star)}{\pi_c(\boldsymbol{\theta}-\boldsymbol{x}^\star)} \qquad \boldsymbol{\theta}\in\boldsymbol{R}^p,\ \boldsymbol{x}\in\boldsymbol{R}^p, \tag{29}$$

where $\boldsymbol{x}^\star$ lies between $\boldsymbol{\theta}$ and $\boldsymbol{x}$. Combining (28) and (29) and using Cauchy–Schwartz inequality we get

$$\left|\log\frac{\pi_c(\boldsymbol{\theta})}{\pi_c(\boldsymbol{x})}\right| \leq ||\boldsymbol{\theta}-\boldsymbol{x}||\frac{||\nabla\pi_c(\boldsymbol{\theta}-\boldsymbol{x}^\star)||}{\pi_c(\boldsymbol{\theta}-\boldsymbol{x}^\star)} \leq ||\boldsymbol{\theta}-\boldsymbol{x}||\frac{k}{c}. \tag{30}$$

Since $||\boldsymbol{z}|| \leq G(\boldsymbol{z})\,\forall \boldsymbol{z}$, we get on straightforward calculation from (30) that

$$\pi_c(\boldsymbol{x})e^{-\frac{k}{c}G(\boldsymbol{\theta}-\boldsymbol{x})} \leq \pi_c(\boldsymbol{\theta}) \leq \pi_c(\boldsymbol{x})e^{\frac{k}{c}G(\boldsymbol{\theta}-\boldsymbol{x})} \ \ \forall\, \boldsymbol{\theta},\boldsymbol{x}\in\boldsymbol{R}^p. \tag{31}$$

Notice that since $M_f(\boldsymbol{t})$ exists and is finite in an open neighbourhood of $\mathbf{0}$, the same is true of $M_f^\star(\boldsymbol{t})$, the MGF of $\boldsymbol{X}^\star = (|X_1|,\cdots,|X_p|)'$. So there exists $t_0(>0)$ such that $M_f^\star(\boldsymbol{t})$ exists and is finite for all $||\boldsymbol{t}|| < t_0$. Therefore, if we choose $c_0 = \frac{k}{t_0}+1$, then whenever $c \geq c_0$

$$M_f^\star\left(\frac{k}{c}\right) = \int_{\boldsymbol{R}^p} e^{\frac{k}{c}G(\boldsymbol{z})} f(\boldsymbol{z})\,d\boldsymbol{z} \tag{32}$$

exists and is finite and hence by using (31) we get the following bounds on the posterior density $\pi_c(\boldsymbol{\theta}|\boldsymbol{x})$:

$$\frac{e^{-\frac{k}{c}G(\boldsymbol{\eta})}f(\boldsymbol{\eta})}{\int_{\boldsymbol{R}^p} e^{\frac{k}{c}G(\boldsymbol{z})}f(\boldsymbol{z})\,d\boldsymbol{z}} \leq \pi_c(\boldsymbol{\theta}|\boldsymbol{x}) \leq \frac{e^{\frac{k}{c}G(\boldsymbol{\eta})}f(\boldsymbol{\eta})}{\int_{\boldsymbol{R}^p} e^{-\frac{k}{c}G(\boldsymbol{z})}f(\boldsymbol{z})\,d\boldsymbol{z}} \quad \forall\,\boldsymbol{\eta}\in\boldsymbol{R}^p, \tag{33}$$

whenever $c \geq c_0$, where $\boldsymbol{\eta} = \boldsymbol{x} - \boldsymbol{\theta}$.

Step II: Using (33) we get

$$-h_c(\boldsymbol{\eta}) \leq \pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - f(\boldsymbol{x} - \boldsymbol{\theta}) \leq g_c(\boldsymbol{\eta}), \tag{34}$$

for any $\boldsymbol{x}$ and $\boldsymbol{\theta}$, and for any $c \geq c_0$, where the nonnegative functions $g_c(.)$ and $h_c(.)$ are defined as

$$g_c(\boldsymbol{\eta}) = \frac{e^{\frac{k}{c}G(\boldsymbol{\eta})} f(\boldsymbol{\eta})}{u_f(c)} - f(\boldsymbol{\eta}), \qquad u_f(c) = \int_{\boldsymbol{R}^p} e^{-\frac{k}{c}G(\boldsymbol{z})} f(\boldsymbol{z})\, d\boldsymbol{z}$$

$$\text{and } h_c(\boldsymbol{\eta}) = f(\boldsymbol{\eta}) - \frac{e^{-\frac{k}{c}G(\boldsymbol{\eta})} f(\boldsymbol{\eta})}{w_f(c)}, \qquad w_f(c) = \int_{\boldsymbol{R}^p} e^{\frac{k}{c}G(\boldsymbol{z})} f(\boldsymbol{z})\, d\boldsymbol{z}.$$

