Journal of Statistical Planning and Inference 71 (1998) 151–162

journal of statistical planning and inference

# Stability and infinitesimal robustness of posterior distributions and posterior quantities

Sanjib Basu[a,*], Sreenivasa Rao Jammalamadaka[b], Wei Liu[c,1]

[a] *Northern Illinois University, Division of Statistics, DeKalb, IL 60115-2854, USA*

[b] *University of California, Santa Barbara, USA*

[c] *Syntex Corporation, Palo Alto, USA*

Received 28 June 1996; accepted 17 February 1998

**Abstract**

Infinitesimal sensitivities of the posterior distribution $P(\cdot|X)$ and posterior quantities $\rho(P)$ w.r.t. the choice of the prior $P$ are considered. In a very general setting, the posterior $P(\cdot|x)$ and posterior quantities $\rho(P)$ are treated as functions of the prior $P$ on the space $\mathscr{M}$ of all probability measures. Qualitative robustness and stability, loosely, then amount to checking if these functions satisfy continuity and Lipschitz condition of order 1. They thus depend on the underlying topology and metric on the space $\mathscr{M}$. It is proved that posterior $P(\cdot|X)$ and posterior quantity $\rho(P)$ are qualitatively robust in the total variation topology as well as in the weak topology under mild conditions. Qualitative robustness of the Bayes risk, on the other hand, requires rather strong conditions. Stability of posteriors and posterior quantities are also established. An intriguing example shows that simply continuity and boundedness of the likelihood are not enough to guarantee stability of the posterior under the weak convergence metrics. 

*AMS classification*: 62F15; 62G35; 60B05

*Keywords*: Bayes risk; Infinitesimal sensitivity; Lipschitz condition; Posterior robustness; Probability metrics; Stability; Weak convergence

## 1. Introduction

Bayesian paradigm requires the specification of two inputs; the sampling density $f(X|\theta)$, and the prior $P(\theta)$. A decision theoretic framework requires one to further specify a third quantity, namely the loss function $L(\theta, a)$. Since such specifications are only approximations or guesses at best, one would like to study the sensitivity of the

* Corresponding author. E-mail: basu@niu.edu.

[1] Some part of this work was part of his Ph.D. dissertation at University of California, Santa Barbara.

0378-3758/98/$19.00 
PII: S0378-3758(98)00090-1

final actions to these various inputs. In this article, we concentrate on studying the effect of imprecision of the prior distribution $P(\theta)$.

Sensitivity of Bayesian analysis to the choice of the prior has received considerable attention. We refer the reader to review article by Berger (1994) and the references therein. The global sensitivity approach considers a class of all plausible priors and measures the sensitivity by finding the ranges of relevant posterior quantities. Many recent works are directed towards infinitesimal or local sensitivity which examines the effect of infinitesimal perturbations of the prior (see Basu, 1996).

The well developed robustness theory in frequentist statistics (see Huber, 1981; Hampel et al., 1986), on the other hand, is built on three central concepts: qualitative robustness, influence function, and the breakdown point. As Huber explains them: (1) qualitative robustness – a small perturbation (in the model input) should have a small effect; (2) the influence function measures the effect of infinitesimal perturbations (in the model input); and (3) the breakdown point tells us how big the perturbation can be before the statistical quantity (on which inference is based) becomes totally unreliable. Loosely, these ideas correspond to (and depend on) the continuity, first derivative, and the distance of the nearest singularity, respectively, of the posterior quantity. The global and local sensitivity approach of Bayesian robustness are similar to the concept of the breakdown point and the influence curve. One major focus of this article is towards examining (1) in the context of Bayesian analysis, i.e., studying qualitative robustness or continuity of posteriors and posterior quantities. However, this is not the only emphasis. We further develop results that are applicable towards both (2) and (3) in the context of Bayesian analysis.

### *1.1. Notations and definitions*

Before proceeding further, we clarify notations. Let $\Theta$ be the parameter space. In typical parametric problems $\Theta \subseteq \mathrm{R}^k$ whereas in nonparametric problems $\Theta$ is often a subset of the space of probability measures. In general, we assume $\Theta$ is a Polish (complete, separable and metrizable) space, equipped with a metric $\kappa(\cdot,\cdot)$. Let $\mathscr{M}$ be the space of all probability measures on $(\Theta,\mathscr{B})$ where $\mathscr{B}$ is the Borel-$\sigma$-field on $\Theta$. Let $d(\cdot,\cdot)$ be an appropriate metric on $\mathscr{M}$ (to be specified later). We assume that the sampling distributions $Q_\theta$ of data $X|\theta$ form a dominated family with density (w.r.t. the common dominating measure) $f(X|\theta)$. We use standard terminology, $X$ denotes a random variable while $x$ denotes observed data. We use $P(\cdot)$ ($\in \mathscr{M}$) to denote a prior probability measure. Let $m(P|x) = \int_\Theta f(x|\theta)\,\mathrm{d}P(\theta)$ denote the marginal w.r.t. prior $P$. If $m(P|x) > 0$, $P(\cdot|x)$, defined as $P(A|x) = (1/m(P|x)) \int_A f(x|\theta)\,\mathrm{d}P(\theta)$ for any $A \in \mathscr{B}$, is the posterior probability measure corresponding to the prior $P(\cdot)$. We will implicitly assume that $m(P|x) > 0$ whenever we mention posterior $P(\cdot|x)$ or posterior quantity $\rho_P$. $\pi(\cdot)$ and $\pi(\cdot|x)$ will denote the prior and the posterior densities, respectively (whenever appropriate). We use $\rho_P$ or $\rho(P)$ to denote a posterior quantity (such as the posterior mean) corresponding to the prior $P$. Since our focus here is on imprecise specification of the prior $P$, we suppress the dependence on $f$ in all posterior notations.

