Journal of Statistical Planning and Inference 40 (1994) 375–382

Bayesian robustness for multiparameter problems

Mohan Delampady

Economic Analysis Unit, Indian Statistical Institute, Mysore Road, R.V.C.E. Post, Bangalore 560 009, India

Dipak K. Dey

Department of Statistics, Box U-120, University of Connecticut, Storrs, CT 06209, USA

Received 1 June 1992; revised manuscript received 10 April 1993

Abstract

Bayesian robustness is studied for ε-contamination classes of prior distributions. Nonparametric classes of contaminations such as the class of all unimodal spherically symmetric densities are considered here. Posterior ϕ-divergence and its curvature are used to measure the sensitivity of priors on the resulting posterior densities. Examples are provided to illustrate our results.

AMS Subject Classification: Primary 62A15; secondary 62F15.

Key words: ϕ-divergence; ε-contamination; curvature.

1. Introduction

Sensitivity of inferences when uncertainties exist regarding the prior being used is of substantial interest. In this paper we study the robustness of posterior densities when the prior varies in a reasonably large class of distributions that are in a neighborhood of a subjectively elicited prior π_0. Specifically, we consider the ε-contamination class

$$\Gamma = \{\pi\colon\ \pi(\theta) = (1-\varepsilon)\pi_0(\theta) + \varepsilon q(\theta),\ q \in Q\}, \tag{1.1}$$

where Q is a large class of plausible contaminations.

A large literature exists on robust Bayesian studies for ε-contamination classes of priors. Berger (1985), Berger and Berliner (1986) and Sivaganesan and Berger (1989) are some of the papers related to the discussion that follows. The major focus of these articles has been the study of the range of some of the posterior quantities.

Correspondence to: Dr. D.K. Dey, Department of Statistics, Box U-120, University of Connecticut, Storrs, CT 06209, USA.

We, however, take a different approach and study the sensitivity of the posterior density using the distance between the posterior arising from π_0 and that arising from any π in Γ. This approach with Kullback–Leibler divergence has recently been used by Gelfand and Dey (1990). Dey and Birmiwal (1990) (referred to as D&B in the following discussion) extend those results to other divergence measures and apply them to parametric classes of contaminations. Our main focus is the study of robustness when the class Q of contaminations is nonparametric. We consider the class of all spherically symmetric unimodal densities.

Let us first develop some definitions and notations. Suppose X is the random variable with density (or mass function) $f(x\,|\,\theta)$, where θ is the parameter vector of interest. The marginal distribution of X with respect to any prior π is given by $m(x\,|\,\pi)=\int f(x\,|\,\theta)\pi(\theta)\,d\theta$, and hence for a prior density $\pi\in\Gamma$ given in (1.1), $m(x\,|\,\pi)=(1-\varepsilon)m(x\,|\,\pi_0)+\varepsilon m(x\,|\,q)$. Also, the posterior density of θ given x with respect to π is $\pi(\theta\,|\,x)=\lambda(x)\pi_0(\theta\,|\,x)+(1-\lambda(x))q(\theta\,|\,x)$, where $\pi_0(\theta\,|\,x)$ and $q(\theta\,|\,x)$ are, respectively, the posterior densities with respect to π_0 and q, and $\lambda(x)\in[0,1]$ is given by $\lambda(x)=(1-\varepsilon)m(x\,|\,\pi_0)/m(x\,|\,\pi)$.

Following D&B, define ϕ-divergence between posterior densities $\pi_0(\theta\,|\,x)$ and $\pi(\theta\,|\,x)$ to be

$$D=D(\pi(\theta\,|\,x),\pi_0(\theta\,|\,x))=\int \pi_0(\theta\,|\,x)\phi(\pi(\theta\,|\,x)/\pi_0(\theta\,|\,x))\,d\theta, \tag{1.2}$$

where ϕ is any continuously differentiable convex function. Several well-known divergence measures, e.g. Kullback–Leibler, Hellinger distance, power divergence, etc., can be obtained by the appropriate choice of the φ-function.