By using (34) and the fact that $g_c(.)$ and $h_c(.)$ are nonnegative, we get

$$c\left|\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - f(\boldsymbol{x} - \boldsymbol{\theta})\right| \;\leq\; c \max\{g_c(\boldsymbol{\eta}), h_c(\boldsymbol{\eta})\} \;\leq\; c g_c(\boldsymbol{\eta}) + c h_c(\boldsymbol{\eta}) \tag{35}$$

Finally, to obtain (27), we will give bounds on $cg_c(\boldsymbol{\eta})$ and $ch_c(\boldsymbol{\eta})$.

Notice that we can rewrite $cg_c(\boldsymbol{\eta})$ as

$$\begin{aligned} cg_c(\boldsymbol{\eta}) &= cf(\boldsymbol{\eta})\left(\frac{e^{\frac{k}{c}G(\boldsymbol{\eta})}}{u_f(c)} - 1\right) \\ &= c\frac{f(\boldsymbol{\eta})}{u_f(c)}\left(e^{\frac{k}{c}G(\boldsymbol{\eta})} - 1\right) + cf(\boldsymbol{\eta})\left(\frac{1 - u_f(c)}{u_f(c)}\right). \end{aligned} \tag{36}$$

Using the fact that $f(\boldsymbol{z})e^{-\frac{k}{c}G(\boldsymbol{z})} \leq f(\boldsymbol{z})\ \forall \boldsymbol{z}$ and $f(\boldsymbol{z})e^{-\frac{k}{c}G(\boldsymbol{z})} \nearrow f(\boldsymbol{z})$ as $c \nearrow \infty$ for any fixed $\boldsymbol{z}$, it follows from the Dominated Convergence Theorem that $0 < u_f(c) \nearrow 1$ as $c \nearrow \infty$. Also we know that $c\left(e^{\frac{k}{c}G(\boldsymbol{\eta})} - 1\right) \downarrow c$ and therefore we get

$$\frac{c}{u_f(c)}\left(e^{\frac{k}{c}G(\boldsymbol{\eta})} - 1\right) \downarrow c (\geq 1). \tag{37}$$

This handles the first term in (36). Next, notice that whenever $c \geq 1$,

$$\begin{aligned} c(1 - u_f(c)) &= \int_{\boldsymbol{R}^p} c\left(1 - e^{-\frac{k}{c}G(\boldsymbol{z})}\right) f(\boldsymbol{z})\, d\boldsymbol{z} \\ &\leq \int_{\boldsymbol{R}^p} c\left(\frac{k}{c}G(\boldsymbol{z})\right) f(\boldsymbol{z})\, d\boldsymbol{z} \qquad [\text{since } 1 - e^{-x} \leq |x|\ \forall\, x \in \boldsymbol{R}] \\ &= k\int_{\boldsymbol{R}^p} G(\boldsymbol{z}) f(\boldsymbol{z})\, d\boldsymbol{z} \;=\; m_1 \text{ (a fixed constant)} \end{aligned} \tag{38}$$

So combining (38) and (37), we get from (36) that whenever $c \geq c_0 (\geq 1)$

$$cg_c(\boldsymbol{\eta}) \leq \frac{c_0}{u_f(c_0)}\left(e^{G(\boldsymbol{\eta})\frac{k}{c_0}} - 1\right) f(\boldsymbol{\eta}) + \frac{m_1}{u_f(c_0)} f(\boldsymbol{\eta}) \qquad \forall\, \boldsymbol{\eta} \in \boldsymbol{R}^p, \tag{39}$$

A similar argument shows that

$$ch_c(\boldsymbol{\eta}) \;\leq\; kG(\boldsymbol{\eta})f(\boldsymbol{\eta}) + c_0\left(w_f(c_0) - 1\right) f(\boldsymbol{\eta}). \tag{40}$$

(27) now follows by combining (35), (39), and (40).