Hampel (1971) first used the term "qualitative robustness" in the context of classical robustness. He defined it as equicontinuity of the distribution of the relevant statistic as the sample size $n$ changes. Since both the sample size $n$ and data $x$ are fixed in Bayesian analysis, we use the term qualitative robustness to imply continuity in term of the prior $P$. The formal definition follows.

**Definition 1.** (i) The posterior measure $P(\cdot|x)$ is called *qualitatively robust* at $P_0$ if $\forall \varepsilon > 0, \exists \delta > 0$ such that $d(P,P_0) < \delta \Rightarrow d(P(\cdot|x), P_0(\cdot|x)) < \varepsilon$.
(ii) Let $\rho_P$ be a posterior quantity taking values in some space $\Omega$ equipped with a metric $v(\cdot,\cdot)$. $\rho_P$ is called *qualitatively robust* at $P_0$ if $\forall \varepsilon > 0, \exists \delta > 0$ such that $d(P,P_0) < \delta \Rightarrow v(\rho_P, \rho_{P_\circ}) < \varepsilon$.

The main concern of qualitative robustness is that if we know priors $P$ and $P_0$ are "close", can we guarantee that the two resulting posteriors (or posterior quantities) will also be "close"? Is this a serious concern? To answer this, we quote the following from Berger's (1985) discussion on Diaconis and Ylvisaker's (1985) article on prior approximation: "There is a very serious issue concerning such an approximation, however, namely the issue of whether this good approximation to the prior ensures that the posterior will also be well approximated. I think the answer, in general, is no". For other studies on whether approximately correct priors lead to approximately correct posteriors, see Diaconis and Freedman (1986) and Kadane and Chuang (1978). Some of our results on qualitative robustness have already been mentioned in these articles, either for special cases or as casual remarks without proofs. For example, in the special case when $\Theta = \mathrm{R}$, part of Result 6 is available in Kadane and Chuang (1978). Our Result 1 is implicit in Diaconis and Freedman's (1986) work on the derivative of posteriors. We unify these widely dispersed special case results in the very general setting of a Polish space, and present them in a natural order.

We next examine if the posterior $P(\cdot|x)$ or a posterior quantity $\rho_P$ satisfy a Lipschitz condition of order 1 in terms of the prior $P$. Following Męczarski and Zieliński (1991), we call this notion *stability*.

**Definition 2.** (i) The posterior measure $P(\cdot|x)$ is *stable* at prior $P_0$ if $\exists \delta > 0$ and $M > 0$ such that for $\forall$ priors $P$ which satisfies $d(P,P_0) < \delta$, we have $d(P(\cdot|x), P_0(\cdot|x)) < M\, d(P,P_0)$.
(ii) A posterior quantity $\rho_P$ is *stable* at prior $P_0$ if $\exists \delta > 0$ and $M > 0$ such that for $\forall P$ satisfying $d(P,P_0) < \delta$, $v(\rho_P, \rho_{P_\circ}) < M d(P,P_0)$.

Why is the concept of stability of interest? First, a Lipschitz condition for the posterior, a notion which is in between continuity and existence of derivatives, is clearly of mathematical interest. The second interest comes from a robustness viewpoint. If our imprecisely elicited prior lies within a $\delta$-neighborhood of the "true" prior, what stability guarantees is that the oscillation of the posterior (or a posterior quantity) is $\mathrm{O}(\delta)$ as $\delta \to 0$. Męczarski and Zieliński (1991) explore this notion in a specific parametric problem, our examination here is completely general. Thirdly, stability almost ensures Fréchet differentiability, though not completely. It guarantees that the relevant limit which defines the derivative is bounded.

### 1.2. Probability metrics

We will use the following well-known metrics ($d(\cdot,\cdot)$) on the space $\mathscr{M}$ (see Huber, 1981):

(A) *Total variation metric*: $d_T(P,Q)=\sup_{A\in\mathscr{B}}|P(A)-Q(A)|$. Note that the $L_1$ metric $d_{L_1}(P,Q)=\int_\Theta|\mathrm{d}P(\theta)-\mathrm{d}Q(\theta)|$ is equivalent to the total variation metric due to the relation $d_{L_1}(P,Q)=2\,d_T(P,Q)$;

(B) *Prohorov metric*: $d_P(P,Q)=\inf\{\varepsilon>0: P(A)\leqslant Q(A^\varepsilon)+\varepsilon\ \forall A\in\mathscr{B}\}$, where $A^\varepsilon=\{\eta\in\Theta:\inf_{\theta\in A}\kappa(\theta,\eta)<\varepsilon\}$;

(C) *Dudley metric*: Let $\mathscr{F}=\{f:\Theta\to\mathrm{R}$; such that $\sup_{\theta\in\Theta}|f(\theta)|\leqslant 1$, and $|f(\theta)-f(\eta)|\leqslant\kappa(\theta,\eta)\ \ \forall\theta,\eta\in\Theta\}$. Then, $d_D(P,Q)=\sup\{|\int f(\theta)\,\mathrm{d}P(\theta)-\int f(\theta)\,\mathrm{d}Q(\theta)|: f\in\mathscr{F}\}$ (see Dudley, 1976); and