One possible approach to study the robustness of posterior densities is by obtaining the range of ϕ-divergence as q varies in Q for different choices of ϕ. It will be seen later that this is not always easy. However, a measure of this range is given by the range of the local curvature of the ϕ-divergence. As in D&B we define the local curvature of the ϕ-divergence to be

$$C=C(q)=\frac{d^2}{d\varepsilon^2}D\bigg|_{\varepsilon=0}. \tag{1.3}$$

In Section 2, we study the variation of this local curvature for an appealing nonparametric class of contaminations. The results will be illustrated using examples. Proofs of the main results are given in the appendix.

2. Variation of curvature for nonparametric classes of contaminations

It can be shown (Theorem 4.1, D&B) that

$$C(q)=\phi''(1)\operatorname*{Var}_{\pi_0(\theta\,|\,x)}\left[\frac{q(\theta)}{\pi_0(\theta)}\right], \tag{2.1}$$

where ϕ'' denotes the second derivative and $\mathrm{Var}_{\pi_0(\theta|x)}$ denotes the variance with respect to $\pi_0(\theta|x)$.

Note that

$$\operatorname*{Var}_{\pi_0(\theta|x)}\left[\frac{q(\theta)}{\pi_0(\theta)}\right]=\left\{\int\frac{f(x|\theta)}{\pi_0(\theta)}q^2(\theta)\,\mathrm{d}\theta-\frac{1}{m(x|\pi_0)}\left[\int f(x|\theta)q(\theta)\,\mathrm{d}\theta\right]^2\right\}\Big/m(x|\pi_0)$$

$$=A(q)/m(x|\pi_0)\quad\text{(say)}.\tag{2.2}$$

Variations of local curvature for various classes of contaminations Q are of interest. Consequently, it is necessary to compute $\sup_{q\in Q} C(q)$. It can be seen that if all densities are allowed as contaminations, the supremum above is unbounded. Also, robustness with respect to a huge class such as this is neither expected nor interesting. Reasonable constraints need to be imposed on this class to obtain interesting classes of contaminations. Unimodality and spherical symmetry are some reasonable shape constraints on the contaminations. Even with these constraints, the resulting classes of contaminations include infinitely spiky densities which are clearly unreasonable. With this in mind, we impose a further constraint that $\max_\theta q(\theta)\leqslant h$, for any allowable contamination where h is a prespecified height for q. A reasonable choice for h is $\pi_0(0)$ (see Sivaganesan (1991) for other choices and their justifications). Define Q_{US}, the class of all unimodal spherically symmetric densities q, which satisfies the above constraint on the maximum height of q. Before we can derive results on the supremum of curvature for these classes, we need to make assumptions on f and π_0.

Suppose

$$B(r)=\int_{S(r)}\frac{f(x|\theta)}{\pi_0(\theta)}\mathrm{d}\theta-\frac{1}{m(x|\pi_0)}\left[\int_{S(r)}f(x|\theta)\,\mathrm{d}\theta\right]^2,$$

where $S(r)$ denotes the sphere of radius r around 0. Let $V(r)$ denote the volume of a sphere of radius r and $U(S(r))$ denote the uniform density on $S(r)$. Clearly, $B(r)/V^2(r)=A(U(S(r)))$.

Assumption 1. (i) $B(r)$ is increasing in r in the range where $h\leqslant 1/V(r)$.

(ii) If $\max_{V(r)\geqslant 1/h}A(U(S(r)))=A(U(S(t)))$, then the function l defined in the range $0\leqslant V(r)\leqslant 1/h$, by

$$l(r)=\left[\frac{1/h-V(r)}{V(t)-V(r)}\right]^2[B(t)-B(r)]+B(r),$$

has its global maximum at $r=0$.