Proof of Lemma 4.1: We have to prove that there exist fixed positive constants L_1 and U_1, such that

$$S_{\pi_c}^{-\frac{L_1}{c}}(\boldsymbol{x}) \subseteq S_{\pi_c}(\boldsymbol{x}) \subseteq S_{\pi_c}^{\frac{U_1}{c}}(\boldsymbol{x}), \quad \forall\, \boldsymbol{x} \in \boldsymbol{R}^p,\ \forall\, c \geq c_2, \tag{41}$$

for some $c_2\,(\geq 1)$.

The following simple lemmas will be used in the proof; their proofs are omitted.

Lemma **7.1** $z\varphi^p(z) \leq \frac{1}{\sqrt{p}}\varphi^p(\frac{1}{\sqrt{p}})\ \forall\ z \in \boldsymbol{R}$.

Lemma **7.2** *For any* $p \geq 1$, $V_p'(r) \uparrow r$.

In the following, we supply the proof of the right side inclusion of (41), i.e., we prove that there exist constants U_1 and c_2 such that $S_{\pi_c}(\boldsymbol{x}) \subseteq S_{\pi_c}^{\frac{U_1}{c}}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \boldsymbol{R}^p$ whenever $c \geq c_2$. The proof of the left inclusion is completely analogous and is therefore omitted.

Notice that for a given $\varepsilon > 0$, $S_{\pi_c}(\boldsymbol{x}) \subseteq S_{\pi_c}^{\varepsilon}(\boldsymbol{x})$ if and only if $\int_{S_{\pi_c}^{\varepsilon}(\boldsymbol{x})} \pi_c(\boldsymbol{\theta}|\boldsymbol{x})\, d\boldsymbol{\theta} \geq 1-\alpha$. So, in order to prove the required result, we simply show that if $\varepsilon = \frac{U_1}{c}$, for suitable choice of positive constants U_1 and $c_2(\geq 1)$, then $\int_{S_{\pi_c}^{\varepsilon}(\boldsymbol{x})} \pi_c(\boldsymbol{\theta}|\boldsymbol{x})\, d\boldsymbol{\theta} \geq 1-\alpha$ for all $c \geq c_2$.

Applying Theorem 3.1 and Theorem 3.2, we get respectively

$$\sup_{\boldsymbol{x}} \int_{\boldsymbol{R}^p} |\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})|\, d\boldsymbol{\theta} \ \leq\ \frac{M_0}{c} \quad \forall\, c \geq c_0, \tag{42}$$

$$\text{and} \quad \sup_{\boldsymbol{x}} \sup_{\boldsymbol{\theta}} |\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})| \ \leq\ \frac{M_1}{c} \quad \forall\, c \geq c_0 \vee kc_1, \tag{43}$$

where c_0, c_1, M_0, M_1 are as in Theorem 3.1 and Theorem 3.2.

Consider $\varepsilon > 0$ and any $c \geq c_0 \vee kc_1$; let

$$\begin{aligned} \delta(\varepsilon) &= \int_{S_{\pi_0}^{\frac{\varepsilon}{2}}(\boldsymbol{x})} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})d\boldsymbol{\theta} - (1-\alpha) \\ &= \int_{S_{\pi_0}^{\frac{\varepsilon}{2}}(\boldsymbol{x})} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})d\boldsymbol{\theta} - \int_{S_{\pi_0}(\boldsymbol{x})} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})d\boldsymbol{\theta}; \end{aligned}$$

then we have

$$\begin{aligned} \int_{S_{\pi_c}^{\varepsilon}(\boldsymbol{x})} \pi_c(\boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{\theta} &\geq \int_{S_{\pi_c}^{\varepsilon}(\boldsymbol{x})} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})d\boldsymbol{\theta} - \int_{S_{\pi_c}^{\varepsilon}(\boldsymbol{x})} |\pi_c(\boldsymbol{\theta}|\boldsymbol{x}) - \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})|d\boldsymbol{\theta} \\ &\geq \int_{S_{\pi_c}^{\varepsilon}(\boldsymbol{x})} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})d\boldsymbol{\theta} - \frac{M_0}{c} \quad \text{[by (42)]} \end{aligned}$$

$$= 1-\alpha+\delta(\varepsilon)+\int_{S^{\varepsilon}_{\pi_c}(\boldsymbol{x})} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})$$

$$- \int_{S^{\frac{\varepsilon}{2}}_{\pi^0}(\boldsymbol{x})} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})d\boldsymbol{\theta} - \frac{M_0}{c} \tag{44}$$