(D) *Lévy metric*: This metric is typically used when $\Theta\subseteq\mathrm{R}^n$. We have $d_L(P,Q)=\inf\{\varepsilon>0: P(X\in(-\infty,\underset{\sim}{\theta}])\leqslant Q(Y\in(-\infty,\underset{\sim}{\theta}]^\varepsilon)+\varepsilon$, and $Q(Y\in(-\infty,\underset{\sim}{\theta}])\leqslant P(X\in(-\infty,\underset{\sim}{\theta}]^\varepsilon)+\varepsilon,\ \ \forall\underset{\sim}{\theta}\in\Theta\}$ where $\underset{\sim}{\theta}=(\theta_1,\ldots,\theta_n)^\mathrm{T}\in\Theta\subseteq\mathrm{R}^n$, $(-\infty,\underset{\sim}{\theta}]$ denotes $\bigotimes_{i=1}^n(-\infty,\theta_i]$, and $A^\varepsilon$ is as in (B) with $\kappa(\cdot,\cdot)$ being the Euclidean distance on $\Theta$.

It is well known that the later three metrics metrize the weak convergence topology whereas the total variation metric metrizes the total variation or strong topology. The following inequalities are also well known: $d_L\leqslant d_P\leqslant d_T$ and $d_P(P,Q)^2\leqslant 2d_D(P,Q)\leqslant 4d_P(P,Q)$ (see Dudley, 1976; Huber, 1981). We will use these inequalities in Corollary 1.

The rest of the paper is organized as follows. In Section 3, we explore qualitative robustness and stability of posterior distributions, posterior risks and Bayes risks in total variation metric. Qualitative robustness and stability under weak convergence metrics are studied in Section 4. We conclude with a brief discussion in Section 5.

## 2. Motivation

Real-life data are either univariate or at most multivariate. What then is the usefulness of putting the abstract structure of a Polish space on the parameter space $\Theta$ while analyzing such data? To answer this question, we point to the recent vigorous growth of Bayesian nonparametric analysis. In such analysis, the parameter under consideration is often a function, such as the sampling density or the sampling cdf. Hence, the resulting parameter space turns out to be an abstract function space. To point out a specific case, Bayesian models using Dirichlet process priors have received considerable attention recently. In these models, typically, data $x$ are assumed to be observed from a probability measure $Q(x)$. No parametric structure is assumed on $Q$. Rather the whole probability measure $Q$ is treated as the parameter with parameter space $\Theta = \mathscr{M}$ or a subset of $\mathscr{M}$. Finally, a Dirichlet process (DP) prior or a mixture of Dirichlet process (MDP) prior is assumed on $\Theta$. In this setup, a prior robustness investigation requires consideration of the Polish space structure of $\Theta$.

In nonparametric Bayes, an alternative to the Dirichlet process prior is a Gaussian process prior. The logistic transformation of the Gaussian process prior has been used by Leonard (1978), Thornburn (1986) and Lenk (1988, 1991). A Gaussian process prior is generally indexed by a mean function $\mu(\cdot)$ and a variance function $v(\cdot)$.

Suppose we have a base Gaussian process prior $GP_0(\mu_0, v_0)$ with mean function $\mu_0(\cdot)$, variance function $v_0(\cdot)$. Let $GP(\mu, v)$ be an approximating Gaussian process prior. It is known that if the variance functions are not identical, i.e., $v_0(\cdot) \neq v(\cdot)$, then the total variation distance between the base and the approximating prior processes is $d_T(GP_0(\mu_0, v_0),\ GP(\mu, v)) = 2$, i.e., the maximum possible value under the total variation distance. If (i) $v_0(\cdot) \equiv v(\cdot)$; (ii) the function $\mu_0(\cdot) - \mu(\cdot)$ is absolutely continuous with respect to the common variance $v_0(\cdot)$; and (iii) another technical condition is satisfied, then

$$2\left\{1 - \exp\left(-\frac{1}{8}\int p^2 dv_0\right)\right\} \leqslant d_T(GP_0(\mu_0, v_0),$$
$$GP(\mu, v_0)) \leqslant 2\sqrt{1 - \exp\left(-\frac{1}{4}\int p^2\, \mathrm{d}v_0\right)} \tag{1}$$

where $p(\cdot)$ is the Radon–Nikodym derivative of $\mu_0(\cdot) - \mu(\cdot)$ with respect to $v_0(\cdot)$. If any of the above three conditions (i)–(iii) is not satisfied, then $d_T(GP_0(\mu_0, v_0), GP(\mu, v)) = 2$, the maximum possible value. For proofs and further discussion of these results, see Liese and Vajda (1987) (pp. 61–63).

Once again, suppose we have a base Gaussian process prior $GP(\mu_0, v_0)$ and an approximating Gaussian process prior $GP(\mu, v_0)$ where $\mu_0(\cdot) - \mu(\cdot)$ is absolutely continuous with respect to $v_0(\cdot)$ and $p(\cdot) = d(\mu_0(\cdot) - \mu(\cdot))/\mathrm{d}v_0(\cdot)$. If $p(\cdot) \xrightarrow{L^2(v_0)} 0$, i.e., if $\int p^2\, \mathrm{d}v \to 0$, it then follows from (1) that $d_T(GP_0(\mu_0, v_0),\ GP(\mu, v_0)) \to 0$. Can we guarantee from here that the two posteriors resulting from these two Gaussian process priors will also be close to each other in total variation topology? The full strength of our results in the general setup of abstract Polish spaces come into force here. Our Result 1 guarantees that as long as the likelihood is bounded, the posteriors resulting out of the two Gaussian process priors will also converge (no matter how complicated the resulting posteriors are). Our Result 4 further guarantees that the resulting posteriors will, in fact, be stable in the total variation distance.

