APPLICATIONES MATHEMATICAE
22,1 (1993), pp. 139–143

A. BORATYŃSKA and R. ZIELIŃSKI (Warszawa)

BAYES ROBUSTNESS VIA THE KOLMOGOROV METRIC

Abstract. An upper bound for the Kolmogorov distance between the posterior distributions in terms of that between the prior distributions is given. For some likelihood functions the inequality is sharp. Applications to assessing Bayes robustness are presented.

1. Introduction and notations. Given a sample space $\mathcal{X}$ and a parameter space Θ, let $l : \mathcal{X} \times \Theta \to \mathbb{R}^1$ be a given function such that $(\forall\theta) l(\cdot, \theta)$ is a density of a probability distribution function on $\mathcal{X}$ and $(\forall x)\, l_x(\cdot) = l(x, \cdot)$ is the likelihood function on Θ. Throughout the paper we assume that Θ is an interval (θ_L, θ_U) in $\mathbb{R}^1$, $-\infty \le \theta_L < \theta_U \le +\infty$, and that for every $x \in \mathcal{X}$ the likelihood function $l_x(\cdot)$ is of finite variation; also, we define $s(x) = \sup_{\theta\in\Theta} l_x(\theta)$.

All integrals are Lebesgue–Stieltjes integrals over (θ_L, θ_U) unless stated otherwise. To avoid some technical difficulties we assume that the distribution functions F and G appearing below are continuous. Actually, it is enough to assume that points of discontinuity of F and G do not coincide with those of $l_x(\cdot)$.

Let $l_x(\cdot) = l_x^+(\cdot) - l_x^-(\cdot)$ be the Jordan decomposition of $l_x(\cdot)$ and let $l_x^*(\cdot) = l_x^+(\cdot) + l_x^-(\cdot)$. We assume that for every $x \in \mathcal{X}$,

$$u(x) = \int dl_x^*(\theta) < \infty\,.$$

Observe that if $l_x(\cdot)$ is differentiable and $\int |\partial l_x(\theta)/\partial\theta|\, d\theta < \infty$, then $u(x) = \int |\partial l_x(\theta)/\partial\theta|\, d\theta$.

If F and G are any cdf's then

$$\varrho(F, G) = \sup_{\theta\in\Theta} |F(\theta) - G(\theta)|$$

1991 *Mathematics Subject Classification*: Primary 62F15; Secondary 62C10, 62F35.
Key words and phrases: Bayes robustness, Kolmogorov metric, stability of Bayes procedures.
Supported by KBN Grant 2-1168-91-01 Gr-101.

denotes the Kolmogorov distance between F and G. For cdf's F and G of prior distributions on Θ and for a given $x \in \mathcal{X}$, let F_x and G_x be cdf's of the corresponding posterior distributions.

2. The main result. The following theorem gives us an estimate for the Kolmogorov distance between posterior distributions F_x and G_x in terms of the Kolmogorov distance between the appropriate prior cdf's F and G. Given a cdf H, let $m_x(H) = \int l_x(\theta)\, dH(\theta)$.

THEOREM. *Let F be a given prior distribution and let $x \in \mathcal{X}$ be a fixed point in the sample space. For every likelihood function $l_x(\cdot)$,*

$$\varrho(F_x, G_x) \le \frac{\varrho(F,G)}{\max\{m_x(F), m_x(G)\}}(s(x)+u(x))\,. \tag{1}$$

There exists a likelihood function for which the inequality is sharp.