Remark. These assumptions are stronger than what we require. However, these are much easier to check. Also, note that $B(r)$ is always increasing near 0. It can be very easily seen in one dimension that

$$\frac{\mathrm{d}}{\mathrm{d}r}B(r)\bigg|_{r=0}=2\frac{f(x|\theta)}{\pi_0(\theta)}>0.$$

Further, if π_0 is symmetric,

$$\frac{\mathrm{d}}{\mathrm{d}r}B(r)=\frac{f(x\,|\,r)+f(x\,|\,-r)}{\pi_0(r)m(x\,|\,\pi_0)}\left[\int_{-\infty}^{\infty}f(x\,|\,\theta)\pi_0(\theta)\,\mathrm{d}\theta-2\pi_0(r)\int_{-r}^{r}f(x\,|\,\theta)\,\mathrm{d}\theta\right].$$

For π_0 which are reasonably sharp,

$$2\pi_0(r)\int_{-r}^{r}f(x\,|\,\theta)\,\mathrm{d}\theta<\int_{-r}^{r}\pi_0(\theta)f(x\,|\,\theta)\,\mathrm{d}\theta<\int_{-\infty}^{\infty}\pi_0(\theta)f(x\,|\,\theta)\,\mathrm{d}\theta,$$

so that $(\mathrm{d}B(r)/\mathrm{d}r>0$ for all r.

Assumption (ii) is harder to check. We have noticed in many examples that $l(r)$ is quite often a monotone decreasing function of r. However, we have found cases where $l(r)$ attains a minimum near $V(r)=1/h$, and increases gradually in the rest of the range.

Theorem 2.1. *Under Assumption* 1,

$$\sup_{q\in Q_{US}}\int\frac{f(x\,|\,\theta)}{\pi_0(\theta)}q^2(\theta)\,\mathrm{d}\theta-\frac{1}{m(x\,|\,\pi_0)}\left[\int f(x\,|\,\theta)q(\theta)\,\mathrm{d}\theta\right]^2$$
$$=\sup_{V(r)\geqslant 1/h}A(U(S(r))). \tag{2.3}$$

Proof. The details are given in the appendix; the main features of the proof are the following. Note, first of all, that any $q\in Q_{US}$ is a mixture of uniform densities on spheres centered at the origin, i.e.

$$q(\theta)=\int_0^{\infty}\frac{1}{V(r)}I(\theta\in S(r))\,\mathrm{d}\mu(r),$$

where μ is any probability measure on $[0,\infty)$ satisfying

$$q(0)=\int_0^{\infty}\frac{1}{V(r)}\,\mathrm{d}\mu(r)\leqslant h. \tag{2.4}$$

The proof is done in two steps. In the first step, consider r_i, $i=1,2$, such that $U(S(r_i))$ satisfy (2.3). Then both r_i satisfy $V(r_i)\geqslant 1/h$. We show that if $A(U(S(r_1)))>A(U(S(r_2)))$, then $A(U(S(r_1)))\geqslant A(\alpha U(S(r_1))+(1-\alpha)U(S(r_2)))$, for $0\leqslant\alpha\leqslant 1$, where $\alpha(U(S(r_1))+(1-\alpha)U(S(r_2))$ is the mixture of densities $(U(S(r_1))$ and $(U(S(r_2))$. In step 2, let t be such that $\max_{V(r)\geqslant 1/h}A(U(S(r)))=A(U(S(t)))$. Then consider $\alpha U(S(r))+(1-\alpha)U(S(t))$, where $V(r)<1/h$. We show that $A(U(S(t)))\geqslant A(\alpha U(S(r))+(1-\alpha)U(S(t)))$, for $0\leqslant\alpha\leqslant 1$, under Assumption 1 that we made above. □

The result above is now illustrated using two interesting examples.