## 3. Total variation metric

In this section, we establish qualitative robustness and stability of posterior distributions and posterior quantities in total variation metric. The typical assumption required for all results of this section is that the involved functions be bounded. For example, to establish either qualitative robustness or stability of the posterior distribution $P(\cdot|x)$, we require the likelihood $f(x|\theta)$ to be a bounded function of $\theta$. To establish the same for a ratio-linear posterior quantity $\rho^h(P)=\int_\Theta h(\theta)\,f(x|\theta)\,\mathrm{d}P(\theta)/m(P|x)$, we additionally need to assume that $h(\theta)\,f(x|\theta)$ is a bounded function in $\theta$. Similarly, to establish qualitative robustness of the posterior risk, we assume that $f(x|\theta)\cdot L(\theta,a)$ is bounded in $\theta$ where $L(\theta,a)$ is the loss function. We repeat these assumptions again in the stated results for clarity.

### 3.1. Qualitative robustness

Qualitative robustness w.r.t. the prior $P$ is the question of continuity, i.e., does approximate priors guarantee approximate posteriors? In the following three results, we establish qualitative robustness of posterior distributions, posterior quantities, posterior risks and Bayes risks in total variation topology.

**Result 1.** *Let $\mathscr{M}$ be equipped with the total variation topology, generated by $d_T$. Suppose the likelihood $f(x|\theta)$, for observed data x, is a bounded function of $\theta$. Then the posterior $P(\cdot|x)$ is qualitatively robust at every $P_0\in\mathscr{M}$.*

*Additionally, suppose $\rho^h(P)=\int_\Theta h(\theta)f(x|\theta)\,\mathrm{d}P(\theta)/m(P|x)$ is a real valued ratio-linear posterior quantity where the function $h(\theta)\cdot f(x|\theta)$ is bounded in $\theta$. Then, under the Euclidean metric on the real line, the posterior quantity $\rho^h(P)$ is qualitatively robust at every $P_0\in\mathscr{M}$.*

**Proof.** See Basu et al. (1993). □

For the following two results, we slightly change our direction and focus on a decision theoretic framework. Let $L(\theta,a)$ denote the loss function for parameter $\theta$ and action $a$. Under observed data $x$ from the sampling distribution $f(x|\theta)$ and prior $P$ on the parameter space $\Theta$, the expected posterior loss (or posterior risk) of an action $a$ is given by $r(P,a)=\int L(\theta,a)f(x|\theta)\,\mathrm{d}P(\theta)/m(P|x)$. Our next result examines the qualitative robustness of the posterior risk. This result is an immediate corollary of Result 1.

**Result 2.** *Suppose $f(x|\theta)\cdot L(\theta,a)$ is a bounded function of $\theta$ for observed data x and fixed action a. Then the posterior risk $r(P,a)$ is qualitatively robust under the total variation metric $d_T$.*

Our next investigation focuses on the posterior Bayes risk $r(P)=\inf_a r(P,a)$. Since $r(P)$ involves infimum over all actions instead of a single fixed $a$, its robustness requires stronger assumptions.

**Result 3.** *Suppose $L(\theta,a)\cdot f(x|\theta)$ is uniformly bounded in $\theta$ and a, i.e., $|L(\theta,a)\cdot f(x|\theta)|\leqslant M$ for all $\theta\in\Theta$ and for all actions a. The posterior Bayes risk $r(P)$ is then qualitatively robust under the total variation metric $d_T$.*

**Proof.** Take a prior sequence $\{P_n\}$ converging to a prior $P_0$ in the total variation topology. Quick calculations yield $|r(P_0,a)-r(P_n,a)| \leqslant (M/m(P_0)) \int_\Theta |\mathrm{d}P_0(\theta)-\mathrm{d}P_n(\theta)| + M|1/m(P_n) - 1/m(P_0)|$ for every action $a$. However, the r.h.s. is independent of $a$, and tends to zero as $n \to \infty$. Hence $r(P_n,a) \to r(P_0,a)$ uniformly in $a \in \mathscr{A}$.

For every $n$ and $a$, $r(P_n,a) \geqslant r(P_n)$. Hence, $r(P_0) = \inf_a r(P_0,a) = \inf_a \lim_{n\to\infty} r(P_n,a) \geqslant \lim_{n\to\infty} r(P_n)$. To prove the other inequality, fix $\varepsilon > 0$, and for every fixed $n$, find $a_n$ such that $r(P_n) \geqslant r(P_n,a_n) - \varepsilon$. Next, find $N$ such that $n \geqslant N \Rightarrow |r(P_n,a)-r(P_0,a)| < \varepsilon \; \forall a \in \mathscr{A}$. Hence, for $n \geqslant N$, $r(P_n) \geqslant r(P_n,a_n) - \varepsilon \geqslant r(P_0,a_n) - 2\varepsilon \geqslant r(P_0) - 2\varepsilon$. This completes the proof since $\varepsilon$ is arbitrary. □

### 3.2. Stability

Stability provides us with a stronger robustness criterion than qualitative robustness. It is also a big step forward, towards establishing the differentiability of the posterior w.r.t the prior. The next result establishes stability of posteriors and posterior quantities under the total variation metric.

