The Annals of Statistics
1996, Vol. 24, No. 1, 174–195

LOCAL SENSITIVITY OF POSTERIOR EXPECTATIONS[1]

BY PAUL GUSTAFSON

University of British Columbia

We investigate the degree to which posterior expectations are sensitive to prior distributions, using a local method based on functional differentiation. Invariance considerations suggest a family of norms which can be used to measure perturbations to the prior. The sensitivity measure is seen to depend heavily on the choice of norm; asymptotic results suggest which norm will yield the most useful results in practice. We find that to maintain asymptotically sensible behaviour, it is necessary to reduce the richness of the class of prior perturbations as the dimension of the parameter space increases. Jeffreys' prior is characterized as the prior to which inference is least sensitive.

1. Introduction. Statistical inference always requires various assumptions. The assumptions are not always directly verifiable; hence statisticians are interested in the degree to which inferences are sensitive to the assumptions. In Bayesian statistics there is a good deal of interest in the degree to which inferences are sensitive to prior distributions. The common approach to assessing sensitivity is to measure the size of the class of posteriors (or perhaps just a particular posterior quantity) that arises from a specified class of priors. This is referred to as global sensitivity analysis; see Berger (1990, 1994), Lavine (1991a, b) and Wasserman (1992b) for reviews. The fact that global analyses often entail a large computational burden has lead to a recent surge of interest in local sensitivity analysis; see Basu, Jammalamadaka and Liu (1993), Berger (1986), Cuevas and Sanz (1988), Delampady and Dey (1994), Dey and Birmiwal (1994), Diaconis and Freedman (1986), Gelfand and Dey (1991), Gustafson and Wasserman (1995), Ruggeri and Wasserman (1993, 1995), Sivaganesan (1993) and Srinivasan and Truszczynska (1990). The idea of a local analysis is to examine the rate at which the posterior changes, relative to the prior. This paper examines several methods of assessing the local sensitivity of posterior expectations.

Often it might be desirable to assess the sensitivity of the posterior as a whole rather than the sensitivity of particular posterior expectations. However, as discussed in Gustafson and Wasserman (1995), it is very difficult to do so in a local manner, without obtaining results which are asymptotically unsatisfactory. In particular, local measures of the overall posterior sensitiv-

Received November 1993; revised August 1995.

[1] Research supported by a Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada.

AMS 1991 *subject classifications*. Primary 62F15; secondary 62F35.

Key words and phrases. Classes of probabilities, Fréchet derivative, Jeffreys' prior, local sensitivity, L^p-norm, robustness.

ity tend to diverge to infinity as the sample size grows. This is in contrast to our knowledge that, starting from different priors, posteriors tend to agree as the data accumulate. Here our attention lies with the sensitivity of posterior expectations. In this setting we can obtain sensitivity measures with more satisfactory asymptotic behaviour.

Both global and local analyses depend on having some quantified notion of the difference between two prior distributions. In the global case, one often starts with a base prior distribution and then forms the class of all priors with limited difference, or distance, from the base prior. As detailed in Section 2, we consider a continuum of ways to quantify distance from the base prior. Different quantifications can lead to different results, most noticeably in an asymptotic sense. We also investigate various properties of the classes of prior defined by our notion of distance.

Usually, we wish to consider classes of priors that are too large to be indexed by a finite-dimensional parameter, to allow for a wide variety of prior distributions, especially in terms of tail behaviour. This necessitates having an infinite-dimensional notion of derivative to measure the rate at which posterior expectations change with respect to the prior. In particular, we consider Fréchet derivatives and their norms, applied to the mapping from prior distribution to posterior expectation. Implementation of this methodology is described in Section 3. When the derivative (or more precisely its norm) is large, inference is sensitive to the prior in that small changes in the prior can cause large changes in the posterior mean. We provide some simple examples, as a first step toward elucidating the properties of these posterior derivatives.

In almost any statistical context, it is illuminating to consider what happens when the sample size grows to infinity. Section 4 contains asymptotic results suggesting which versions of the posterior derivative are most useful in practice. Additionally, these results provide an asymptotic answer to the question of which prior leads to the least sensitive inference. This is a previously unexplored starting point for the determination of noninformative priors.

For the sake of clarity, proofs of the results are relegated to Section 6.

2. Prior perturbations.

2.1. *Linear perturbations.* We start with a parameter space, Θ, assumed to be a subset of $\mathbb{R}^k$, for some k. Additionally, let $\mathscr{B}$ be a σ-field of subsets of Θ and let μ be a σ-finite measure on $(\Theta, \mathscr{B})$. Attention is restricted to probability measures on $(\Theta, \mathscr{B})$ which are absolutely continuous with respect to μ. Let π be a probability density with respect to $(\Theta, \mathscr{B}, \mu)$ (inducing probability measure Π). We refer to π as the base prior or base density. We assume π has support Θ (else we could reduce Θ). Along the lines of most Bayesian robustness investigations, we consider classes of densities which contain π.

A most straightforward way to construct a small class of densities containing π is by perturbation, and the most common form of perturbation is linear. The result of perturbation is represented as w_L:

$$w_L(\cdot; \pi, u) = \pi(\cdot) + u(\cdot). \tag{1}$$

We restrict u to be nonnegative, to ensure that w_L is nonnegative. When w_L is integrable, its normalized version is denoted as w_L^*.

To progress further, we must quantify the magnitude of perturbations. Since π is normalized, any reasonable measure of the size of u should reflect the degree to which w_L^* can differ from π. Some desirable features of a size functional are as follows:

D1. Size($\cdot$) is a norm.

D2. Size($\cdot$) is invariant under a change of dominating measure.

D3. Size($\cdot$) is invariant under one-to-one transformation.

D4. Size(u) $< \infty \Rightarrow w(\cdot, \pi, u)$ is integrable.

D1 is necessary for the study of local sensitivity. D2 and D3 are minimal invariance requirements, D2 being necessary as the perturbations are defined in terms of densities. D4 ensures that members of a class of densities based on the size function are proper.

One possibility is to define

$$\text{size}(u) = \|u/\pi; \Pi\|_p = \begin{cases} \left(\int_\Theta (u/\pi)^p \, d\Pi\right)^{1/p}, & p < \infty, \\ \text{ess sup}_\Theta \, u/\pi, & p = \infty, \end{cases} \tag{2}$$

the L^p norm of u/π with respect to dominating measure Π, for some $p \in [1, \infty]$.

RESULT 1. *For linear perturbations, defining size as in* (2) *satisfies* D1 *through* D4.

Now a class of densities containing π can be defined. In particular, for $p \in [1, \infty]$, let

$$\Gamma_L^p(\Pi; \delta) = \{w_L^*(\cdot; \pi, u): \|u/\pi; \Pi\|_p \le \delta\}, \tag{3}$$

the class of all limited size perturbations of π.

2.2. *Nonlinear perturbations.* There is no compelling reason to study only linear perturbations. Indeed, linear perturbations are usually specified for simplicity. We can generalize (1) by selecting a strictly increasing continuous function $t: [0, \infty) \to [-\infty, \infty)$ and defining

$$w_N(\cdot; \pi, u) = t^{-1}(t(\pi(\cdot)) + u(\cdot)).$$

Restricting u to be nonnegative ensures w_N is nonnegative, since t is increasing. Again, when w_N is integrable, we denote its normalized version by w_N^*.

A basic class of increasing functions is the class of power functions. It is convenient to consider a class of matched power functions. Specifically, we set the function and its first derivative to be 1 at 1. The resulting class can be indexed as $\{t_p\colon p \in [1, \infty)\}$, where

$$t_p(y) = py^{1/p} - (p-1).$$

It is natural to extend to the case $p = \infty$ via the pointwise limit of t_p. Thus we define

$$t_\infty(y) = \log y + 1,$$

giving

$$w_N(\cdot; \pi, u, p) = \begin{cases} \left(\pi^{1/p}(\cdot) + u(\cdot)/p\right)^p, & p < \infty, \\ \pi(\cdot)\exp(u(\cdot)), & p = \infty. \end{cases}$$

Note that when $p = 1$, the nonlinear perturbations are in fact the linear perturbations of the previous section. It turns out that the L^p norm with respect to the dominating measure is a natural way to quantify the size of these perturbations.

RESULT 2. *For nonlinear perturbations based on* $p \in [1, \infty]$, D1 *through* D4 *are satisfied by taking* $\operatorname{size}(u) = \|u; \mu\|_p$.

