ACTA MATHEMATICA VIETNAMICA
Volume 28, Number 2, 2003, pp. 201-207

BAYESIAN ESTIMATION UNDER ESTIMATION CONSTRAINT

PHAM GIA THU AND TRAN LOC HUNG

Abstract. We suppose that a constraint is imposed on the value of the estimator in Bayesian Estimation Theory and study the distribution of this estimator under various hypotheses, using the two main loss functions: the absolute value and the quadratic.

1. Introduction

In classical Bayesian Estimation Theory the prior distribution $f(\theta)$ provides the first information on the parameter to be estimated. With the loss function considered, $L(\theta, w)$, the estimator is the value that would minimize the average loss or $\int_{\Re} L(\theta, w) f(\theta) d\theta$.

Considering the statistical model $h(x|\theta)$, let $\vec{x}$ be the observations obtained on the model and $f(\theta|\vec{x})$ the posterior distribution of θ. The optimal estimation is now the value that would minimize the average posterior loss.

In the case of a quadratic loss (Pham-Gia and Tukkan [7]) that estimation is the posterior mean, with the corresponding average loss equal the posterior variance. For the absolute value loss function, the corresponding quantities are the posterior median and the posterior mean absolute deviation (Pham-Gia, Duong and Tukkan, [6]).

In this article we suppose that the estimator is subject to a certain constraint, as it is often in real applications, and cannot take the optimal value which is either the mean or the median. That constraint could be a fixed interval [a,b], for example. Alternately, we can also suppose that the estimator itself is subject to a family of probability distributions, and will look for estimation optimality under this constraint. The behavior of the model under this non-optimal (in the classical sense) estimation will also be studied... We consider the two common loss functions $L_i(\theta, w) = K_i|\theta - w|^i$; $(i = 1, 2)$ only since difficulties associated with cases $i > 2$, as presented in Bar-Lev et al. (see [1]), do not allow the deviation of the associated solutions.

Received July 15, 2002.

1991 *Mathematics Subject Classification.* 62C10, 62D05.

Key words and phrases. Constraint, variance function, posterior risk, Bayes risk, dispersion function.

Research partially supported by NSERC Grant A9249 (Canada) and FICU Grant 2000/Pas/13, awarded jointly with the Hue Univeversity (Vietnam).

In Section 2 we will recall the main properties of the two dispersion functions that will play fundamental roles in the study of the above problem: the mean absolute deviation and the variance, both taken about an arbitrary point in $\Re$ and considered as functions of this point. In Section 3 results on the optimal estimator, the posterior risk and the Bayes risk, are presented according to two hypothesis: the constraint is a fixed interval $[a, b]$ within which the estimation must lie, and a finite set of distributions $\{W_j\}$ from which to choose the best distribution of the estimator.

2. The mean absolute deviation and the variance functions

Let Θ be a random variable defined on $\Re$, with distibution $F_\Theta(w)$ and finite mean μ. We defined the absolute deviation function of Θ at a point $w \in \Re$, also known as the dispersion function of Θ at w, as $\delta_\Theta(w) = E(|\Theta - w|)$. According to [6], we have

$$\delta_\Theta(w) = \mu - w + 2[wF_\Theta(w) - \mu\phi(w)], \tag{2.1}$$

where ϕ is the incomplete first moment of Θ, defined by

$$\phi(w) = \mu^{-1} \int_{-\infty}^{w} t dF_\Theta(t).$$

On the other hand, Munoz-Perez and Sanchez-Gomez (see for details in [4]) have given the expression of the distrbution $F_\Theta(w)$ of Θ, in terms of the derivative of $\delta_\Theta(w)$ as follows

$$F_\Theta(w) = \frac{1}{2}[\delta'_\Theta(w) + 1], \tag{2.2}$$

which can also be obtained from (2.1).