Observe now that

$$\delta(\varepsilon) = \int_{k^0_\alpha-\frac{\varepsilon}{2}\le\varphi_p(\boldsymbol{x}-\boldsymbol{\theta})<k^0_\alpha} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$= \int_{k^0_\alpha-\frac{\varepsilon}{2}\le\varphi_p(\boldsymbol{\theta})<k^0_\alpha} \varphi_p(\boldsymbol{\theta})d\boldsymbol{\theta} \ (\ge 0 \text{ for any } \varepsilon>0) \tag{45}$$

So, by (44) we get

$$\int_{S^{\varepsilon}_{\pi_c}(\boldsymbol{x})} \pi_c(\boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{\theta} \ge 1-\alpha+\delta(\varepsilon)+\int_{S^{\varepsilon}_{\pi_c}(\boldsymbol{x})} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta}) - \int_{S^{\frac{\varepsilon}{2}}_{\pi^0}(\boldsymbol{x})} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})d\boldsymbol{\theta} - \frac{M_0}{c}$$

$$= 1-\alpha+\delta(\varepsilon)+\int_{\substack{\pi_c(\boldsymbol{\theta}|\boldsymbol{x})\ge k^0_\alpha-\varepsilon\\ \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})<k^0_\alpha-\frac{\varepsilon}{2}}} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$- \int_{\substack{\pi_c(\boldsymbol{\theta}|\boldsymbol{x})< k^0_\alpha-\varepsilon\\ \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})\ge k^0_\alpha-\frac{\varepsilon}{2}}} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})d\boldsymbol{\theta} - \frac{M_0}{c}$$

$$= 1-\alpha+\delta(\varepsilon) - \int_{\substack{\pi_c(\boldsymbol{\theta}|\boldsymbol{x})< k^0_\alpha-\varepsilon\\ \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})\ge k^0_\alpha-\frac{\varepsilon}{2}}} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})d\boldsymbol{\theta} - \frac{M_0}{c}$$

$$\ge 1-\alpha+\delta(\varepsilon) - \int_{|\pi_c(\boldsymbol{\theta}|\boldsymbol{x})-\varphi_p(\boldsymbol{x}-\boldsymbol{\theta})|>\frac{\varepsilon}{2}} \varphi_p(\boldsymbol{x}-\boldsymbol{\theta})d\boldsymbol{\theta} - \frac{M_0}{c}. \tag{46}$$

Therefore, for any $\frac{\varepsilon}{2} > \frac{M_1}{c}$, we have from (43) and (46) that

$$\int_{S^{\varepsilon}_{\pi_c}(\boldsymbol{x})} \pi_c(\boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{\theta} \ge 1-\alpha+\delta(\varepsilon)-\frac{M_0}{c}. \tag{47}$$

So, to show the right-side inclusion of (41), it is enough to prove that for some constant $U_1 \ge 2M_1$, $\delta\left(\frac{U_1}{c}\right) \ge \frac{M_0}{c}$.

Let $\gamma>0$ be the unique solution of $\varphi^p(\chi_\alpha(p)) - \varphi^p(\chi_\alpha(p)+\gamma) = \frac{\varepsilon}{2}$. Then, we get from (45) that

$$\delta(\varepsilon) = \int_{k^0_\alpha-\frac{\varepsilon}{2}\le\varphi_p(\boldsymbol{\theta})<k^0_\alpha} \varphi_p(\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$\ge \Big(V_p(\sqrt{p}(\chi_\alpha(p)+\gamma)) - V_p(\sqrt{p}\chi_\alpha(p))\Big)(k^0_\alpha-\frac{\varepsilon}{2})$$