In analogy to (3), define

$$\Pi_N^p(\Pi; \delta) = \left\{w_N^*(\cdot; \pi, u, p) \colon \|u; \mu\|_p \le \delta\right\}. \tag{4}$$

It is worth noting that satisfying D2 depends crucially on using the same value of p for the power transformation and the norm. While other size functionals may satisfy D1 through D4 in both the linear and nonlinear case, the choices presented here appear to be the only natural quantifications. As is commonly the case when working with L^p norms, principally we are interested in the values 1, 2 and ∞ for p. Much of the treatment, however, allows for arbitrary values of p.

2.3. *Properties.* Now we wish to make connections to some familiar classes of probability measures. The most common classes in both frequentist and Bayesian robustness studies are based on ε-contamination [Berger (1990); Berger and Berliner (1986)]. The ε-contamination of Π by Π^* is the mixture distribution $(1 - \varepsilon)\Pi + \varepsilon\Pi^*$.

RESULT 3. *The class* $\Gamma_L^1(\Pi; \delta) = \Gamma_N^1(\Pi; \delta)$ *is the class of all* ε*-contaminations of* π, *where the contaminants are absolutely continuous with respect to* μ *and* $\varepsilon \le 1/(1 + (1/\delta))$.

Of course metrics can be used to define classes of probabilities. One such metric, based on densities, is the Hellinger distance

$$d_H(\pi_1, \pi_2) = \left(\int_\Theta \left(\pi_1^{1/2} - \pi_2^{1/2} \right)^2 d\mu \right)^{1/2}.$$

This bears some resemblance to the definition of $p = 2$ nonlinear perturbations. Indeed, we have the following result.

RESULT 4. *$\Gamma_N^2(\Pi; \delta)$ is properly contained in the Hellinger ball of radius $\delta/2$ centred about π.*

A metric that has attracted attention in the Bayesian robustness literature is the density ratio metric [DeRobertis (1978); DeRobertis and Hartigan (1981); Wasserman (1992a)], also defined in terms of densities:

$$d_{\mathrm{DR}}(\pi_1, \pi_2) = \begin{cases} \operatorname*{ess\,sup}_{\theta, \phi \in S(\pi_1)} \log \dfrac{\pi_1(\theta)/\pi_1(\phi)}{\pi_2(\theta)/\pi_2(\phi)}, & S(\pi_1) = S(\pi_2), \\ \infty, & \text{otherwise}, \end{cases}$$

where $S(\pi)$ is the support of π. In turns out that density ratio neighbourhoods are generated when $p = \infty$, using linear or nonlinear perturbations.

RESULT 5. *$\Gamma_L^\infty(\Pi; \delta)$ is identically the density ratio ball of size* $\log(1 + \delta)$ *about π, while $\Gamma_N^\infty(\Pi; \delta)$ is identically the density ratio ball of size δ about π.*

Thus our endpoints $p = 1$ and $p = \infty$ correspond to the familiar ε-contamination class and density ratio class, respectively.

Since the L^p norm with respect to a probability measure satisfies $\|u\|_{p_1} \le \|u\|_{p_2}$ whenever $1 \le p_1 < p_2 \le \infty$, it follows that $\Gamma_L^{p_2}(\Pi; \delta) \subset \Gamma_L^{p_1}(\Pi; \delta)$ whenever $p_1 < p_2$. For the nonlinear classes a limiting result of this form obtains.

RESULT 6. *If $1 \le p_1 < p_2 \le \infty$, $\Gamma_N^{p_2}(\Pi; \delta) \subset \Gamma_N^{p_1}(\Pi; \delta + o(\delta))$.*

This result has repercussions for the local sensitivity measure introduced in the next section.

It is instructive to quantify the notion of richness of a class of perturbations. For a class Γ of probability measures on $(\Theta, \mathscr{B})$ containing Π, define a richness function $R_\Gamma\colon [1, \infty) \to [1, \infty)$ according to

$$R_\Gamma(x) = \sup_{\{A \in \mathscr{B}:\, \Pi(A) \ge 1/x\}} \sup_{\Pi' \in \Gamma} \frac{\Pi'(A)}{\Pi(A)}. \tag{5}$$

The richness function bears some similarity to concentration functions [Cifarelli and Regazzini (1987)]. Note that R_Γ is invariant under transformation. As well, R_Γ is nondecreasing and $R_\Gamma(x) \in [1, x]$. The rate at which R_Γ increases gives an indication of how much mass can be added to small Π-probability sets, and thus reflects the richness of the class.

A way to quantify the rate of increase is via the notion of regular variation. A function h is of regular variation with index of variation p if

$$\lim_{x \to \infty} \frac{h(\lambda x)}{h(x)} = \lambda^p$$

for all $\lambda > 0$. As examples, x^p has index p, as does $x^p \log x$. We say Γ is a regular class if R_Γ is a function of regular variation. For such Γ define $\eta(\Gamma)$ as the index of variation of R_Γ. It is clear that $\eta(\Gamma) \in [0, 1]$ and that $\eta(\{\Pi\}) = 0$, while $\eta(\{\text{all probability measures}\}) = 1$. Furthermore, if $\Gamma_1 \subset \Gamma_2$, then $\eta(\Gamma_1) \le \eta(\Gamma_2)$. There are connections to Wasserman's (1992c) definition of tail richness. In particular, a regular class that is tail rich must have an index greater than zero.

This notion of richness can be applied to the linear and nonlinear classes of priors.

RESULT 7. *If* Π *is absolutely continuous with respect to Lebesgue measure, then*

$$\eta(\Gamma_L^p(\Pi, \delta)) = \frac{1}{p}, \qquad p \in [1, \infty],$$

$$\eta(\Gamma_N^p(\Pi, \delta)) = \begin{cases} 1, & p < \infty, \\ 0, & p = \infty. \end{cases}$$

Thus the linear classes span the whole range of possible richness; however, the nonlinear classes are either minimally or maximally rich.

3. Posterior derivatives.

3.1. *Theory*. Before considering the derivative of a posterior expectation with respect to the prior, the mapping from the prior to the posterior mean must be represented as a mapping between normed linear spaces. Specifically, for either linear or nonlinear perturbations let $\mathscr{U} = \{u \ge 0: \text{size}(u) < \infty\}$ and $\mathscr{V} = \mathbb{R}$. Define $T^g: \mathscr{U} \to \mathscr{V}$ by

$$T^g u = E^x_{w(\cdot;\, \pi,\, u)} g(\theta),$$

the posterior expectation of $g(\theta)$ when the prior is the perturbation of π by u. The norms are the appropriate size functional on $\mathscr{U}$ and absolute value on $\mathscr{V} = \mathbb{R}$. Note that $\mathscr{U}$ is not a proper linear space as its elements are restricted to be everywhere nonnegative; however, it is linear if only nonnegative scalars are admitted. If, for $u_0 \in \mathscr{U}$, there exists a linear functional $\dot{T}(u_0): \mathscr{U} \to \mathscr{V}$ satisfying

$$\left\| T(u_0 + u) - T(u_0) - \dot{T}(u_0) u \right\|_{\mathscr{V}} = o(\|u\|_{\mathscr{U}}),$$

then T is Fréchet differentiable at u_0, with derivative $\dot{T}(u_0)$. In this case, $\dot{T}(u_0)$ is also the derivative in the weak (Gâteaux) sense, satisfying

$$\dot{T}(u_0)u = \lim_{\varepsilon \downarrow 0} \frac{T(u_0 + \varepsilon u) - T(u_0)}{\varepsilon}.$$

Thus $\dot{T}(u_0)u$ is the rate of change of T at u_0 in direction u.

In the present context $\dot{T}^g(0)$ measures the rate of change of $E^x(g(\theta))$ as the prior is perturbed (since $u_0 = 0$ corresponds to the base prior π). Letting L be the likelihood function and using Cov^x to denote posterior covariance, the derivative can be computed and expressed as follows.

RESULT 8. *If L and gL are bounded, then*

$$\dot{T}_L^g(0)u = \mathrm{Cov}_\pi^x\left(g(\theta), \frac{u(\theta)}{\pi(\theta)}\right).$$

RESULT 9. *If L and gL are bounded, then*

$$\dot{T}_N^g(0)u = \mathrm{Cov}_\pi^x\left(g(\theta), \frac{u(\theta)}{\pi^{1/p}(\theta)}\right).$$

In the $p = 1$ case (both linear and nonlinear), the derivative can be rewritten as $(m'/m)(\rho_g' - \rho_g)$, where m and ρ_g are the marginal density of the data and posterior mean of g under the base prior π, while m' and ρ_g' are the analogous quantities when the prior density is u. This form of the expression is familiar, having been derived (under a variety of conditions) and used by Ruggeri and Wasserman (1993), Sivaganesan (1993) and Srinivasan and Truszczynska (1990). Diaconis and Freedman (1986) considered the overall posterior rather than a particular mean and obtained (m'/m) as the derivative norm. For a related diagnostic based on posterior covariance, see Kass, Tierney and Kadane (1989).