**Result 4.** *Let $\mathscr{M}$ be metrized by the total variation metric $d_T$. If the likelihood $f(x|\theta)$ is a bounded function of $\theta$ then the posterior $P(\cdot|x)$ is stable at every $P_0 \in \mathscr{M}$.*

**Proof.** Let $f(x|\theta) \leqslant M \; \forall\theta$. Take a prior $P_0 \in \mathscr{M}$. Since $m(P_0) > 0$ and $m(P|x)$ is continuous at $P_0$, $\exists$ a neighborhood $\mathscr{P}_0$ of $P_0$ such that $m(P|x)$ is bounded away from zero on $\mathscr{P}_0$, i.e., $m(P|x) \geqslant \gamma$ (say) $> 0 \; \forall P \in \mathscr{P}_0$. Choose a prior $Q$ from this neighborhood $\mathscr{P}_0$. It follows that $d_T(P_0(\cdot|x), Q(\cdot|x)) \leqslant (M/m(P_0)) \int |\mathrm{d}P_0(\theta) - \mathrm{d}Q(\theta)| + M|m(P_0) - m(Q)|/(m(Q)\,m(P_0)) \leqslant (M/m(P_0))\, d_{L_1}(P_0,Q) + 2M^2\, d_{L_1}(P_0,Q)/\gamma^2$. Since $d_{L_1}(P_0,Q) = 2d_T(P_0,Q)$, this proves that posterior $Q(\cdot|x)$ is stable at the prior $P_0$. □

**Remark.** A similar result holds for ratio-linear posterior quantities.

## 4. Weak convergence metrics

In this section, we investigate the qualitative robustness and stability of posterior distributions and posterior quantities under the Prohorov, Dudley or Lévy metrics which generate the weak convergence topology.

### 4.1. Non-stable posteriors and posterior quantities

We begin this section with two examples. These examples relate to the work of Diaconis and Freedman (1986) who establish Fréchet differentiability of posteriors under total variation topology and comment that the same holds under weak convergence topology if likelihood $f(x|\theta)$ is bounded and continuous in $\theta$. Krasker and Pratt (1986), in their discussion, mention without proof that Fréchet differentiability under the weak topology "require the further assumption that $f(x|\theta)$ satisfy a Lipschitz condition in $\theta$ (continuity does not appear to be enough)". Our Example 1 in the following, explicitly shows that even when the likelihood $f(x|\theta)$ is continuous and uniformly bounded in $\theta$, the posterior may not be stable under the Lévy or the Prohorov metric. This example thus refutes Diaconis and Freedman's (1986) statement and establishes that the Lipschitz condition on the likelihood $f(x|\theta)$ proposed by Krasker and Pratt (1986) is not relaxable for establishing stability and Fréchet differentiability.

**Example 1.** For brevity, we will state this example only in terms of the Lévy metric $d_L$ (the same example works for the Prohorov metric). We will also skip many technical details which can be found in Basu et al. (1993).

Let $\Theta = [0,1]$ be the parameter space. Let $\{\eta_n\}$ be a sequence on $\Theta$ converging to 0 (in Euclidean norm). Assume $0<\eta_n<\gamma/2\ \forall n$ where $0<\gamma<1$ is a fixed number. Define $P_0 = (1-\gamma)U[0,1] + \gamma\delta_{\{0\}}$ and $P_n = (1-\gamma)U[0,1] + \gamma\delta_{\{\eta_n\}}$ where $\delta_{\{\eta\}}$ denotes the dirac measure at $\eta$. Let $F_0$ and $F_n$ be the cdfs corresponding to $P_0$ and $P_n$. To simplify notations, we will often use $\eta$ to denote a generic element of the sequence $\{\eta_n\}$ and $P_\eta$ or $F_\eta$ to denote $(1-\gamma)U[0,1] + \gamma\delta_{\{\eta\}}$.

By definition, $d_L(P_0,P_n) = \inf\{\varepsilon>0 : F_n(\theta-\varepsilon)-\varepsilon \leqslant F_0(\theta) \leqslant F_n(\theta+\varepsilon)+\varepsilon\ \forall\theta\}$. It is easy to see that the inside relation holds if we take $\varepsilon=\eta_n$. On the other hand, if we take $\varepsilon<\eta_n$ then for $\theta=0$, $F_0(0) > F_n(\varepsilon)+\varepsilon$. This shows that $d_L(P_n,P_0)=\eta_n$. Since $\eta_n\to 0$, we have the prior sequence $P_n$ converging to the prior $P_0$ in terms of the Lévy metric. Note, however, that for $A=\{0\}$, $P_0(A)=\gamma>0$, but $P_n(A)=0\ \forall n$. Hence $P_n$ does not converge to $P_0$ in the total variation metric.

We next fix the likelihood function. Suppose we observe $X=1$ from a Bernoulli distribution with parameter $p$. However, suppose our interest is not in the parameter $p$, but in $\theta=p^2$. In terms of $\theta$, we then have the likelihood function $f(x|\theta) = 1-\sqrt{\theta}$, $\theta\in[0,1]$. Notice that the likelihood $f(x|\theta)$ is bounded and, in fact, uniformly continuous on $\Theta=[0,1]$.

Once we have the likelihood $f(x|\theta)$ set up, we proceed to find the Lévy distance between $P_0(\cdot|x)$ and a generic member $P_\eta(\cdot|x)$ of the posterior sequence $\{P_n(\cdot|x)\}$, i.e. $d_L(P_\eta(\cdot|x),P_0(\cdot|x))$ defined as $\inf\{\varepsilon>0 : F_\eta(\theta-\varepsilon|x)-\varepsilon \leqslant F_0(\theta|x) \leqslant F_\eta(\theta+\varepsilon|x)+\varepsilon\ \forall\theta\}$.