It is worth pointing out that the formula (2.1) can be as follows (see for instance [12])

$$\begin{aligned} \delta_\Theta(w) &= w - \mu + 2 \int_{x \geq w} (x - w) dF_\Theta(x) \\ &= \mu - w + 2 \int_{x < w} (w - x) dF_\Theta(x). \end{aligned} \tag{2.3}$$

or, if X be a discrete random variable with distributions $p_n = P(X = x_n)$, $n \geq 0$, then

$$\begin{aligned}\delta_\Theta(w) &= \sum_n \mid x_n - w \mid p_n \\ &= w - \mu + 2 \sum_{n: x_n \geq w} (x_n - w) p_n \\ &= \mu - w + 2 \sum_{n: x_n < w} (w - x_n) p_n. \end{aligned} \tag{2.4}$$

The $\delta_\Theta(w)$ is shown to have the following properties:

a. It is continuously differentiable a.e. and convex on $\Re$,
b. $\lim_{w \to +\infty} \delta'_\theta(w) = 1$ and $\lim_{w \to -\infty} \delta'_\theta(w) = -1$.
c. It is the L^1 distance (with respect to the Lebesgue measure on $\Re$) between F_Θ and F_w, where F_w is the distribution of the degenerate variable at the point w. We have

$$\int_\Re |F_\Theta(x) - F_w(x)| dx = \delta_\Theta(w). \tag{2.5}$$

d. The L^1 norm (with respect to the Lebesgue measure) of the difference $(\delta_\Theta(w) - \delta_\Theta(\mu))$ is the variance σ^2_Θ of Θ, where $\delta_\Theta(\mu) = |\mu - w|$ is the dispersion function of the degenerate variable at $E(\Theta) = \mu$, i.e. we have

$$\int_\Re |\delta_\Theta(w) - \delta_\Theta(\mu)| dw = \sigma^2_\Theta. \tag{2.6}$$

It is to be noticed that according to (2.6) some results related to upper bounds of the L^1 distances of two dispresion functions have been established in [18].

When μ and Md exist and are unique, the value of $\delta_\Theta(w)$ at $w = \mu$ and at $w = Md$ play particular roles in Applied Statistics and Economics, being denoted respectively by $\delta_1(\mu)$ and $\delta_2(\Theta) = \delta_\Theta(Md)$. Pham-Gia and Tukkan [7], used these measures in the study of income distributions associated with the beta distribution, while Gastwirth [2] studied the statistical properties of the standard measure of relative uniformity in tax assessments which is based on $\delta_2(\Theta)$.

Furthermore, Munoz-Perez and Sanchez-Gomez [5] have proved that $\delta_X(w) - \mu_X \leq \delta_Y(w) - \mu_Y$ is equivalent to dilation ordering, by which $X \overset{\text{dil}}{\leq} Y$ if $E[\Phi(X - \mu_X)] \leq E[\Phi(Y - \mu_Y]$, for any convex function Φ.

Some interesting results concerning the connection of the weak convergence of the random variables with the convergence of the dispersion functions have been investigated in [17]. The results obtained in [17] can be applied to some problems of theory of limit theorems for sum of independent random variables.

On the other hand, the variance of a random variable Θ, with distribution $F_\Theta(\theta)$, about any point w, denoted $V_\Theta(w)$ is defined as

$$V_\theta(w) = \int_{\Re} (\theta - w)^2 dF_\Theta(\theta) = \sigma_\Theta^2 + (w-\mu)^2, \tag{2.7}$$

where σ_Θ^2 is the variance of Θ about its mean μ. Although both δ_Θ and $V_\Theta(w)$ are convex functions, $V_\Theta(w)$ displays a symmetry about $w = \mu$ whereas $\delta_\Theta(w)$ is asymmetric, with minimum value of $\delta_2(\Theta)$ at $w = Md$. For values k_1, k_2 within some ranges, we then have 2 distinct values c_1, c_2 and d_1, d_2 such that