Rather than focussing on a particular direction for perturbations to the prior, we consider the worst case direction—that in which the posterior expectation changes most rapidly. Specifically, $\dot{T}_g(0)$ is a linear operator; hence we can compute its norm:

$$\left\|\dot{T}^g(0)\right\| = \sup_{u \in \mathscr{U}} \frac{\left\|\dot{T}^g(0)u\right\|_{\mathscr{V}}}{\|u\|_{\mathscr{U}}} = \sup_{\|u\|_{\mathscr{U}}=1} \left\|\dot{T}^g(0)u\right\|_{\mathscr{V}}.$$

Thus $\|\dot{T}^g(0)\|$ reflects the overall sensitivity of the posterior expectation. Letting π^x be the posterior density on θ and denoting by q the extended real satisfying $q^{-1} + p^{-1} = 1$, the following result obtains.

RESULT 10. *Under the conditions of Result* 8,

$$\left\|\dot{T}_L^g(0)\right\|_p = \max\{\|a_L^+; \Pi\|_q, \|a_L^-; \Pi\|_q\}, \tag{6}$$

where

$$a_L(\theta) = \frac{\big(g(\theta) - \rho_g\big)\pi^x(\theta)}{\pi(\theta)}$$

and $a^+ = \max(a, 0)$, $a^- = -\min(a, 0)$.

RESULT 11. *Under the conditions of Result* 9,

$$\big\|\dot{T}_N^g(0)\big\|_p = \max\{\|a_N^+ ; \mu\|_q, \|a_N^- ; \mu\|_q\}, \tag{7}$$

where

$$a_N(\theta) = \frac{\big(g(\theta) - \rho_g\big)\pi^x(\theta)}{\pi^{1/p}(\theta)}.$$

An immediate corollary is the following result.

RESULT 12. *For* $p \in [1, \infty]$,

$$\big\|\dot{T}_L^g(0)\big\|_p = \big\|\dot{T}_N^g(0)\big\|_p.$$

Note that the derivative norm is subscripted by p to indicate the norm used to measure prior perturbations. The linear and nonlinear classes give the same indication of local sensitivity, despite their differing natures when $p \in (1, \infty)$. In particular recall that for these intermediate values of p, the nonlinear class is richer than the corresponding linear class. The $p = 1$ norm was arrived at by Ruggeri and Wasserman (1993) using a slightly different argument. The $p = \infty$ norm was obtained by Ruggeri and Wasserman (1995) by an entirely different method, namely, differentiation of the endpoints of the global range with respect to the size of the density ratio class.

Note that computationally, some simplification of (6) and (7) occurs. Specifically, when $p = 1$, $\max\{\|a^+\|_\infty, \|a^-\|_\infty\} = \|a\|_\infty$ (with respect to any dominating measure). Additionally, when $p = \infty$, a integrates to 0 with respect to the requisite measure; hence $\|a^+\|_1 = \|a^-\|_1 = (1/2)\|a\|_1$. Only for intermediate values of p is it necessary to compute and compare two norms.

The proofs of Results 10 and 11 give us some indication of the local shape of these classes. If g, L and π are continuous functions of the parameter, then the prior direction in which the norm is achieved is itself a continuous function if $1 < p < \infty$, but discontinuous if $p = 1$ or $p = \infty$. If discontinuous priors are deemed to be inappropriate for the problem at hand, then a choice of p other than 1 or ∞ might be more consistent with the goal of specifying a class of reasonable priors. This sort of concern about the plausibility of priors which achieve extreme inferences over a class has been voiced by Hartigan [(1983), Section 12.8] and Lindley (1992). In this regard, a choice of p which does not correspond to the commonly used cases ($p = 1, \infty$) might lead to a more sensible analysis. For example, in a one-dimensional problem, when $p = 2$ and $g(\theta) = \theta$, the maximizing u in the nonlinear case might be proportional to $(\theta - \rho_\theta)^+ L(\theta)\pi^{1/2}(\theta)$, which is typically smooth, and not very peaked, as $(\theta - \rho_\theta)$ is small when $L(\theta)$ is large.

Results 10 and 11 also lead to approximations to the posterior range. In particular, for either linear or nonlinear classes,

$$\sup_{\pi' \in \Gamma^p(\Pi;\,\delta)} E^x_{\pi'} g(\theta) = E^x_{\pi} g(\theta) + \delta \|a^+\|_q + o(\delta),$$

$$\inf_{\pi' \in \Pi^p(\Pi;\,\delta)} E^x_{\pi'} g(\theta) = E^x_{\pi} g(\theta) - \delta \|a^-\|_q + o(\delta).$$

Note that in a truly linear space (i.e., one admitting negative scalars), we produce approximations to the global range that are symmetric about the target posterior mean. Thus we have the added benefit of being able to represent asymmetry in the local approximation. These approximations to the global range depend on having the stronger (Fréchet) notion of derivative apply.

The form of (6) immediately reveals that the norm is decreasing in p, by Jensen's inequality. Indeed this phenomenon is predicted by Result 6 in the nonlinear case and the remark preceding it in the linear case.

In some contexts the fact that changing the function of interest g can change the value of the norm $\|\dot{T}^g(0)\|$ dramatically might be undesirable. To address this issue, Ruggeri and Wasserman (1995) suggested dividing the norm by the posterior standard deviation of g, which is denoted here by ξ_g. This can be motivated in two distinct ways. First, this is equivalent to computing the norm when the function of interest is the centred and scaled quantity $\xi_g^{-1}(g(\theta) - \rho_g)$, which is invariant under affine transformations of g. Second, $\xi_g^{-1}\|\dot{T}^g(0)\|$ can be interpreted as a ratio of between prior uncertainty about g to within prior uncertainty about g. Particularly, if one entertains a class of priors of size δ, then $\delta\|\dot{T}^g(0)\|$ is approximately the half-width of the range of $E^x g(\theta)$, while $2\xi_g$ is roughly an analogous range of posterior uncertainty about $g(\theta)$, given the base prior. Thus, operationally, sensitivity to the prior might only be a concern when the ratio of $\delta\|\dot{T}^g(0)\|$ to $2\xi_g$ is large. Otherwise, uncertainty about g due to the prior is dwarfed by uncertainty due to the sample. Henceforth the quantity $\xi_g^{-1}\|\dot{T}^g(0)\|$ is referred to as the scaled sensitivity of g.

3.2. *Examples*. Here some simple examples of the posterior derivative norm are given. Throughout a normal model, $X \sim N(\theta, 1/n)$, is assumed. Three prior distributions for θ are considered: normal, Cauchy and log-gamma. The particular hyperparameters are chosen so that the three priors share the same first and third quartiles. The normal prior is conjugate, while the Cauchy and log-gamma are chosen to illustrate the effects of a thicker-tailed distribution and a skewed distribution, respectively. In the case of the normal prior, the derivative norms for $p = 1, 2, \infty$ can be computed analytically (up to the normal distribution function). In the Cauchy and log-gamma cases one-dimensional numerical integration ($p = 2, \infty$) and maximization ($p = 1$) is used to calculate the norms. The norms are plotted as a function of the data value x, for several sample sizes n, in Figures 1, 2 and 3. Generally, the $p = 1$ norms are an order of magnitude larger than their $p = 2$ counterparts, which in turn are an order of magnitude larger than $p = \infty$ norms.