$$= \sqrt{p}\gamma V_p'(\sqrt{p}(\chi_\alpha(p)+\gamma^\star))(k^0_\alpha-\frac{\varepsilon}{2}), \tag{48}$$

where the last step uses Taylor expansion and $\gamma^\star$ is between 0 and γ. Again, by definition of γ,

$$\begin{aligned}
\frac{\varepsilon}{2} &= \varphi^p(\chi_\alpha(p)) - \varphi^p(\chi_\alpha(p)+\gamma) \\
&= p\gamma(\chi_\alpha(p)+\gamma^{\star\star})\varphi^p(\chi_\alpha(p)+\gamma^{\star\star}) \quad \text{[by Taylor expansion, with } 0<\gamma^{\star\star}<\gamma] \\
&\leq \gamma p \frac{1}{\sqrt{p}}\varphi^p(\frac{1}{\sqrt{p}}) \qquad \text{[by Lemma 7.1]} \\
\Rightarrow \gamma &\geq \frac{\varepsilon}{2}\frac{1}{\sqrt{p}}\varphi^{-p}(\frac{1}{\sqrt{p}}) \qquad (49)
\end{aligned}$$

Combining (48) and (49), we get

$$\begin{aligned}
\delta(\varepsilon) &\geq \frac{\varepsilon}{2}\varphi^{-p}(\frac{1}{\sqrt{p}})V_p'(\sqrt{p}(\chi_\alpha(p)+\gamma^\star))(k_\alpha^0-\frac{\varepsilon}{2}) \\
&\geq \frac{\varepsilon}{2}\varphi^{-p}(\frac{1}{\sqrt{p}})V_p'(\sqrt{p}\,\chi_\alpha(p))(k_\alpha^0-\frac{\varepsilon}{2}) \quad \text{[by Lemma 7.2]} \\
&= \frac{\varepsilon}{2}(k_\alpha^0-\frac{\varepsilon}{2})\lambda \qquad \left[\text{by denoting } \lambda = \varphi^{-p}(\frac{1}{\sqrt{p}})V_p'(\sqrt{p}\,\chi_\alpha(p))\right]. \qquad (50)
\end{aligned}$$

Now choose $U_1 = 2(M_1+1)\vee 2\left(\frac{M_0+1}{\lambda k_\alpha^0}\right)$; clearly, $U_1 > 2M_1$; now using (50), we get

$$\begin{aligned}
\delta\left(\frac{U_1}{c}\right) &\geq \frac{U_1}{2c}k_\alpha^0\lambda - \frac{U_1^2}{4c^2}\lambda \\
&\geq \frac{M_0+1}{c} - \frac{1}{c^2}\left(\frac{M_0+1}{\lambda k_\alpha^0}+M_1+1\right)^2\lambda \\
&\geq \frac{M_0}{c} \qquad \forall\, c \geq c_{01} = \left(\frac{M_0+1}{\lambda k_\alpha^0}+M_1+1\right)^2\lambda. \qquad (51)
\end{aligned}$$

So, with the special choice of $\varepsilon = \frac{U_1}{c}$ ($> \frac{2M_1}{c}$, as is required from before), we get from (47) that

$$\begin{aligned}
\int_{S_{\pi_c}^\varepsilon(\boldsymbol{x})} \pi_c(\boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{\theta} &\geq 1-\alpha+\delta(\varepsilon)-\frac{M_0}{c} \quad \text{[for all } c\geq c_0\vee kc_1] \\
&\geq 1-\alpha \qquad \text{[for all } c\geq c_2 = c_{01}\vee c_0\vee kc_1].
\end{aligned}$$

Hence,

$$S_{\pi_c}(\boldsymbol{x}) \subseteq S_{\pi_c}^{\frac{U_1}{c}}(\boldsymbol{x}) \quad \forall \boldsymbol{x}\in \boldsymbol{R}^p \quad \forall c\geq c_2. \qquad (52)$$