Suppose we take $\theta=0$ in the above and write the function $g_\eta(\varepsilon) = F_0(0|x) - F_\eta(\varepsilon|x) - \varepsilon$. Then, it can be shown through analytical arguments and numerical evaluations that: (i) $g_\eta(\varepsilon)>0$ for $0<\varepsilon\leqslant\eta$ whenever $0\leqslant\eta<0.01367$; and (ii) let $\varepsilon^*_\eta$ be the smallest nonnegative root of $g_\eta(\varepsilon)=0$. Then $\varepsilon^*_\eta \geqslant 0.1\sqrt{\eta}$ for all $\eta<0.0001$. This shows that $g_\eta(\varepsilon) = F_0(0|x) - F_\eta(\varepsilon|x) - \varepsilon > 0$ for $0<\varepsilon\leqslant\varepsilon^*_\eta$ and hence $d_L(P_\eta(\cdot|x),P_0(\cdot|x)) \geqslant \varepsilon^*_\eta \geqslant 0.1\sqrt{\eta}$ whenever $0<\eta\leqslant 0.0001$. Notice that $\sqrt{\eta}$ has an unbounded derivative near 0. Thus, it cannot be dominated by any straight line, i.e., there does not exist $K>0$ such that $\sqrt{\eta}\leqslant K\eta$ for $\eta$ near 0.

On the other hand, we have shown that the Lévy distance between the two priors $P_\eta$ and $P_0$ is $d_L(P_\eta,P_0)=\eta$. This proves that there does not exist any $K>0$ such that

$d_L(P_\eta(\cdot|x), P_0(\cdot|x)) < Kd_L(P_\eta, P_0)$, i.e., $P(\cdot|x)$ is *not* stable at $P_0$ in terms of Lévy metric. However, $P(\cdot|x)$ is clearly qualitatively robust in the Lévy metric (which follows immediately from Result 6).

**Example 2.** The same example also shows that even if both $h(\theta)$ and $f(x|\theta)$ are continuous and uniformly bounded, the ratio-linear posterior quantity $\rho^h(P)$ may not be stable in terms of the Lévy metric. Consider $h(\theta) = \theta$ in the above example. Then $h(\theta)$ is bounded and (uniformly) continuous on $\Theta = [0, 1]$. Take $\gamma = 0.5$, i.e., $P_\eta = 0.5U[0, 1] + 0.5\delta_{\{\eta\}}$. The posterior mean for this prior $P_\eta$ and likelihood $f(x|\theta) = 1 - \sqrt{\theta}$ (as specified in Example 1) is $\rho^h(P_\eta) = \{3 + 30\eta(1 - \sqrt{\eta})\}/\{10(4 - 3\sqrt{\eta})\}$. Quick algebra shows that $\lim_{\eta\to 0}\{\mathrm{d}\rho^h(P_\eta)/\mathrm{d}\eta\} = \infty$ which implies that the posterior mean $\rho^h(P)$ is *not* stable at the prior $P_0$.

### *4.2. Stability under Dudley and Prohorov metrics*

In our next result, we establish that when the likelihood $f(x|\theta)$ satisfies a Lipschitz condition, the resulting posterior is stable under the Dudley metric $d_D$. Since the Dudley metric itself is defined in terms of Lipschitz functions (Huber, 1981), in fact, calls it the bounded Lipschitz metric), it blends easily with a Lipschitz likelihood to give us stability of the posterior. We then use this result and the relation between Dudley and Prohorov metrics to show stability of ratio-linear posterior quantities in terms of the Prohorov metric in Corollary 1.

**Result 5.** *Let the prior space $\mathscr{M}$ be metrized by the Dudley metric $d_D(\cdot,\cdot)$. Assume* (i) *the likelihood function is a bounded function of $\theta$, i.e., $f(x|\theta) \leqslant M_1\ \forall\theta$; and* (ii) *the likelihood is a Lipschitz function of $\theta$, i.e., $\exists K_1 > 0$ such that $|f(x|\theta) - f(x|\eta)| \leqslant K_1\kappa(\theta,\eta)\ \forall\theta,\eta$ where $\kappa(\cdot,\cdot)$ is the metric on the parameter space $\Theta$. Then the posterior $P(\cdot|x)$ is stable at every $P_0 \in \mathscr{M}$.*

*If moreover, we have a real valued function $h(\theta)$ which satisfies* (iii) $|h(\theta)f(x|\theta)| \leqslant M_2\ \forall\theta\in\Theta$, *and* (iv) $|h(\theta)f(x|\theta) - h(\eta)f(x|\eta)| \leqslant K_2\kappa(\theta,\eta)\ \forall\theta,\eta$, *then the ratio-linear posterior quantity $\rho^h(P)$ is also stable at every $P_0 \in \mathscr{M}_0$.*

**Remark.** The conditions assumed on the likelihood $f(x|\theta)$ for Result 5 to hold may appear to be too strong. However, they are often easy to check and hold in many common situations. If the parameter space is a subset of $\mathrm{R}^n$, then ($a$) $f(x|\theta)$ is differentiable; and ($b$) $\mathrm{d}/\mathrm{d}\theta f(x|\theta)$ is uniformly bounded are sufficient to guarantee that the likelihood $f(x|\theta)$ satisfy condition (ii). Similar remarks hold for the function $h(\theta)$.

**Proof of Result 5.** For brevity, we will only prove stability of the posterior $P(\cdot|x)$. The proof for stability of the posterior quantity $\rho^h(P)$ is similar.