Proof. Since

$$F_x(\theta) - G_x(\theta) = \frac{\int_{\theta_L}^{\theta} l_x(t)\, dF(t)}{m_x(F)} - \frac{\int_{\theta_L}^{\theta} l_x(t)\, dG(t)}{m_x(G)}\,,$$

adding and subtracting $\int_{\theta_L}^{\theta} l_x(t)\, dF(t)/m_x(G)$, we obtain

$$F_x(\theta) - G_x(\theta) = \frac{1}{m_x(G)}\Big(\int_{\theta_L}^{\theta} l_x(t)\, d(F(t)-G(t)) - F_x(\theta)\int l_x(t)\, d(F(t)-G(t))\Big)\,.$$

Integrating by parts gives

$$F_x(\theta) - G_x(\theta) = \frac{1}{m_x(G)}\Big(l_x(\theta)(F(\theta)-G(\theta)) + \int (F_x(\theta) - \mathbf{1}_{(-\infty,\theta)}(t))(F(t)-G(t))\, dl_x(t)\Big)\,.$$

Hence

$$F_x(\theta) - G_x(\theta) \le \frac{1}{m_x(G)}\varrho(F,G)(s(x)+u(x))\,.$$

Similarly,

$$G_x(\theta) - F_x(\theta) \le \frac{1}{m_x(F)}\varrho(F,G)(s(x)+u(x))\,,$$

which gives (1).

For the second statement, see Example 3 below. ∎

3. How sharp is inequality (1)? The following three examples answer that question. In each of them ε is a fixed positive number, $l_x(\cdot)$ is a fixed likelihood function, and F is a fixed prior distribution. The prior distribution G is chosen in such a way that $\varrho(F,G)=\varepsilon$.

Let *RHS* and *LHS* denote the right and left hand sides of (1), respectively.

EXAMPLE 1. Suppose that $l_x(\cdot)$ is the likelihood function of the normal distribution $N(\theta,\sigma^2)$ and that F is normal $N(0,\tau^2)$. Define

$$G(\theta)=\begin{cases}0 & \text{if } \theta<F^{-1}(\varepsilon),\\ F(\theta)-\varepsilon & \text{if } F^{-1}(\varepsilon)\le\theta<0,\\ F(\theta)+\varepsilon & \text{if } 0\le\theta<F^{-1}(1-\varepsilon),\\ 1 & \text{if } \theta\le F^{-1}(1-\varepsilon).\end{cases}$$

Then $F(\theta)-G(\theta)\equiv\varepsilon$ on the support of G and

$$RHS\le\frac{3\varepsilon}{\dfrac{\sigma}{\sqrt{\sigma^2+\tau^2}}\exp\left[-\dfrac{1}{2}\dfrac{x^2}{\sigma^2+\tau^2}\right]},$$

$$LHS\ge\Phi\left(\sqrt{\frac{1}{\sigma^2}+\frac{1}{\tau^2}}\left[\tau\Phi^{-1}(\varepsilon)-\frac{\dfrac{x}{\sigma^2}}{\dfrac{1}{\sigma^2}+\dfrac{1}{\tau^2}}\right]\right).$$

Both *RHS* and *LHS*, as well as their difference of course, tend to zero as $\varepsilon\to0$. For large σ^2 and small τ^2 the right hand side is approximately 3ε and the left hand side equals ε, and hence $RHS/LHS\approx3$. ∎

EXAMPLE 2. Let $l_x(\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x}$ and let $F(\theta)=\theta^\alpha,\ \alpha>0$. Constructing G as in Example 1 we obtain $RHS/LHS\approx2$ for small ε. ∎

EXAMPLE 3. To see that the inequality is sharp take $\mathbf{1}_{(\theta-1/2,\theta+1/2)}(x)$, $\theta\in\Theta=(0,1)$ as a family of densities on the sample space $\mathcal{X}=\mathbb{R}^1$. Then $l_x(\theta)=\mathbf{1}_{(x-1/2,x+1/2)}(\theta)$ and for $x=1/2$ one obtains $u(x)\equiv0$ and, for every cdf F on Θ, $m_x(F)=1$. Now $F_x=F$ and $G_x=G$, $s(x)=1$ and hence $LHS=RHS$. ∎

4. Bayes robustness. For a given prior distribution F consider the class of prior distributions $\mathcal{G}_\varepsilon=\{G:\varrho(G,F)\le\varepsilon\}$ (see, for example, the class Γ_1 in Berger (1985)) and the class of the corresponding posterior distributions.