$$\delta_\Theta(c_1) = \delta_\Theta(c_2) = k_1$$

and

$$V_\Theta(d_1) = V_\Theta(d_2) = k_2.$$

3. Estimation under constraint of the estimator

3.1. Let us consider the simple case where the estimator of θ belongs to an interval on $\Re$, say [a,b] and let $L(\Theta, w)$ be the loss function. The posterior risk is then $\int_{\Re} L(\Theta, w) f(\theta|\vec{x}) d\theta$ and let θ_0 be the value that would minimize that risk inside [a,b], i.e. $\lim\limits_{w\in[a,b]} \int_{\Re} L(\Theta, w) f(\theta|\vec{x}) d\theta$ is attained for $w = w_0$.

We then take $\hat{\Theta} = w_0$. The posterior risk is $\int_{\Re} L(\Theta, w_0) f(\theta|\vec{x}) d\theta$ and the Bayes risk is $\int_{\Re} \Psi_n(x)dx = \int_{\Re} L(\Theta, w_0) f(\theta|\vec{x}) d\theta$, where Ψ_n is the predictive distribuiton of x, based on $\vec{x}$, a vector of observations of size n.

Consider the following loss functions: the absolute value one of the form

$$L_1(\Theta, w) = K_1|w - \theta|, \quad K_1 > 0,$$

and the quadratic one of the form

$$L_2(\Theta, w) = K_2|w - \theta|^2, \quad K_2 > 0.$$

Without loss of generality, we can take $K_1 = K_2 = 1$. The associated loss functions are then called unit loss functions and results for the cases $K_1 \neq 1$ can be deduced from these cases by multiplication by appropriate constants.

For a fixed Θ, $\delta_\Theta(w_0)$ represents the average loss (also called risk) that results from taking w_0 as the estimate of Θ, using the L_1 loss function. A similar conclusion holds for $V_\Theta(w_0)$ and L_2.

3.2. The problem of building a normed linear space structure on a subset of probability distributions has not found a satisfactory answer yet. The dispersion measure $\delta_2(\Theta)$ can provide a solution to that problem.

Let us consider the class of infinitely divisible distributions on $\Re$ with the following operations on its independent members: convolution product $F_1 * F_2$, inverse function $F^{-1}(a) = \inf\{x : F(x) > a\}$, with $Md(F) = F^{-1}(\frac{1}{2})$, and degenerate distributions at any point are all identified to the neutral element.

Theorem 1. *Let Ξ be the class of infinitely divisible distributions. The Ξ is a closed normed linear space, with $\|F\| = \delta_2(F)$.*

Proof. We know that Ξ is closed under convolution and passage to the limit. Moreover, for independent X, F_1 and F_2 in Ξ, such that $F_1 * X = F_2 * X$, we have $F_1 = F_2$. Hence Ξ is a linear space. Moreover, defining the norm of F as above, we have $\|F\| = 0$ iff F is degenerate. Also, for $a \in \Re$, we have $\|aF\| = \delta_2(aX) = a\delta_2(X) = a\|F\|$ since $\delta_2(X)$ is a measure of dispersion. Then

$$\|F_1 * F_2\| = \delta_2(X_1 + X_2) \leq \delta_2(X_1) + \delta_2(X_2) = \|F_1\| + \|F_2\|.$$

The first part of the double inequality

$$\delta_2(X_1) \leq \delta_2(X_1 + X_2) \leq \delta_2(X_1) + \delta_2(X_2)$$

is given by the dilation ordering (see [5]). For the second part, we note that

$$Pr(X + Y \leq Md(X) + Md(Y)) \geq Pr(X \leq Md(X))) + Pr(Y \leq Md(Y)).$$

Hence

$$Md(X + Y) \leq Md(X) + Md(Y)$$

and

$$E(|X + Y - Md(X + Y)|) \leq E(|X - Md(X)|) + E(|Y - Md(Y)|).$$

□

3.3. There are several possible constraints that could be imposed on an estimator in real applications and they could be of deterministic or stochastic types. In the following section we will consider the simple deterministic constraint that the value of the estimations is to be within a fixed interval $[a, b]$, with $a < b$.