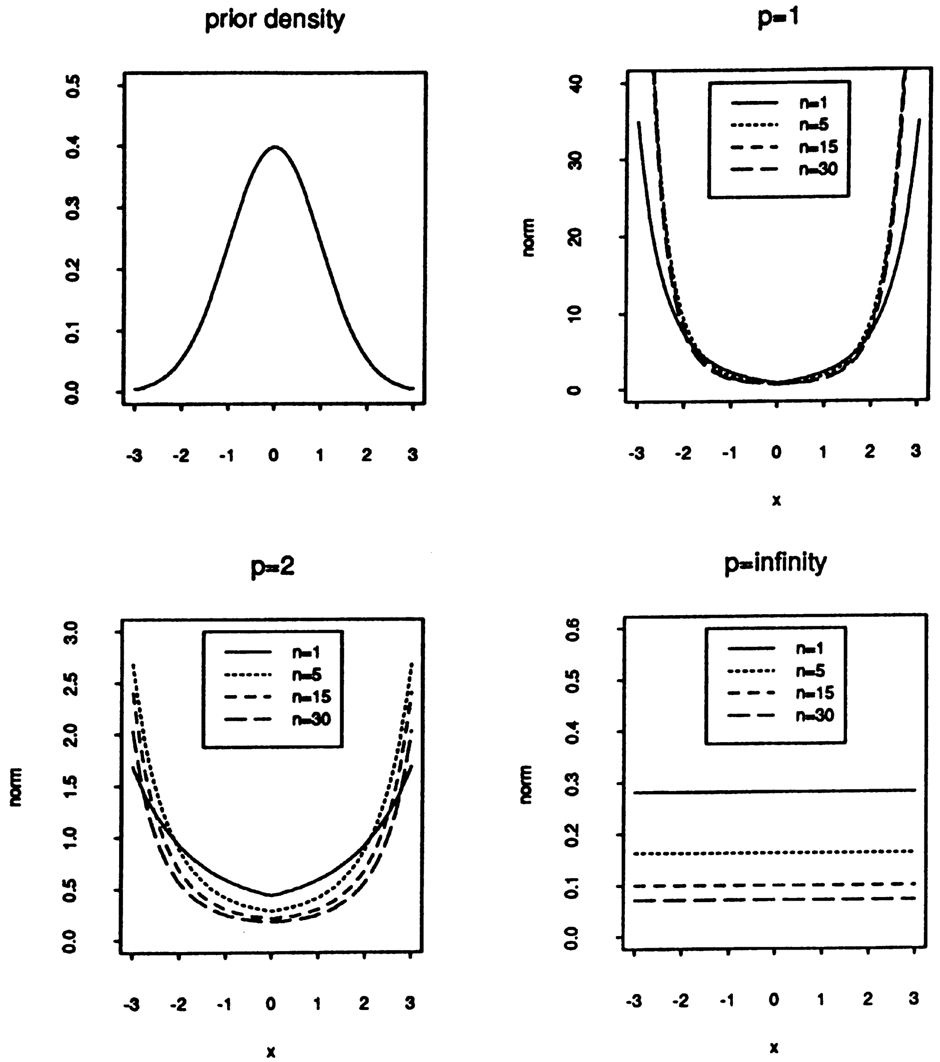

FIG. 1. *Posterior derivative norms: normal model and standard normal prior.*

For the normal prior, in the $p = 1$ case we can distinguish only the $n = 1$ norm. It is surprising that amongst the remaining sample sizes the norms are visually indistinguishable. As expected, the norm increases as the data value lies further from the prior mean. The $p = 2$ norms take on a similar shape as a function of the data, but exhibit a stronger dependence on the sample size. Note here that the sample size ordering is not preserved for extreme data. In particular, when the data value lies far from the prior mean, we see larger sensitivity associated with the larger sample sizes. The sample size depen-

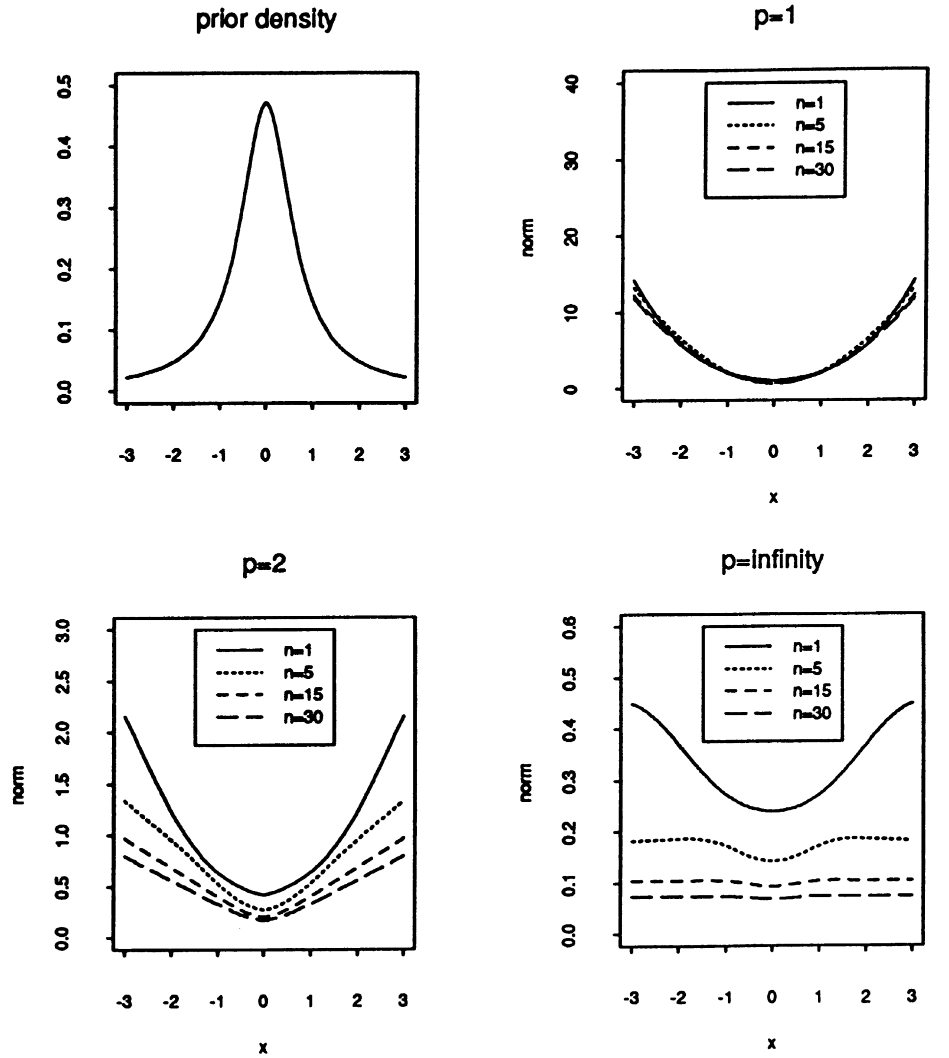

FIG. 2. *Posterior derivative norms: normal model and Cauchy prior. The prior is centred, with scale parameter* 0.674.

dence appears stronger still for the $p = \infty$ norms, but the striking feature here is a complete absence of data dependence.

With the Cauchy prior, in the $p = 1$ case we see even less dependence on the sample size than for the normal prior. Also, the norms are smaller than in the normal prior case, as might be expected with a thick-tailed prior. The $p = 2$ norms again have a shape similar to that of the $p = 1$ norms, but exhibit much more dependence on the sample size. This time, the norm decreases in the sample size even for extreme data values. When $p = \infty$, we