Example 1. Suppose $X\,|\,\theta\sim\mathrm{N}(\theta,1)$, and under π_0, $\theta\sim\mathrm{N}(0,\tau^2)$, $\tau^2>1$. Then $m(\cdot\,|\,\pi_0)$ is the density of $\mathrm{N}(0,\tau^2+1)$, and $f(x\,|\,\theta)/\pi_0(\theta)$ is $[\sqrt{2\pi\tau^2}/\sqrt{\tau^2-1}]\exp(x^2/2(\tau^2-1))$

times the $N(\tau^2 x/(\tau^2-1), \tau^2/(\tau^2-1))$ density for θ. Assumption 1 can be verified for x values which are sufficiently large enough in magnitude for the given value of τ^2. For example, we have verified that when $\tau^2=1.1$, Assumption 1 is valid for $|x|\geqslant 2$, and when $\tau^2=2$, it is valid for $|x|\geqslant 3$, and so on. Let $\mu(x,\tau^2)=\tau^2 x/(\tau^2-1)$, $\sigma^2(x,\tau^2)=\tau^2/(\tau^2-1)$ and ψ and Φ denote the standard normal density and c.d.f., respectively, then we obtain

$$\sup_{q\in Q_{US}} C(q)=\frac{\phi''(1)}{m(x\,|\,\pi_0)}\sup_{r\geqslant 1/(2h)}\frac{1}{4r^2}D(r,x,\tau^2),$$

where

$$D(r,x,\tau^2)=\frac{\tau^2/\sqrt{\tau^2-1}}{\psi(x/\sqrt{\tau^2-1})}\left\{\Phi\left[(r-\mu(x,\tau^2))/\sigma(x,\tau^2)\right]\right.$$

$$\left.-\Phi\left[(-r-\mu(x,\tau^2))/\sigma(x,\tau^2)\right]\right\}-\frac{\sqrt{\tau^2+1}}{\psi(x/\sqrt{\tau^2+1})}\{\Phi(r-x)-\Phi(-r-x)\}^2.$$

For selected values of τ and x, the corresponding upper bounds $A(U[-t,t])$ and $C^*=A[U[-t,t]]/m(x\,|\,\pi_0)$ are listed in Table 1. The upper bounds on $C(q)$ corresponding to these values can be obtained by finding $\phi''(1)C^*$.

From the extremely large values for C^* corresponding to $\tau=1.1$ and $x=3,4$, we note that the data are not compatible with $\pi_0(\theta)$. However, the same data are quite compatible with $\pi_0(\theta)$ if $\tau=1.5$ or 2. Of course, these observations are not anything new to Bayesians. Our point here is simply that we have developed the new tool consisting of upper bounds on $C(q)$ to make these observations, and that the same technique can be used in other more complicated situations.

Table 1
Upper bounds on curvature in the normal case

τ	x	t	$A(U[-t,t])$	C^*
1.1	2	13.514	98.7180	909.3
	3	20.015	7293530	2.08225×10^8
	4	26.226	7.648×10^{13}	1.06395×10^{16}
1.5	2	4.760	0.1306	1.0918
	3	6.489	0.7605	13.7237
	4	8.461	8.5764	454.3244
2.0	3	5.448	0.0811	1.1186
	4	6.700	0.2556	7.0946

Table 2
Upper bounds on curvature in the binomial case

τ	$\hat{p}$	t	$A(U[0.5-t,0.5+t])$	C^*
0.224	0.80	0.357	660.69	6998.37
0.224	0.833	0.436	1346.86	18570.33
0.288	0.9	0.436	1224.42	13468.66

Example 2. Suppose $X\,|\,\theta\sim\text{binomial}(n,\theta)$ and under π_0, $\theta\sim\text{beta}(\alpha_0,\alpha_0)$. Then Theorem 2.1 gives

$$\sup_{q\in Q_{US}} C(q)=\frac{\phi''(1)}{m(x\,|\,\pi_0)}\sup_{1/2\geqslant r\geqslant 1/(2h)}\frac{1}{4r^2}D(r,x,\tau^2),$$