Theorem 2. *Let Θ be the parameter of interest in the statistical model $\varphi(x|\theta)$ and let $f(\theta)$ be its prior. If the estimation of Θ must be in $[a, b]$ then it can only take the value a, or b, or $Md(\Theta)$.*

Proof. Under the unit absolute value loss function, we have to consider the posterior distribution of Θ, and find the value of the estimation that minimizes the posterior risk: $\min\limits_{w\in[a,b]}(\delta_\theta(w)) \geq \delta_2(\theta)$. Hence, due to the convexity of the function $\delta_\Theta(w)$, which has a minimum at Md_{post}, with the corresponding posterior risk as $\delta_{2,\text{post}}(\Theta)$ or, for the second case, b (if $Md_{\text{post}} > b$), with as posterior risk $\delta_{\Theta_{\text{post}}}(b)$, or a ($Md_{\text{post}} < a$), with as posterior risk $\delta_{\Theta_{post}}(a)$.

Let $\Psi = \int_\Re f(\theta)h(x|\theta)d\theta$ be the predictive distribution of X. The Bayes risk is hence $E_X(\delta_{\Theta_{\text{post}}}(.))$, with $. = a$ or b, or Md_{post}.

Similarly, for the quadratic loss function, we have $\min\limits_{w\in[a,b]}(V_\Theta(w)) \geq \sigma^2(\Theta)$, where the convex function $V_\Theta(w)$ attains its minimum $\sigma^2(\Theta)$ at $w = \mu_{\text{post}}$. The distributions of the corresponding errors $U = |\Theta - w|$ and $V = (\Theta - w)^2$, for the estimated values determined above, have their means and variances given by

$$E(U) = \delta_\Theta(w), \quad Var(U) = \sigma_\Theta^2 - \delta_\Theta^2(w) + (\mu_\Theta - w)^2,$$

while

$$E(V) = \sigma_\Theta^2 - (\mu_\Theta - w)^2 \quad \text{and} \quad Var(V) = E[(\Theta - w)^4] + (\sigma_\Theta^2 + (\mu_\Theta - w)^2)^2,$$

with, in both cases, $w = a$ or b, or μ_{post}. □

3.4. For a constraint of stochastic type, we now consider the case when W itself could be a random variable with domain Ω. In this case, we are often in presence of a family of random estimators $\{W_j\}$, $j = 1, 2, \ldots, n$ and would like to identify the best distribution of the family. In the posterior distribution, the estimation of Θ now follows the distribution of W, does not depend on the pointwise minimization of the posterior risk, but depends rather on the global minimization.

Theorem 3. *Let $\{W_j\}$, $j = 1, 2, \ldots, n$ be a family of distributions of the estimator of Θ. We have, for the absolute value loss function, the worst random decision distribution given by $\max\limits_{1 \le j \le n} \{\delta_\Theta(E(W_j))\}$, and under the square-error loss function, by $\max\limits_{1 \le j \le n} \{V_\Theta(E(W_j))\}$.*

Proof. As shown by Munoz-Perez and Sanchez-Gomez in [4], $\delta_\Theta(w)$ is a convex function of w. Hence we have $EW_j(\delta_\Theta(w_j)) \ge \delta_\Theta(E(W_j))$, for $j = j_0$ identifies W_{j_0} as the worst distribution. By step-to-step elimination, we can obtain the best distribution W^*.

Under the square error loss function, we have $E_W(V_\Theta(w)) \ge V_\Theta(E(W))$, and a similar conclusion is valid.