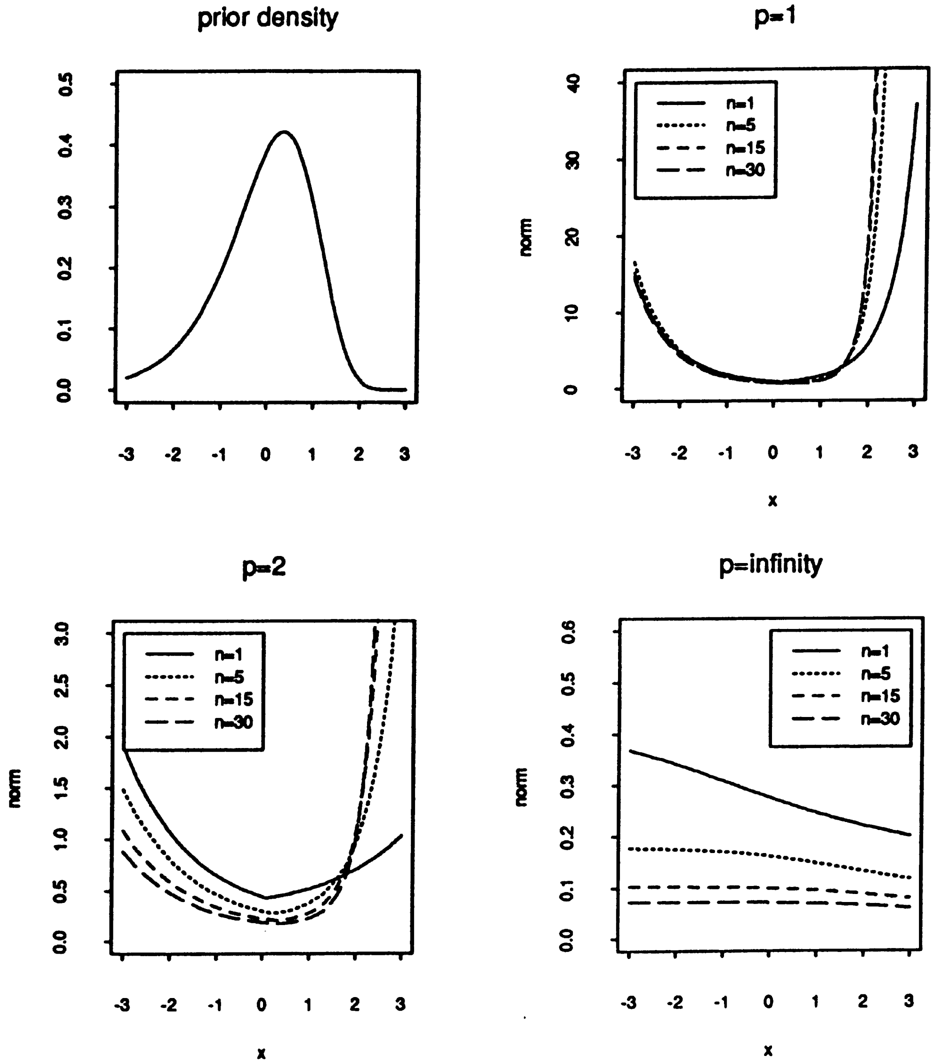

FIG. 3. *Posterior derivative norms*: *normal model and log-gamma prior. The underlying gamma distribution has shape* 1.272 *and scale* 1.119.

see a dependence, albeit weak, on the data value. This dependence grows weaker with larger sample sizes.

With the skewed log-gamma prior, the $p = 1$ and $p = 2$ norms indicate much greater sensitivity for data values in the right (thin) tail of the prior as opposed to the left (thick) tail, in accord with intuition. Again, for sample sizes larger than 1, the dependence on the sample size is minimal when $p = 1$. Paradoxically, when $p = \infty$ the norm is large for data in the left prior tail and small for data in the right prior tail, contrary to the other cases. For larger sample sizes, this phenomenon is tempered, and again the data dependence is extremely weak.

Since a reasonable measure of sensitivity should depend upon both the sample size and the sample mean, these examples suggest that the $p = 2$ norms are the most useful gauge of sensitivity amongst the three norms.

4. Asymptotics. In order to further understand the behaviour exhibited in the examples of the previous section, we consider the asymptotic behaviour of the posterior derivative norms. It is assumed that the model is parameterized so that the first component of the k-dimensional parameter is the function of interest, that is, $g(\theta) = \theta_1$. As in Chen (1985), we consider a fixed data sequence $x_1, x_2, \dots,$ assumed to be an iid realization from $f(\cdot|\theta^*) \in \{f(\cdot|\theta)\colon \theta \in \Theta\}$. Let $I(\theta)$ be the (expected) Fisher information at θ and let $l_n(\cdot)$ denote the log likelihood function based on $x_1, \dots, x_n$. We assume l_n has a strict local maximum at $\hat{\theta}^{(n)}$; hence $l'_n(\hat{\theta}^{(n)}) = 0$ and $-l''_n(\hat{\theta}^{(n)})$ is positive definite. Define

$$A_1(I) = |I_{11} - I_{12} I_{22}^{-1} I_{21}|^{-1},$$

where I_{11}, I_{22} and $I_{12} = I_{21}^T$ are the usual submatrices of the information matrix obtained by partitioning according to the parameter of interest θ_1 and the remaining parameters $(\theta_2, \dots, \theta_k)$. As additional notation, let $B_\delta(\theta)$ be the size δ Euclidean neighbourhood of θ.

We require four conditions for the asymptotic results.

R1. $\hat{\theta}^{(n)} \to \theta^*$ and $-n^{-1} l''(\hat{\theta}^{(n)}) \to I(\theta^*)$.

R2. For every $\varepsilon > 0$, there exists $\delta > 0$ such that the eigenvalues of $[l''_n(\hat{\theta}^{(n)})]^{-1}[l''_n(\theta)]$ lie in the interval $(1 - \varepsilon, 1 + \varepsilon)$ for $\theta \in B_\delta(\hat{\theta}^{(n)})$ and sufficiently large n.

R3. For any $\delta > 0$, there exist positive constants c_1 and c_2 such that for $\theta \in B_\delta^C(\hat{\theta}^{(n)})$,

$$l_n(\theta) - l_n(\hat{\theta}^{(n)}) \le c_1 \Big\{ (\theta - \hat{\theta}^{(n)})^T l''_n(\hat{\theta}^{(n)}) (\theta - \hat{\theta}^{(n)}) \Big\}^{c_2}$$

for sufficiently large n.

R4. π is bounded and positive and continuous at θ^*.

R1, R2 and R4 are common, easily satisfied, regularity conditions. R3, which is essentially Chen's (1985) condition C3.1, is stronger than common tail conditions for asymptotic results (for example, it is stronger than Chen's condition C3). This strength is needed to deal with the supremum involved in the $p = 1$ derivative norm. R3 holds when θ is the natural parameter of an exponential family [Chen (1985)].

Now we can consider the asymptotic behaviour of the norms.

RESULT 13. *Under* R1 *through* R4,

$$n^{(1/2)[1-(k/p)]}\|\dot{T}^{\theta_1}(0)\|_p \to c_{p,k}\left[\frac{|I(\theta^*)|^{1/2}}{\pi(\theta^*)}\right]^{1/p} A_1^{1/2}(I(\theta^*)),$$

where

$$c_{p,k} = \begin{cases} (2\pi)^{-k/2}e^{-1/2}, & p = 1, \\ (2\pi)^{-k/2p}(1/q)^{(1/2)[1+(k/q)]}\left(\int_0^\infty z^q\phi(z)\,dz\right)^{1/q}, & p > 1, \end{cases}$$

with ϕ *being the standard normal density.*

A striking feature here is the dependence of the rate at which the norm shrinks (or grows) upon the dimension of the parameter space. Generally, the norm increases with the dimension of the parameter space. To ensure that the norm vanishes asymptotically, one must choose p larger than k. Thus with larger parameter spaces, one must use smaller classes of perturbations to attain sensible results.

Considering the $k = 1$ case, the asymptotics make clear the problems with the $p = 1$ and $p = \infty$ norms. When $p = 1$ the norm asymptotically does not depend on the sample size; this is consistent with the finite sample behaviour in the examples. The range of a posterior expectation with respect to an ε-contamination class vanishes as n increases [Sivaganesan (1988)], so in an asymptotic sense the local approximation to the global range is poor. Conversely, when $p = \infty$, asymptotically the norm does not depend on the prior, meaning that asymptotically there is no consideration as to whether or not the data are consistent with the prior. Hence in finite samples the norm depends only weakly, if at all, on the data. Only when $p = 2$ does the asymptotic sensitivity depend jointly on the sample size and the prior. This is consistent with the superior performance of the $p = 2$ norm in the examples.

Result 13 easily gives the asymptotic behaviour of the scaled sensitivity as well. Specifically, R1 through R4 imply that the posterior standard deviation of θ_1 satisfies $n^{1/2}\xi_{\theta_1}^{(n)} \to A_1^{1/2}(I(\theta^*))$. Therefore, modulo constants the scaled sensitivity behaves like

$$\left[n^{k/2}\frac{|I(\theta^*)|^{1/2}}{\pi(\theta^*)}\right]^{1/p}.$$

In particular, all components of the parameter are equally sensitive asymptotically after scaling.

Before observing data, a natural goal is to choose the prior so that (at least asymptotically) inference will be insensitive to the prior. As is common when seeking noninformative priors, the parameter space is assumed to be compact, giving the following result.

RESULT 14. *For $p < \infty$ and compact Θ, the maximum (over $\theta \in \Theta$) asymptotic scaled sensitivity is minimized (as a functional of the proper prior) by taking $\pi(\theta)$ to be the probability density proportional to $|I(\theta)|^{1/2}$.*

Thus we have an interpretation of Jeffreys' (1961) prior as the prior to which inference will be minimally sensitive, in an asymptotic minimax sense. This is yet another sense in which Jeffreys' prior is noninformative. For other minimax motivations for the use of Jeffreys' prior, see Clarke and Barron (1994) and Good (1969). Since all parameters are asymptotically equally sensitive after scaling, we obtain a prior which makes no distinction between parameter components. In particular, we do not obtain a prior satisfying Stein's (1985) condition for being noninformative about a particular parameter.