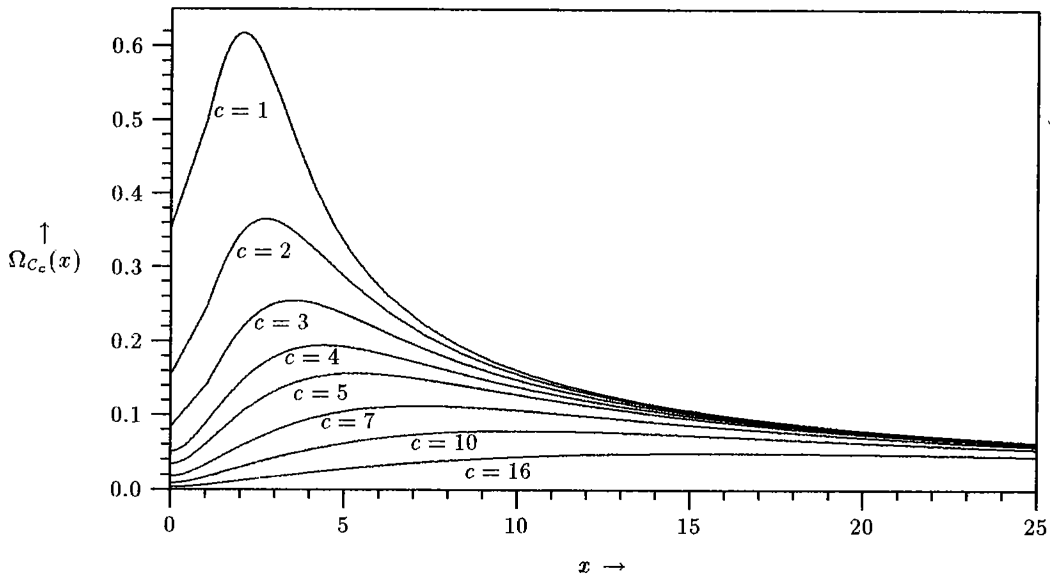

Figure 1: Plots of $\Omega_{\mathcal{C}_c}(x)$ as functions of x for different values of c