Thus, for the corresponding error variables $U^* = |\Theta - W^*|$ and $V^* = (\Theta - W^*)^2$ with Θ^* and W^* being random variables, we have

$$E(U^*) = \iint\limits_{\Re^2} |\theta - w| dF(w) dF(\theta),$$

$$Var(U^*) = \iint\limits_{\Re^2} |\theta - w|^2 dF(w) dG(\theta) - [E(U^*)]^2$$

while

$$E(V^*) = \iint\limits_{\Re^2} |\theta - w| dF(w) dG(\theta),$$

$$Var(V^*) = \iint\limits_{\Re^2} |\theta - w|^4 dF(w) dG(\theta) - [E(V^*)]^2.$$

□

References

[1] S. K. Bar -Lev, B. Boukai and P. Enis, *On the mean square error, the mean absolute error and the like*, Comm. Statist. -Theory Meth. **28** (1999), 1813-1822.

[2] J. J. Gastwirth, *Statistical properties of a measure of tax assessment uniformly*, J. Stat. Planning Infer. **6** (1982), 1-12.

[3] M. Loeve, *Probability Theory I*, 4th edition, Springer-Verlag, New York, 1977.

[4] J. Munoz-Perez and Sanchez-Gomez, *A characterization of the distribution function: the Dispersion function*, Statistic and Probability Letters **10** (1990), 235-239.

[5] J. Munoz-Perez and Sanchez-Gomez, *Dispersion ordering by dilation*, J. Applied Probability **27** (1990), 440-444.

[6] T. Pham-Gia, Q. P. Duong and N. Tukkan, *Using the mean deviation to determine the prio distribution*, Stat. Prob. Letters **13** (1993), 273-281.

[7] T. Pham-Gia and N. Tukkan, *Sample size determination in Bayesian analysis*, The Statisticians **4** (1992), 389-397.

[8] T. Pham-Gia and N. Tukkan, *Estimating the beta distribution from its Lorenz curve*, Mathematical and Computer Modelling (2001).

[9] T. Pham-Gia, and T. L. Hung, *The mean and median absolute deviations*, Mathematical and Computer Modelling **34** (2001), 921-936.

[10] T. Pham-Gia, *Some applications of the Lorenz curve in decision analysis*, Amer. J. Math. Manag. Sciences **15** (1995), 1-34.

[11] T. Pham-Gia, *Noise effects on Bayesian decisian criteria under absolute value loss function*, Theory and Decision (submitted).

[12] Tran Loc Hung, *On results concerning the mean absolute deviations*, Bulletin of Science, College of Pedagogy, Hue University, N^0 1(32) (1999), 39-45.

[13] Tran Loc Hung and Pham-Gia Thu, *On the mean absolute deviation of the random variables*, Vietnam National University, Journal of Science t.XV (1999) $N^0$5, 36-44.

[14] Pham-Gia Thu and Tran Loc Hung, *On results concerning the dispersion function of the random variables*, Bulletin of Science, College of Pedagogy, Hue University. (Submited).

[15] T. Pham-Gia and Tran Loc Hung, *On the dispersion function as a generalization of the absolute deviations of random variables*, 2002 (to appear)

[16] Pham-Gia Thu and Tran Loc Hung, *On the L^1-norm approach and its application in some statistical problems*, Proc. Second National Conference on Probability and Statistics, Research, Applications and Training, Hanoi, 2-4 November, 2001, 165-181, (in Vietnamese).

[17] Tran Loc Hung and Nguyen Van Son, *On the connection of the weak convergence of the random variables with convergence of dispersion functions*, Vietnam J. Math. (to appear)

[18] Tran Loc Hung, *Some remarks related to L^1-norm distances of the dispersion functions*, The Hue University Journal of Research, $N^0$10 (2002), 17-20, (in Vietnamese).

Department of Mathematics and Statistics
University of Moncton,
Moncton, New Brunswik, Canada E1A 3E9

Department of Mathematics, College of Sciences
77 Nguyen Hue str., Hue, Vietnam

E-mail address: `tlhung@hueuni.edu.vn`