5. Discussion. Perhaps the most important finding here is that by considering classes of priors lying between the extremes of ε-contamination and density ratio classes, we can obtain local measures of sensitivity which behave reasonably in simple problems and asymptotically. Berger (1992) and Wasserman (1992c) have criticized the density ratio class for not being rich enough for robustness studies, while Berger (1990) and others have suggested that (unrestricted) ε-contamination classes are too rich. Often the problem is addressed by using ε-contamination with a restricted class of contaminants. However, this can entail either unrealistic restrictions or difficult calculations. The local sensitivity measure of posterior expectations with respect to $p = 2$ perturbations (linear or nonlinear) appears to have desirable properties, as one might expect from classes of priors that are smaller than ε-contamination classes but larger than density ratio classes. In particular, we see a sensible reliance of the $p = 2$ norm upon the sample size and prior. Of course the fact that both the linear and nonlinear classes yield the same measure of local sensitivity, despite having different indices of richness, underscores the fundamental limitation of local analysis.

While classes of priors based on L^2 norms show promise for use in assessing sensitivity (at least locally), they lack the interpretability of other classes. The size of ε-contamination classes can be gauged by thinking about mixtures, while density ratio classes have a natural interpretation in terms of odds: $d_{\mathrm{DR}}(\pi_1, \pi_2) < k$ implies $\Pi_1(A)/(1 - \Pi_1(A)) < e^k \Pi_2(A)/(1 - \Pi_2(A))$ for all $A \in \mathscr{B}$. We can, however, gain an indication of the size of L^2 based classes by comparing them to the more familiar classes. In particular, Result 6 in the nonlinear case and the remark preceding it in the linear case, give a density ratio class as a lower bound and an ε-contamination class as an upper bound. This provides a way to think about the size of the $p = 2$ classes, which is necessary if one wishes to consider local approximations to global ranges.

The simple examples considered here were introduced to gain some familiarity with the various posterior derivative norms. However, local methods are useful in multiparameter problems, where (i) the model and prior structure are sufficiently complicated that sensitivity measures are really required and (ii) global calculations become intractable. Though the treatment here allows for parameters of arbitrary dimension, it is limited to the case where the whole prior is perturbed. Typically in multiparameter models one would like to know if aspects of the posterior are sensitive to various aspects of the prior, and so one would like to perturb only parts of the prior. Gustafson (1996) shows how local methods can be used for this sort of investigation.

6. Proofs. Proofs of Results 1, 2, 3, 12 and 14 are straightforward and omitted.

PROOF OF RESULT 4. Assume u is nonnegative, $\|u; \mu\|_2 \le \delta$. By the definition of $w_N^*(\cdot; \pi, u, 2)$ we have

$$\frac{u(\cdot)}{2} = \left(w_N^*(\cdot; \pi, u, 2)\right)^{1/2} \left\| \pi^{1/2} + \frac{u}{2}; \mu \right\|_2 - \pi^{1/2}(\cdot).$$

Squaring both sides and integrating gives

$$\begin{aligned}\frac{\|u; \mu\|_2^2}{4} &= \left\| \pi^{1/2} + \left(\frac{u}{2}\right); \mu \right\|_2^2 + 1 \\ &\quad + \left\| \pi^{1/2} + \left(\frac{u}{2}\right); \mu \right\|_2 \left(2 - d_H^2(\pi(\cdot), w_N^*(\cdot; \pi, u, 2))\right).\end{aligned}$$

Further rearrangement yields

$$d_H^2(\pi(\cdot), w_N^*(\cdot; \pi, u, 2)) = \frac{\|u; \mu\|_2^2/4}{\|\pi^{1/2} + (u/2); \mu\|_2} - \frac{\left(\|\pi^{1/2} + (u/2); \mu\|_2 - 1\right)^2}{\|\pi^{1/2} + (u/2); \mu\|_2}$$

and, in particular,

$$d_H^2(\pi(\cdot), w_N^*(\cdot; \pi, u, 2)) \le \frac{\|u; \mu\|_2^2}{4},$$

the desired result. To see that the containment is proper, note that Hellinger neighbourhoods, unlike Γ_N^2, can contain densities with smaller support than the base density. □

PROOF OF RESULT 5. We consider the nonlinear case first. Note that the density ratio distance is invariant to scalar multiplication of densities, so we need not work with normalized densities. First say $u \ge 0$, $\|u\|_\infty \le \delta$. We want to show $d_{\mathrm{DR}}(\pi, \pi e^u) \le \delta$. However,

$$d_{\mathrm{DR}}(\pi, \pi e^u) = \operatorname*{ess\ sup}_{\theta, \phi \in S(\pi_0)} (u(\theta) - u(\phi)) \le \operatorname*{ess\,sup}_{\theta} u(\theta) \le \delta,$$

since u is nonnegative. Next assume that $d_{\mathrm{DR}}(\pi, \pi') \le \delta$. π and π' must then have common support, which we will denote S. Define $u^* = \log(\pi'/\pi)$ on S, $u^* = 0$ on S^C. Then

$$\delta \ge d_{\mathrm{DR}}(\pi, \pi') = \operatorname*{ess\,sup}_{\theta \in S} u^*(\theta) - \operatorname*{ess\ inf}_{\phi \in S} u^*(\phi),$$

which implies u^* is essentially bounded. In particular, taking $u = u^* - (\inf_\phi u^*(\phi))$ (except on a set of μ-measure 0, to ensure nonnegativity of u) implies that $\pi'(\cdot)$ and $w_N^*(\cdot; \pi, \infty)$ define the same probability measure. Since $u \geq 0$ and $\|u\|_\infty \leq \delta$, $\pi' \in \Gamma_N^\infty(\Pi; \delta)$, completing the proof for the nonlinear case. For the linear case, we relate the linear class to a nonlinear class. In particular, $\pi + u = \pi e^{u'}$, where $u' = \log(1 + (u/\pi))$. This immediately implies the result. □

PROOF OF RESULT 6. Fix u such that $\|u; \mu\|_{p_2} \leq \delta$. First we consider the case $p_2 < \infty$. By simple algebra we can verify that $w_N(\cdot; \pi, u, p_2) = w_N(\cdot; \pi, u', p_1)$, where $u' = p_1[(\pi^{1/p_2} + (u/p_2))^{p_2/p_1} - \pi^{1/p_1}]$. Since u' is convex in u, by the mean value theorem we have $0 \leq u' \leq (\pi^{1/p_2} + (u/p_2))^{(p_2/p_1)-1}u$; hence $\|u'; \mu\|_{p_1} \leq [\int_\Theta (\pi^{1/p_2} + (u/p_2))^{p_2 - p_1} u^{p_1}\, d\mu]^{1/p_1}$. Applying Hölder's inequality to the integral gives $\|u'; \mu\|_{p_1} \leq \|\pi^{1/p_2} + (u/p_2)\|_{p_2}^{(p_2/p_1)-1} \|u\|_{p_2}$, so by subadditivity we have $\|u'; \mu\|_{p_1} \leq (1 + \delta/p_2)^{(p_2/p_1)-1}\delta$, giving the required result. The proof when $p_2 = \infty$ proceeds similarly. Perturbation by u with respect to $p_2 = \infty$ is equivalent to perturbation by u' with respect to p_1, where $u' = p_1 \pi^{1/p_1}[e^{(u/p_1)} - 1]$, which is bounded by $\pi^{1/p_1} e^{u/p_1} u$. Hence $\|u'\|_{p_1} \leq e^{\delta/p_1}\delta$, yielding the required result. □

PROOF OF RESULT 7. First consider the linear case, with $p < \infty$. Note that for $A \in \mathscr{B}$,

$$\text{(8)} \qquad \sup_{\Pi' \in \Gamma_L^p(\Pi, \delta)} \frac{\Pi'(A)}{\Pi(A)} = \sup_{\{u:\, \|u/\pi : \Pi\|_p \leq \delta\}} \left(\Pi(A)\left[1 + \frac{\Pi(A^c) + \int_{A^c} u\, d\mu}{\Pi(A) + \int_A u\, d\mu}\right]\right)^{-1},$$

which is increasing in $\int_A u\, d\mu$ and decreasing in $\int_{A^c} u\, d\mu$. Therefore the aim is to maximize $\int_A u\, d\mu$ amongst u which vanish on A^c. If u has size less than δ, then by Hölder's inequality

$$\int_A u\, d\mu \leq \left(\int_A (u/\pi)^p\, d\Pi\right)^{1/p} \Pi^{1/q}(A)$$
$$\leq \delta \Pi^{1/q}(A),$$

where equality can be obtained by taking u proportional to π. Applying this to (8) and rearranging gives

$$R_\Gamma(x) = \sup_{\{A \in \mathscr{B}:\, \Pi(A) \geq x^{-1}\}} \left[\frac{1 + \delta \Pi^{-1/p}(A)}{1 + \delta \Pi^{1/q}(A)}\right].$$

Noting that the quantity in square brackets is decreasing in $\Pi(A)$, and using absolute continuity of Π with respect to Lebesgue measure (which guarantees the existence of small Π-probability sets), it follows that

$$r_\Gamma(x) = \frac{1 + \delta x^{1/p}}{1 + \delta x^{-1/q}}$$
$$\sim x^{1/p} \qquad (x \to \infty),$$

from which the result follows immediately. In light of Result 5, the result in the linear $p = \infty$ case follows from that for the nonlinear $p = \infty$ case, proved below.