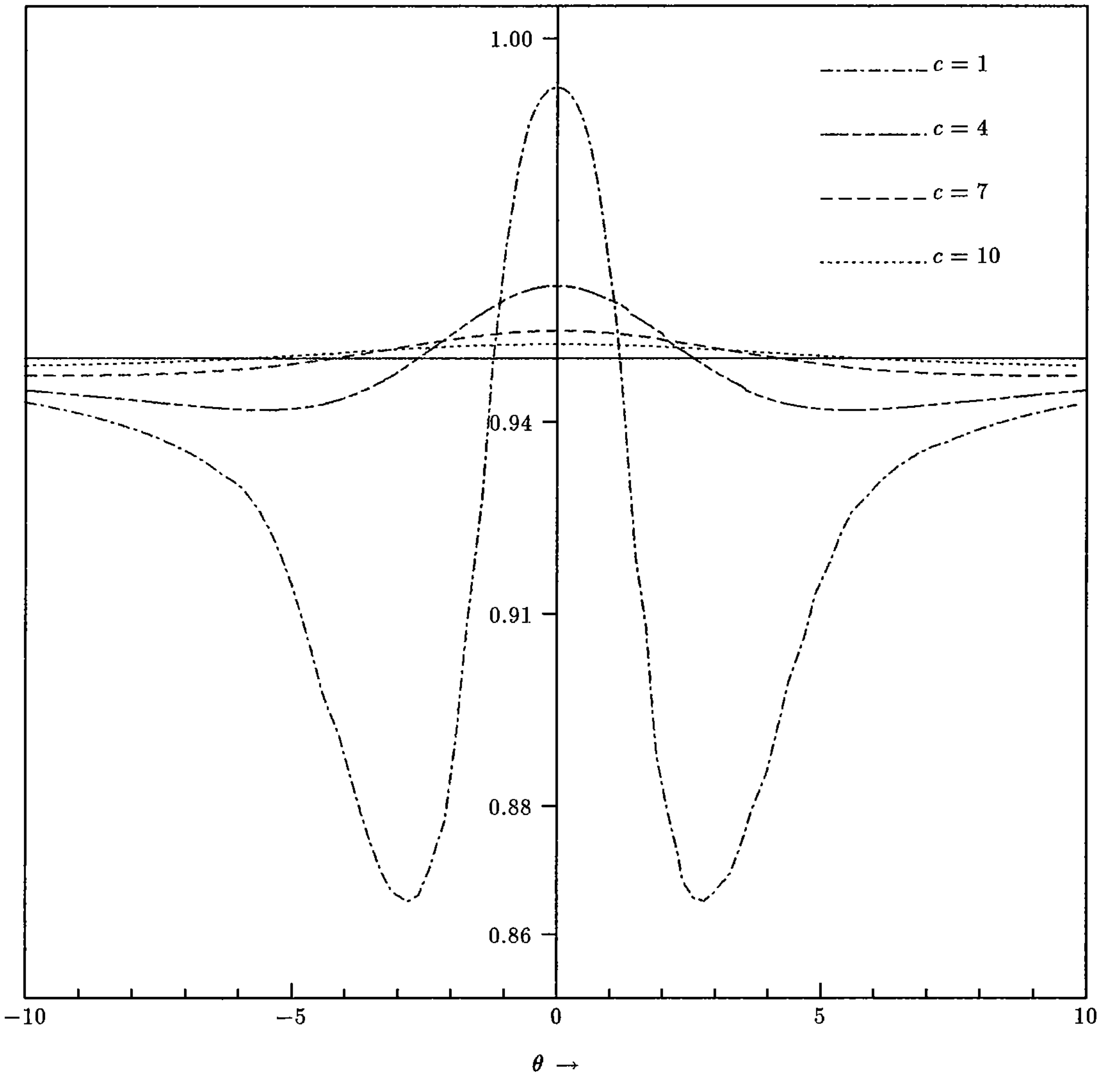

Figure 2: Plots of $\Psi_{\mathcal{C}_c}(\theta)$ for $c = 1, 4, 7$ and 10

Table 1:

c	1	2	3	4	5	7	10	16
$\Omega^\star_{\mathcal{C}_c}$	0.61854	0.36599	0.25535	0.19489	0.15721	0.11311	0.07949	0.04979

Table 2:

c	1	2	3	4	5	7	10	16
$\Psi^\star_{\mathcal{C}_c}$	0.86523	0.91980	0.93571	0.94191	0.94483	0.94735	0.94882	0.94988

Table 3:

c	1	2	3	4	5	7	10	16
$H^\star_{\mathcal{C}_c}$	0.82399	0.48623	0.33277	0.25020	0.20003	0.14301	0.09999	0.05812

References

[1] Berger, J. (1980), "A robust generalized Bayes estimator and confidence region for a multivariate normal mean ," *Ann. Statist.*, **8** , pp. 716-761.

[2] Berger, J. (1985), "Statistical Decision Theory and Bayesian Analysis," 2nd ed., New York: Springer,.

[3] Berger, J. and Robert, C. (1990), "Subjective hierarchical Bayes estimation of a multivariate normal mean: on the frequentist interface," *Ann. Statist.*, **18**, pp. 617-651.

[4] Bickel, P.J. and Yahav, J.A. (1969), "Some contributions to the asymptotic theory of Bayes solutions," *Z. Wahrsch. verw. Geb.*, **11**, pp. 257-276.

[5] Brown, L.D. (1966), "On the admissibility of invariant estimators of one or more location parameters ," *Ann. Math. Statist.*, **37**, pp. 1087-1136.

[6] Brown, L.D. (1971), "Admissible estimators, recurrent diffusions, and insoluble boundary value problems," *Ann. Math. Statist.*, **42**, pp. 855-904.

[7] Brown, L.D. and Hwang, J.T. (1982), "A unified admissibility proof ," in *Statistical Decision Theory and Related Topics III*,S.S. Gupta and J. Berger (eds.), New York: Academic Press.

[8] Casella, G. and Berger, R.L. (1987), "Reconciling Bayesian and frequentist evidence in the one-sided testing problem," *J. Amer. Statist. Assoc.*, **82**, pp. 106-111.

[9] DasGupta, A. and Rubin, H. (1986), "Bayesian estimation subjeect to minimaxity of the mean of a multivvariate normal distribution in the case of a common unknown variance: a case for Bayesian robustness," *Statistical Decision Theory and Related Topics*, S.S. Gupta and J. Berger (eds.), Vol.2, pp. 325-346, New York: Springer Verlag

[10] DasGupta, A. and Studden, W.J. (1989), "Frequentist behavior of robust Bayes estimates of normal means," *Statist. and Decisions*, **7**, pp. 333-361.