Take a prior $P_0 \in \mathscr{M}$. Since $m(P_0) > 0$ and $m(P|x)$ is continuous in weak topology, $\exists$ a $d_D$-neighborhood $\mathscr{P}_0$ of $P_0$ such that $m(P|x) \geqslant \gamma > 0\ \forall P \in \mathscr{P}_0$. Choose prior $Q$ from this neighborhood. Let $\mathscr{F}$ denote the family of all Lipschitz functions which are bounded by 1 (as described in the definition of the Dudley metric). The Dudley distance between the posteriors $P_0(\cdot|x)$ and $Q(\cdot|x)$ is then given by $d_D(P_0(\cdot|x), Q(\cdot|x)) = \sup_{g\in\mathscr{F}} |\int g(\theta) f(x|\theta)\mathrm{d}P_0(\theta)/m(P_0) - \int g(\theta)f(x|\theta)\,\mathrm{d}Q(\theta)/m(Q)| \leqslant \sup_{g\in\mathscr{F}}\{|\int g(\theta)f(x|\theta)[\mathrm{d}P_0(\theta) - \mathrm{d}Q(\theta)]|/m(P_0) + |\int g(\theta)f(x|\theta)\,\mathrm{d}Q(\theta)|\,|m(P_0) - m(Q)|/\{m(P_0)m(Q)\} \leqslant \sup_{g\in\mathscr{F}}\{1\text{st term} + M_1|m(P_0) - m(Q)|/\gamma^2\}$.

For an arbitrary $g\in\mathscr{F}$, let $g^*(\theta) = g(\theta)f(x|\theta)/(K_1 + M_1)$. It is easy to see that $|g^*(\theta)| \leqslant 1$ and $|g^*(\theta) - g^*(\eta)| \leqslant \{|g(\theta)f(x|\theta) - g(\theta)f(x|\eta)| + |g(\theta)f(x|\eta) - g(\eta)f(x|\eta)|\}/(K_1 + M_1) \leqslant \kappa(\theta,\eta)$ since the likelihood $f(x|\cdot)$ and the function $g(\cdot)$ are both bounded and continuous. Thus $g^*(\theta)$ is a member of the family $\mathscr{F}$.

The first term inside the supremum, $|\int g(\theta)f(x|\theta)[\mathrm{d}P_0(\theta) - \mathrm{d}Q(\theta)]|/m(P_0) = (K_1 + M_1)|\int g^*(\theta)[\mathrm{d}P_0(\theta) - \mathrm{d}Q(\theta)]|/m(P_0) \leqslant (K_1 + M_1)d_D(P_0, Q)/m(P_0)$ by definition of $d_D(P_0, Q)$ since $g^*$ is a member of $\mathscr{F}$. Similar arguments show that the second term inside the supremum, $|m(P_0) - m(Q)| \leqslant \max(K_1, M_1)d_D(P_0, Q)$. It now follows that $d_D(P_0(\cdot|x), Q(\cdot|x)) \leqslant \{(K_1 + M_1)/m(P_0) + M_1\max(K_1, M_1)/\gamma^2\}d_D(P_0, Q)$, i.e., the posterior $Q(\cdot|x)$ is stable in the Dudley metric at the prior $P_0$. This completes the proof. □

As mentioned in Section 1.2, the Dudley distance and the Prohorov distance between two measures $P$ and $Q$ satisfy $d_P(P,Q)^2 \leqslant 2d_D(P,Q) \leqslant 4d_P(P,Q)$. We use the second inequality to establish stability of the posterior quantity $\rho^h(P)$ under the Prohorov metric. However, for the complete posterior $P(\cdot|x)$, we can only show that it is Lipschitz of order $\frac{1}{2}$ under the Prohorov metric.

**Corollary 1.** *Assume the conditions for Result* 5. *Then the posterior $P(\cdot|x)$ satisfies a Lipschitz condition of order $\frac{1}{2}$ in terms of the prior $P$ under the Prohorov metric, i.e., for prior $P_0\in\mathscr{M}$ and prior $Q$ in a neighborhood of $P_0$, we have $d_P(P_0(\cdot|x), Q(\cdot|x)) \leqslant K\sqrt{d_P(P_0, Q)}$ where $K>0$ is a fixed number.*

*Additionally, if $h(\theta)f(x|\theta)$ is a bounded Lipschitz function then the ratio-linear posterior quantity $\rho^h(P)$ satisfies a Lipschitz condition of order* 1 *in terms of the prior $P$ and hence stable under the Prohorov metric.*

**Proof.** Stability of the posterior quantity $\rho^h(P)$ is immediate. The result about the posterior $P(\cdot|x)$ follows from the relation between Prohorov and Dudley metrics: $d_P(P_0(\cdot|x), Q(\cdot|x)) \leqslant \sqrt{2d_D(P_0(\cdot|x), Q(\cdot|x))} \leqslant \sqrt{2Md_D(P_0, Q)}$ (Result 5) $\leqslant \sqrt{4M}\sqrt{d_P(P_0, Q)}$. □

### 4.3. *Qualitative robustness under weak convergence metrics*

We, so far, have refrained from discussing qualitative robustness under weak convergence metrics. For these metrics (Prohorov, Dudley or Lévy), the typical assumption required to establish qualitative robustness is that all involved functions be bounded and continuous (as opposed to only boundedness required for the total variation metric). The following result parallels Results 1, 2 and 3 of Section 3. For proof, see Basu, et al. (1993).