where

$$D(r,x,\tau^2)=\frac{\Gamma^2(\alpha_0)}{\Gamma(2\alpha_0)}\begin{bmatrix}n\\x\end{bmatrix}\int_{-r}^{r}\theta^{x-\alpha_0+1}(1-\theta)^{n-x-\alpha_0+1}\,\mathrm{d}\theta$$
$$-\frac{\Gamma^2(\alpha_0)}{\Gamma(2\alpha_0)}\frac{\Gamma(n+2\alpha_0)}{\left[\begin{smallmatrix}n\\x\end{smallmatrix}\right]\Gamma(x+\alpha_0)\Gamma(n-x+\alpha_0)}\left[\int_{-r}^{r}\begin{bmatrix}n\\x\end{bmatrix}\theta^{x}(1-\theta)^{n-x}\,\mathrm{d}\theta\right]^2.$$

For selected values of $\tau^2=0.25/(2\alpha+1)$ and $\hat{p}=x/n$, we have verified the validity of Assumption 1. The corresponding upper bounds $A(U[0.5-t,0.5+t]$ and $C^*=A(U[0.5-t,$ $0.5+t)]/m(x\,|\,\pi_0)$ are listed in Table 2. The upper bounds on $C(q)$ corresponding to these values can be obtained as in the normal example by finding $\phi''(1)C^*$.

The behavior of these upper bounds is also very similar to that in the normal case. For example, if τ is fixed at 0.224, as $\hat{p}$ increases from 0.8 to 0.833, C^* increases from 6998.37 to 18570.33. However, if τ is now increased from 0.224 to 0.288, C^* will actually decrease to 13468.66 even when $\hat{p}$ is increased to a more extreme value such as 0.9.

Appendix

Proof of Theorem 2.1. As mentioned earlier this will be done in two steps.

Step 1: We shall show here that, whenever r_i, $i=1,2$, satisfy $V(r_i)\geqslant 1/h$ and $A(U(S(r_1)))>A(U(S(r_2)))$, then $A(U(S(r_1)))\geqslant A(\alpha U(S(r_1))+(1-\alpha)U(S(r_2)))$ for any $0\leqslant\alpha\leqslant 1$.

Assume $r_1<r_2$. The other case is exactly similar. Let

$$a_1=\int_{\|\theta\|\leqslant r_1}\frac{f(x\,|\,\theta)}{\pi_0(\theta)}\mathrm{d}\theta,\qquad a_2=\int_{r_1<\|\theta\|\leqslant r_2}\frac{f(x\,|\,\theta)}{\pi_0(\theta)}\mathrm{d}\theta,$$

$$b_1=\int_{\|\theta\|\leqslant r_1}f(x\,|\,\theta)\,\mathrm{d}\theta,\qquad b_2=\int_{r_1<\|\theta\|\leqslant r_2}f(x\,|\,\theta)\,\mathrm{d}\theta.$$

Note that

$$A(U(S(r_1)))=\frac{1}{V(r_1)^2}\left[a_1-\frac{1}{m(x\,|\,\pi_0)}b_1^2\right]$$

and

$$A(\alpha U(S(r_1))+(1-\alpha)U(S(r_2)))=\left[\frac{\alpha}{V(r_1)}+\frac{1-\alpha}{V(r_2)}\right]^2 a_1+\left[\frac{1-\alpha}{V(r_2)}\right]^2 a_2$$

$$-\frac{1}{m(x\,|\,\pi_0)}\left[\left[\frac{\alpha}{V(r_1)}+\frac{1-\alpha}{V(r_2)}\right]b_1+\left[\frac{1-\alpha}{V(r_2)}\right]b_2\right]^2.$$

The desired inequality is now proved by simple algebra.