For nonlinear perturbations with $p < \infty$,

$$\sup_{\Pi' \in \Gamma_L^p(\Pi,\delta)} \frac{\Pi'(A)}{\Pi(A)} = \sup_{\{u:\, \|u;\mu\|_p \le \delta\}} \left(\Pi(A) \left[1 + \frac{\int_{A^c} (\pi^{1/p} + (u/p))^p \, d\mu}{\int_A (\pi^{1/p} + (u/p))^p \, d\mu} \right] \right)^{-1}. \tag{9}$$

As previously stated, it is necessary to maximize $\int_A (\pi^{1/p} + (u/p))^p$ subject to $u = 0$ on A^c, and $\|u; \mu\|_p \le \delta$. By Minkowski's inequality, the maximum is $(\Pi^{1/p}(A) + (\delta/p))^p$, which is obtained by taking u proportional to $\pi^{1/p}$ on A. This implies that (9) simplifies to

$$\left(\Pi(A) \left[1 + \frac{1 - \Pi(A)}{(\Pi^{1/p}(A) + (\delta/p))^p} \right] \right)^{-1}. \tag{10}$$

A straightforward calculus argument shows that (10) is decreasing in $\Pi(A)$, so the richness function is obtained by replacing $\Pi(A)$ with $1/x$ in (10). Upon rearrangement, this gives

$$R_\Gamma(x) = \frac{(1 + (\delta/p) x^{1/p})^p}{1 - (1/x) + (1/x)(1 + (\delta/p) x^{1/p})^p} \sim x \qquad (x \to \infty),$$

implying the desired result. When $p = \infty$,

$$\begin{aligned} R_\Gamma(x) &= \sup_{\{A \in \mathscr{B}:\, \Pi(A) \ge x^{-1}\}} \sup_{\Pi' \in \Gamma} \left(\Pi(A) \left[1 + \frac{\int_{A^c} \pi e^u \, d\mu}{\int_A \pi e^u \, d\mu} \right] \right)^{-1} \\ &= \sup_{\{A \in \mathscr{B}:\, \Pi(A) \ge x^{-1}\}} \left(\Pi(A) \left[1 + \frac{\Pi(A^c)}{e^\delta \Pi(A)} \right] \right)^{-1} \\ &= \left(e^{-\delta} + [1 - e^{-\delta}] \frac{1}{x} \right)^{-1}, \end{aligned}$$

which tends to e^δ as $x \to \infty$, giving the result. □

PROOF OF RESULT 8. For bounded functions h on Θ define $I_h u = \int_\Theta h(\theta) w_L(\theta; \pi, u) \mu(d\theta)$; hence $T_L^g u = I_{gL} u / I_L u$. By the linearity of I_h in u

we easily have $\dot{I}_h(0)u = \int_\Theta hu\,d\mu$ with respect to any norm. By the quotient rule,

$$\dot{T}_L^g(0)u = \frac{\dot{I}_{gL}(0)uI_L0 - I_{gL}0\dot{I}_L(0)u}{(I_L(0))^2} = \mathrm{Cov}_\pi^x\left(g, \frac{u}{\pi}\right),$$

completing the proof. □

Proof of Result 9. Define $I_h u = \int_\Theta h(\theta)w_N(\theta;\pi,u,p)\mu(d\theta)$. Hence $T_g u = I_{gL}u/I_L u$, so we can differentiate T by differentiating I. We need to show that if h is bounded, then $\dot{I}_h(0)u = \int_\Theta h\pi^{1/q}u\,d\mu$. In light of the boundedness assumption, it will suffice to show that $\int_\Theta r_u(\theta)\mu(d\theta) = o(\|u\|_p)$ uniformly, where

$$r_u = \begin{cases} (\pi^{1/p} + u/p)^p - \pi - \pi^{1/q}u, & p < \infty,\\ \pi[e^u - 1 - u], & p = \infty.\end{cases}$$

When $p = 1$, r_u is identically zero, so the result holds. When $p \in (1,2)$, $r_u(\theta)$ is convex in u for each θ, so by the mean value theorem, $r_u \le [(\pi^{1/p} + u/p)^{p-1} - \pi^{1/q}]u$, giving $r_u \le (u/p)^{p-1}u$, and, consequently, $\int_\Theta r_u\,d\mu \le (1/p)\|u;\,\mu\|_p^p = o(\|u;\,\mu\|_p)$. The next case is $p \in [2,\infty)$. Here $r_u(\theta)$ has three nonnegative derivatives with respect to u; hence $r_u \le [(p-1)/(2p)](\pi^{1/p} + u/p)^{p-2}u^2$. When $p = 2$, this directly implies that $\int_\Theta r_u\,d\mu = o(\|u;\,\mu\|_2)$. When $p > 2$, by Hölder's inequality we have

$$\int_\Theta r_u\,d\mu \le \left\|\left(\pi^{1/p} + \frac{u}{p}\right)^{p-2};\mu\right\|_{p/(p-2)} \|u^2;\mu\|_{p/2} = \left\|\pi^{1/p} + \frac{u}{p};\mu\right\|_p^{p-2} \|u;\mu\|_p^2$$
$$\le \left(1 + \frac{1}{p}\|u;\mu\|_p\right)^{p-2} \|u;\mu\|_p^2 = o(\|u;\mu\|_p).$$

Finally, when $p = \infty$,

$$\int_\Theta r_u\,d\mu \le \int_\Theta \pi e^u(u^2/2)\,d\mu \le (1/2)e^{\|u;\mu\|_\infty}\|u;\mu\|_\infty^2 = o(\|u;\mu\|_\infty),$$

as desired. Now we can apply the quotient rule for functional derivatives. In particular,

$$\dot{T}_N^g(0)u = \frac{\dot{I}_{gL}(0)uI_L0 - I_{gL}0\dot{I}_L(0)u}{(I_L0)^2} = \mathrm{Cov}_\pi^x\left(g, \frac{u}{\pi^{1/p}}\right),$$

the required result. □

PROOF OF RESULT 10. The expression $\dot{T}_L^g(0)u$ can be written as $\int_\Theta a(u/\pi)\,d\Pi$, or

$$\dot{T}_L^g(0)u = \int_\Theta (a_L^+)\frac{u}{\pi}\,d\Pi - \int_\Theta (a_L^-)\frac{u}{\pi}\,d\Pi. \tag{11}$$

The boundedness of L and gL ensures that $\|a_L^+;\Pi\|_q < \infty$. Thus by Hölder's inequality, the first term of (11) is bounded by $\|a_L^+;\Pi\|_q\|u/\pi;\Pi\|_p$. In fact, we can either achieve the bound or come arbitrarily close to achieving the bound, since a_L^+ is nonnegative. When $p = 1$ we can get close to the bound by taking u close to the generalized function $\delta_{\arg\max(a_L^+)}$, when $p \in (1,\infty)$ we can achieve the bound by taking $u \propto (a_L^+)^{q-1}\pi$ and when $p = \infty$ we can achieve the bound by taking $u \propto I_{\{a_L > 0\}}\pi$. In all cases, the second term of (11) is zero and, hence, the whole expression is maximized. To minimize (11), we replace a_L^+ with a_L^-. □

PROOF OF RESULT 11. The derivative at 0 is $\int_\Theta a_N^+ u\,d\mu - \int_\Theta a_N^- u\,d\mu$. Again, the boundedness of L and gL guarantees that $\|a_N^+;\mu\|_q$ is finite. Thus the argument in the previous proof applies here as well. In this case, the maximizing directions for the positive part are (proportional to) $\delta_{\arg\max(a_N^+)}$ when $p = 1$, $(a_N^+)^{q-1}$ when $p \in (1,\infty)$ and $I_{\{a_N > 0\}}$ when $p = \infty$. □