**Result 6.** *Let $\mathscr{M}$ be equipped with the weak convergence topology, generated by either the Prohorov metric $d_P$ or the Dudley metric $d_D$ or, if $\Theta \subseteq \mathrm{R}^n$, by the Lévy metric $d_L$.*

(1) *If the likelihood $f(x|\theta)$ is bounded and continuous in $\theta$ then the posterior $P(\cdot|x)$ is qualitatively robust at every $P_0 \in \mathscr{M}$.*

(2) *In addition, if $h(\theta) \cdot f(x|\theta)$ is continuous and bounded, then the ratio-linear posterior quantity $\rho^h(P)$ is qualitatively robust at every $P_0 \in \mathscr{M}$.*

(3) *If the loss function $L(\theta, a)$ is continuous and bounded in $\theta$, the posterior risk $r(P, a)$ is also qualitatively robust.*

(4) *Suppose $L(\theta, a) \cdot f(x|\theta)$ is uniformly bounded in $\theta$ and $a$, and moreover, $L(\theta, a)$ is uniformly continuous in $\theta$ for all actions $a$. Then the posterior Bayes risk $r(P)$ is qualitatively robust.*

## 5. Discussion

The local or infinitesimal approach to sensitivity analysis is an useful technique for checking robustness, and complements the approach based on global sensitivity analysis. Local sensitivity analysis in the context of Bayesian robustness, so far, concentrated on measuring the rate of change of the posterior quantity as the prior varies over the whole class $\mathscr{M}$, or in a subclass of $\mathscr{M}$. Instead, in this article, we look at qualitative robustness and *stability* of posterior distributions and posterior quantities, which are much weaker infinitesimal properties.

Our exploration of qualitative robustness and *stability* yields many interesting results. We show in Result 5 that an additional Lipschitz condition is needed on the likelihood to establish stability in weak convergence metrics. In Example 1 we show that just continuity and boundedness of the likelihood are simply not enough to establish stability under Lévy or Prohorov metrics. Agata Bonotynisho of University of Warsaw has shown us (in written communication) that Example 1 is also true under the Dudley metric, i.e., a continuous bounded likelihood is again not enough to guarantee stability.

After qualitative robustness and stability, the next step in local sensitivity check is the evaluation of the derivative of the posterior (or posterior quantity) w.r.t. the prior. In parametric problems, this leads to considerations of directional and total derivatives whereas in nonparametric problems Fréchet and Gâteaux derivatives come into the picture. Evaluations of these derivatives often lead to many interesting findings. For recent expositions in this area, see Basu et al. (1996) for parametric derivatives and Basu (1996) for nonparametric derivatives.

## Acknowledgements

We have benefited from helpful discussions with Benny Cheng, Anirban DasGupta, Mohit Dayal, and John Hsu. The first author is deeply indebted to Suman Majumdar and Svetlozar T. Rachev for valuable advice and many constructive suggestions. We are grateful to Agata Bonotynisho for pointing out that Example 1 also applies to the Dudley metric. They all have our sincerest appreciation.

## References

Basu, S., 1996. Local sensitivity, functional derivatives and nonlinear posterior quantities. Statist. and Decisions 14, 405–418.

Basu, S., Jammalamadaka, S.R., Liu, W., 1996. Local posterior robustness with parametric priors: maximum and average sensitivity. In: Heidbreder, G. (Ed.), Maximum Entropy and Bayesian Statistics. Kluwer Academic Publisher, Dordrecht pp. 97–106.

Basu, S., Jammalamadaka, S.R., Liu, W., 1993. Qualitative robustness and stability of posterior distributions and posterior quantities. Tech. Rep. 238, University of California, Santa Barbara.

Berger, J., 1994. An overview of robust Bayesian analysis. Test 2.

Berger, J., 1985. Discussion on "Diaconis and Ylvisaker (1985). In: Bernardo, J.M., et al. (Eds.), Bayesian Statistics 2. North-Holland, Amsterdam, pp. 152–154.

Diaconis, P., Freedman, D., 1986. On the consistency of Bayes estimates. Ann. Statist. 14, 1–67.

Diaconis, P., Ylvisaker, D., 1985. Quantifying prior opinion. In: Bernardo, J.M., et al. (Eds.), Bayesian Statistics 2. North-Holland, Amsterdam, pp. 133–156.

Dudley, R.M., 1976. Probabilities and metrics: convergence of laws on metric spaces, with a view to statistical testing, Aarhus University Mathematics Institute Lecture Notes Series, no. 45.

Hampel, F.R., 1971. A general qualitative definition of robustness. Ann. Math. Statist. 42, 1887–1896.

Hampel, F.R., Ronchetti, E.L., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics: The Approach Based on Influence Functions. Wiley, New York.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

Kadane, J.B., Chuang, D.T., 1978. Stable decision problems. Ann. Statist. 6, 1095–1110.

Krasker, W.S., Pratt, J.W., 1986. Discussion on Diaconis and Freedman (1986). Ann. Statist. 14, 55–58.

Lenk, P.J., 1988. The logistic normal distribution for Bayesian, nonparametric, predictive densities. J. Amer. Statist. Assoc. 83, 509–516.

Lenk, P.J., 1991. Towards a practicable Bayesian nonparametric density estimator. Biometrika 78, 531–543.

Leonard, T., 1978. Density estimation, stochastic processes, and prior information. J. Roy. Statist. Soc. Ser. B 40, 113–146.

Liese, F., Vajda, I., 1987. Convex Statistical Distances. Teubner-Texte zur Mathematik, 95.

Męczarski, M., Zieliński, R., 1991. Stability of the Bayesian estimator of the Poisson mean under the inexactly specified gamma prior. Statist. Probab. Lett. 12, 329–333.

Thorburn, D., 1986. A Bayesian approach to density estimation. Biometrika 73, 65–76.