[11] Dugundji, J. (1975), *Topology* , London: Prentice Hall.

[12] Efron, B. and Morris, C. (1971), "Limiting the risk of Bayes and empirical Bayes estimators—Part I: The Bayes case ," *J. Amer. Statist. Assoc.*, **67**, pp. 130-139.

[13] Ghosh, M. (1992), "Hierarchical and Empirical Bayes multivariate estimation," in *IMS Lecture Notes*, M. Ghosh and P.K. Pathak (eds.) pp. 151-177.

[14] Ghosh, J.K., Mukerjee, R. (1991), "Characterization of priors under which Bayesian and frequentist Bartlett corrections are equivalent in the multiparameter case," *J. Multivariate Anal.*, **38** no. 2, pp. 385-393.

[15] Ghosh, J.K., Sinha, B.K., and Joshi, S.N. (1982), "Expansions for posterior probability and integrated Bayes risk ," *Statistical Decision Theory and Related Topics III*,S.S. Gupta and J. Berger (eds.), New York: Academic Press.

[16] Hartigan, J.A. (1966) "Note on the confidence prior of Welch and Peers," *J. Roy. Statist. Soc. Ser. B*, **28**, pp. 55-56.

[17] Ibragimov, I.A. and Khas'minsky, R.Z. (1972), "Asymptotic behavior of statistical estimators. II. Limit theorems for the a posteriori density and Bayes estimators," *Theory of Prob. and Its Appl*, **18**, pp. 76-91.

[18] Ibragimov, I.A. and Khas'minsky, R.Z. (1981), *Statistical Estimation: Asymptotic Theory*, New York: Springer.

[19] Johnson, R.A. (1967), "An asymptotic expansion for posterior distributions ," *Ann. Math. Statist.*, **38**, pp. 1899-1906.

[20] Kadane, J.B. and Chuang, D.T. (1978), "Stable Decision problems," *Ann. Statist.*, **6**, pp. 1095-1110.

[21] Lehmann, E.L. (1983), *Theory of Point Estimation*, New York: Wiley.

[22] Lehmann, E.L. (1986), *Testing Statistical Hypotheses*, New York: Wiley.

[23] Meeden, G. and Isaacson, D. (1977), "Approximate behavior of the posterior distribution for a large observation ," *Ann. Statist.*, **5**, pp. 899-905.

[24] Rubin, D.B. (1984), "Bayesianly justifiable and relevant frequency calculations for applied statistician." *Ann. Statist.*, **12**, pp. 1151-1172.

[25] Stein, C. (1985), "On the coverage probability of confidence sets based on a prior distribution," *Sequential Meth. Statist.*, **16**, pp. 485-514Banach Center Publications.

[26] Strawderman, W.E. (1971), "Proper Bayes minimax estimators of the multivariate normal mean ," *Ann. Math. Statist.*, **42** , pp. 385-388.

[27] Strawderman, W.E. and Cohen, A. (1971), "Admissibility of estimators of the mean vector of a multivariate normal distribution with quadratic loss ," *Ann. Math. Statist.*, **42**, pp. 270-296.

[28] Umbach, D. (1978), "On the approximate behavior of the posterior distribution for an extreme multivariate observation," *J. Mult. Anal.*, **8**, pp. 518-531.

[29] Walker, A.M. (1969), "On the asymptotic behavior of posterior distributions," *J. Roy. Statist. Soc. Ser. B*, **31**, pp. 80-88.

[30] Welch, B. L. and Peers, H. W. (1963) "On formulae for confidence points based on integrals of weighted likelihoods," *J. Roy. Statist. Soc. Ser. B*, **25**, pp. 318-329.

[31] Woodroofe, M. (1976), "Frequentist properties of Bayesian sequential tests ," *Biometrika*, **63**, pp. 101-110.

[32] Woodroofe, M. (1989), "Very weak expansions for sequentially designed experiments: linear models," *Ann. Statist.*, **17**, pp. 1087-1102.

[33] Zidek, J.V. (1970), "Sufficient conditions for the admissibility under squared–error loss of formal Bayes estimators," *Ann. Math. Statist.*, **41**, pp. 446-456.