PROOF OF RESULT 13. Intuitively, the result is obtained by replacing the posterior distribution with a normal distribution (mean $\hat{\theta}^{(n)}$, variance $[-l''(\hat{\theta}^{(n)})]^{-1}$) and then applying R1. This can be made precise in the $p > 1$ case by standard arguments, splitting the region of integration into a neighbourhood of the MLE and its complement. The nonstandard case is $p = 1$, where we must deal with a supremum over the parameter space. As preliminaries, note that R1 through R4 imply the standard asymptotic results

$$\left|-l_n''(\hat{\theta}^{(n)})\right|^{-1/2}\pi_n^x(\hat{\theta}^{(n)}) \to (2\pi)^{-k/2}, \tag{12}$$

$$\left(\rho^{(n)} - \hat{\theta}^{(n)}\right)^T l_n''(\hat{\theta}^{(n)})\left(\rho^{(n)} - \hat{\theta}^{(n)}\right) \to 0, \tag{13}$$

where $\rho^{(n)}$ is the vector of posterior means of θ. Thus we have

$$\lim_{n\to\infty} n^{(1/2)[1-k]}\left\|\dot{T}^{\theta_1}(0)\right\|_1 = \left[\lim_{n\to\infty}\frac{n^{-k/2}\pi_n^x(\hat{\theta}^{(n)})}{\pi(\hat{\theta}^{(n)})}\right]\left[\lim_{n\to\infty}\sup_\theta h_n(\theta)\right], \tag{14}$$

where

$$h_n(\theta) = n^{1/2}|\theta_1 - \hat{\theta}_1^{(n)}|\exp\left(l_n(\theta) - l_n(\hat{\theta}^{(n)})\right).$$

The replacement of ρ_{θ_1} by $\hat{\theta}_1$ is justified by (13). Applying (12) to the first term of (14), it only remains to show that $\sup_\theta h_n(\theta) \to e^{-1/2}A_1^{1/2}(I(\theta^*))$. R3 gives that for any fixed δ the supremum will occur on $B_\delta(\hat{\theta}^{(n)})$ for sufficiently large n. Fixing ε and choosing δ as in R2 gives, for sufficiently large n,

$$\begin{aligned}\sup_{\theta} h_n(\theta) \le \sup_{\theta \in B_\delta(\hat{\theta}^{(n)})} \Bigg[n^{1/2} |\theta_1 - \hat{\theta}_1^{(n)}| \exp\bigg(-\left(\frac{1}{2}\right)[1-\varepsilon](\theta - \hat{\theta}^{(n)})^T \\ \times \left[-l_n''(\hat{\theta}^{(n)}) \right] (\theta - \hat{\theta}^{(n)}) \bigg) \Bigg] \\ = \sup_{z \in B_{n^{1/2}\delta}(0)} |z_1| \exp\left(-\left(\frac{1}{2}\right)[1-\varepsilon] z^T \left[-n^{-1} l_n''(\hat{\theta}^{(n)}) \right] z \right) \\ = e^{-1/2} A_1^{1/2} \left((1-\varepsilon) \left[-n^{-1} l_n''(\hat{\theta}^{(n)}) \right] \right) \\ \to \left(\frac{1}{1-\varepsilon} \right)^{1/2} e^{-1/2} A_1^{1/2} (I(\theta^*)),\end{aligned}$$

with a similar lower bound holding as well. Taking $\varepsilon \downarrow 0$ completes the proof. □

Acknowledgment. This work is based on part of the author's Ph.D. thesis at the Department of Statistics, Carnegie Mellon University. The author thanks his supervisor, Larry Wasserman, for encouragement and helpful suggestions.

REFERENCES

Basu, S., Jammalamadaka, S. R. and Liu, W. (1993). Local posterior robustness with parametric priors: maximum and average sensitivity. In *Maximum Entropy and Bayesian Methods* (G. Heidbreder, ed.) Kluwer, Dordecht.

Berger, J. O. (1986). Comment on "On the consistency of Bayes estimates" by P. Diaconis and D. Freedman *Ann. Statist.* **14** 30–37.

Berger, J. O. (1990). Robust Bayesian analysis: sensitivity to the prior. *J. Statist. Plann. Inference* **25** 303–328.

Berger, J. O. (1992). Comment on "Recent methodological advances in robust Bayesian inference" by L. Wasserman. In *Bayesian Statistics 4* (J. M. Bernando, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds.) 495–496. Oxford Univ. Press.

Berger, J. O. (1994). An overview of robust Bayesian analysis (with discussion). *Test* **3** 5–124.

Berger, J. O. and Berliner, L. M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14** 461–486.

Chen, C. F. (1985). On asymptotic normality of limiting density functions with Bayesian implications. *J. Roy. Statist. Soc. Ser. B* **47** 540–546.

Cifarelli, D. M. and Regazzini, E. (1987). On a general definition of concentration function. *Sankyā Ser. B* **49** 307–319.

Clarke, B. S. and Barron, A. R. (1994). Jeffreys' prior is asymptotically least favourable under entropy risk. *J. Statist. Plann. Inference* **41** 37–60.

Cuevas, A. and Sanz, P. (1988). On differentiability properties of Bayes operators. In *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 569–577. Oxford Univ. Press.

Delampady, M. and Dey, D. (1994). Bayesian robustness for multiparameter problems. *J. Statist. Plann. Inference* **40** 375–382.

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

DEROBERTIS, L. and HARTIGAN, J. (1981). Bayesian inference using intervals of measures. *Ann. Statist.* **9** 235–244.

DEY, D. and BIRMIWAL, L. R. (1994). Robust Bayesian analysis using entropy and divergence measures. *Statist. Probab. Lett.* **20** 287–294.

DIACONIS, P. and FREEDMAN, D. (1986). On the consistency of Bayes estimates (with discussion). *Ann. Statist.* **14** 1–67.

GELFAND, A. and DEY, D. (1991). On Bayesian robustness of contaminated classes of priors. *Statist. Decisions* **9** 63–80.

GOOD, I. J. (1969). What is the use of a distribution? In *Multivariate Analysis* (P. R. Krishnaiah, ed.) **2** 183–203. Academic Press, New York.

GUSTAFSON, P. (1996). Local sensitivity of inferences to prior marginals. *J. Amer. Statist. Assoc.* **91** To appear.

GUSTAFSON, P. and WASSERMAN, L. (1995). Local sensitivity diagnostics for Bayesian inference. *Ann. Statist.* **23** 2153–2167.

HARTIGAN, J. A. (1983). *Bayes Theory*. Springer, New York.

JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Clarendon Press, Oxford.

KASS, R. E., TIERNEY, L. and KADANE, J. B. (1989). Approximate methods for assessing influence and sensitivity in Bayesian analysis. *Biometrika* **76** 663–674.

LAVINE, M. (1991a). Sensitivity in Bayesian statistics: the prior and the likelihood. *J. Amer. Statist. Assoc.* **86** 396–399.

LAVINE, M. (1991b). An approach to robust Bayesian analysis for multidimensional parameter spaces. *J. Amer. Statist. Assoc.* **86** 400–403.

LINDLEY, D. V. (1992). Comment on "Recent methodological advances ion robust Bayesian inference" by L. Wasserman. In *Bayesian Statistics 4* (J. M. Bernando, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 496. Oxford Univ. Press.

RUGGERI, F. and WASSERMAN, L. (1993). Infinitesimal sensitivity of posterior distributions. *Canad. J. Statist.* **21** 195–203.

RUGGERI, F. and WASSERMAN, L. (1995). Density based classes of priors: infinitesimal properties and approximations. *J. Statist. Plann. Inference* **46** 311–324.

SIVAGANESAN, S. (1988). Range of posterior measures for priors with arbitrary contaminations. *Comm. Statist. Theory Methods* **17** 1591–1612.

SIVAGANESAN, S. (1993). Robust Bayesian diagnostics. *J. Statist. Plann. Inference* **35** 171–188.

SRINIVASAN, C. and TRUSZCZYNSKA, H. (1990). On the ranges of posterior quantities. Technical Report 294, Dept. Statistics, Univ. Kentucky, Lexington.

STEIN, C. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics* 485–514. PWN Publishers, Warsaw.

WASSERMAN, L. (1992a). Invariance properties of density ratio priors. *Ann. Statist.* **20** 2177–2182.

WASSERMAN, L. (1992b). Recent methodological advances in robust Bayesian inference (with discussion). In *Bayesian Statistics 4* (J. M. Bernando, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 483–502. Oxford Univ. Press.

WASSERMAN, L. (1992c). The conflict between improper priors and robustness. Technical Report 559. Dept. Statistics, Carnegie Mellon Univ.

DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BRITISH COLUMBIA
CANADA V6T 1Z2