Step 2: As in Assumption 1, let t be such that

$$A(U(S(t)))=\sup_{V(r)\geqslant 1/h} A(U(S(r))).$$

Consider the mixture $\alpha U(S(r))+(1-\alpha)U(S(t))$ of densities $U(S(r))$ and $U(S(t))$ for $0\leqslant V(r)\leqslant 1/h$. We claim that, under Assumption 1,

$$\max A(\alpha U(S(r))+(1-\alpha)U(S(t)))=A(U(S(t))),$$

where the maximum is over all r, $0\leqslant V(r)\leqslant 1/h$, and over all α, $0\leqslant\alpha\leqslant 1$, such that

$$\frac{\alpha}{V(r)}+\frac{1-\alpha}{V(r)}\leqslant h.$$

We obtain

$$A(\alpha U(S(r))+(1-\alpha)U(S(t)))\leqslant\left[\frac{\alpha}{V(r)}+\frac{1-\alpha}{V(t)}\right]^2 B(r)+\left[\frac{1-\alpha}{V(t)}\right]^2[B(t)-B(r)].$$

Let $u=a/V(r)+(1-a)/V(t)$. Then $u\in[1/V(t),h]$. Also for given u and r, $a=V(r)(V(t)u-1)/(V(t)-V(r))$ so that $(1-a)/V(t)=(1-uV(r))/(V(t)-V(r))$. Define

$$H(u,r)=u^2B(r)+\left[\frac{1-uV(r)}{V(t)-V(r)}\right]^2[B(t)-B(r)]. \tag{A1}$$

We shall show that

$$\max H(u,r)\leqslant A(U(S(t))), \tag{A2}$$

where the maximum is over all u and r in the admissible range. It follows that $(\partial^2 H(u,r)/\partial u^2)>0$, for all u, if $B(t)>B(r)$. From Assumption 1, we have $\max_{0\leqslant V(r)\leqslant 1/h}B(r)=B(r^*)$, where $r^*=(hV(1))^{-1/p}$ is such that $V(r^*)=1/h$, p being the dimension of θ. Since

$$\sup_{V(r)\geqslant 1/h}\frac{B(r)}{V(r)^2}=\sup_{V(r)\geqslant 1/h}A(U(S(r)))=A(U(S(t)))=\frac{B(t)}{V(t)^2},$$

we further have that $h^2 B(r^*) \leqslant B(t)/V(t)^2$. Consequently, $B(r^*) \leqslant B(t)$ since $V(t) \geqslant 1/h$. Therefore, $\max_{0 \leqslant V(r) \leqslant 1/h} B(r) \leqslant B(t)$, and hence $H(u,r)$ is a convex function of u for each r. It follows, then, for each r,

$$\max_{1/V(t) \leqslant u \leqslant h} H(u,r) = \max \left\{ H\left[\frac{1}{V(t)}, r\right], H(h,r) \right\}.$$

Finally, it follows that

$$\max_{0 \leqslant V(r) \leqslant 1/h} H(h,r) = h^2 \max_{0 \leqslant V(r) \leqslant 1/h} \left\{ B(r) + \left[\frac{1/h - V(r)}{V(t) - V(r)}\right]^2 [B(t) - B(r)] \right\}.$$

Since $B(0) = 0$, from Assumption 1(ii), we have

$$h^2 \max_{0 \leqslant V(r) \leqslant 1/h} \left\{ B(r) + \left[\frac{1/h - V(r)}{V(t) - V(r)}\right]^2 [B(t) - B(r)] \right\}$$
$$= h^2 \left[\frac{1}{hV(t)}\right]^2 B(t) = A(U(S(t))).$$

References

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer, New York.

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

Dey, D.K. and L.R. Birmiwal (1990). Robust Bayesian analysis using entropy and divergence measures. *Technical Report*, University of Connecticut, Storrs.

Gelfand, A.E. and D.K. Dey (1990). On Bayesian robustness of contaminated classes of priors. *Statist. Decisions* **9**, 63–80.

Sivaganesan, S. and J. Berger (1989). Ranges of posterior measures for priors with unimodal contaminations. *Ann. Statist.* **17**, 868–889.

Sivaganesan, S. (1991). Sensitivity of some Bayesian estimates to prior uncertainty: a comparison. *Technical Report*, University of Cincinnati.