As consequences of inequality (1) we can estimate the oscillation of posterior distributions under (small) violations of the assumed prior distribution and we can conclude that the posterior distribution is infinitesimally robust (in the sense of, e.g., Męczarski and Zieliński (1991) and of the papers quoted therein) under misspecification of the prior distribution.

Corollary 1 (oscillation of the posterior distribution). *For any given prior distribution F and any sample point $x \in \mathcal{X}$,*

$$(2) \qquad \sup_{G\in\mathcal{G}_\varepsilon} \varrho(F_x, G_x) \le \frac{\varepsilon}{\max\{m_x(F), \min_{G\in\mathcal{G}_\varepsilon} m_x(G)\}}(s(x)+u(x))$$

where $s(x)$ and $u(x)$ depend on the likelihood function $l_x(\cdot)$ only, and $m_x(F)$ depends on the likelihood function and the prior distribution F. ■

Corollary 2 (infinitesimal robustness). *For a fixed $x \in \mathcal{X}$, for every prior distribution F and for each $\varepsilon > 0$ there exists $\delta > 0$ such that for every distribution G on Θ,*

$$\varrho(G, F) < \delta \Rightarrow \varrho(G_x, F_x) < \varepsilon\,. \ ■$$

Corollary 3 (uniform infinitesimal robustness). *If there exist positive α, M_1, and M_2 such that*

$$m_x(F) > \alpha, \quad s(x) < M_1, \quad \textit{and} \quad u(x) < M_2$$

for all $x \in \mathcal{X}$, then for each $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x \in \mathcal{X}$ and for all distributions G on Θ,

$$\varrho(G, F) < \delta \Rightarrow \varrho(G_x, F_x) < \varepsilon\,. \ ■$$

Berger and Berliner (1986), Sivaganesan (1988), Sivaganesan and Berger (1989), Gelfand and Dey (1991), to quote but a few, considered the class $\Gamma_\varepsilon = \{(1-\varepsilon)F + \varepsilon Q : Q \in \mathcal{Q}\}$ of distributions, with a given prior distribution F and some specified $\mathcal{Q}$, and discussed the oscillations of some functionals on the appropriate class of posterior distributions. Since $\Gamma_\varepsilon \subset \mathcal{G}_\varepsilon$ we conclude that if the prior distributions belong to a given ε-contamination class Γ_ε then the posterior distributions do not differ substantially in the Kolmogorov metric. A similar conclusion holds if the prior distributions do not differ too much in the total variation metric. On the other hand, if the prior distributions do not differ much in the Kolmogorov metric the appropriate posterior distributions do not differ substantially in the Lévy or Prohorov metric (see, e.g., Zolotarev (1986) or Rachev (1991)). Taking all this into account one can say that under rather general conditions the Bayes inference is infinitesimally robust to small misspecifications of the prior distribution.

Acknowledgments. The authors are greatly indebted to Professor Lesław Gajek for his fruitful comments which enabled us to improve inequality (1).

References

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

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

A. E. Gelfand and D. K. Dey (1991), *On Bayesian robustness of contaminated classes of priors*, Statist. Decisions 9, 63–80.

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

S. T. Rachev (1991), *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester.

S. Sivaganesan (1988), *Ranges of posterior measures for priors with arbitrary contamination*, Comm. Statist. Theory Methods 17 (5), 1591–1612.

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

V. M. Zolotarev (1986), *Contemporary Theory of Summation of Independent Random Variables*, Nauka, Moscow (in Russian).

AGATA BORATYŃSKA
FACULTY OF MATHEMATICS
UNIVERSITY OF WARSAW
BANACHA 2
02-097 WARSZAWA, POLAND

RYSZARD ZIELIŃSKI
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
P.O. BOX 137
00-950 WARSZAWA, POLAND

Received on 1.7.1993